

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 716 867**

51 Int. Cl.:

C12N 9/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.05.2014 PCT/EP2014/061189**

87 Fecha y número de publicación internacional: **04.12.2014 WO14191527**

96 Fecha de presentación y número de la solicitud europea: **28.05.2014 E 14728899 (7)**

97 Fecha y número de publicación de la concesión europea: **02.01.2019 EP 3004338**

54 Título: **Endonucleasa de asentamiento LAGLIDADG que escinde el gen de receptor de células T alfa y usos de la misma**

30 Prioridad:

31.05.2013 DK 201370303

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

17.06.2019

73 Titular/es:

**CELLECTIS SA (50.0%)
8 rue de la Croix Jarry
75013 Paris, FR y
PRECISION GENOME ENGINEERING, INC.
(50.0%)**

72 Inventor/es:

**ASTRAKHAN, ALEXANDER y
JARJOUR, JORDAN**

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 716 867 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

Endonucleasa de asentamiento LAGLIDADG que escinde el gen de receptor de células T alfa y usos de la misma

Campo de la invención

5 La presente divulgación se refiere a biología molecular y celular, genética, genómica y sus aplicaciones en terapéutica humana. Aspectos particulares se refieren a una endonucleasa de corte poco común que escinde una secuencia diana de ácido nucleico a partir del gen TCR-alfa, más particularmente a una nueva variante de meganucleasas de *I-Onu1* u homólogos que es particularmente eficaz en la alternación de la expresión de este gen en células T, y al uso de la misma para terapia contra el cáncer.

Antecedentes de la invención

10 Las nucleasas específicas de sitio son reactivos poderosos para seleccionar como diana y modificar específica y eficazmente una secuencia de ADN dentro de un genoma complejo. Las roturas de ADN bicatenario provocadas por nucleasas específicas de sitio se reparan comúnmente a través de distintos mecanismos de recombinación homóloga o unión de extremos no homólogos (NHEJ). Mientras que la recombinación homóloga usa normalmente la cromátida hermana del ADN dañado como matriz donadora a partir de la cual realizar una reparación perfecta de la
15 lesión genética, NHEJ es un proceso de reparación imperfecto que a menudo da como resultado cambios en la secuencia de ADN en el sitio de la rotura bicatenaria. Los mecanismos implican volver a unir lo que queda de los dos extremos de ADN a través de religamiento directo (Critchlow y Jackson 1998) o por medio de la denominada unión de extremos mediada por microhomología (Ma, Kim *et al.* 2003). La reparación por medio de unión de extremos no homólogos (NHEJ) da como resultado a menudo pequeñas inserciones o deleciones y puede usarse para la
20 creación de inactivaciones génicas específicas. Hay numerosas aplicaciones de modificación por ingeniería genética del genoma mediante nucleasas específicas de sitio que pasan de la investigación básica a aplicaciones bioindustriales y terapéutica humana. Volver a modificar por ingeniería genética una proteína de unión a ADN para este fin se ha visto limitado principalmente a la endonucleasa de direccionamiento LADLIDADG (LHE) que se produce de manera natural, proteínas de dedos de zinc (ZFP) artificiales, las nucleasas efectoras de tipo activador de la transcripción (nucleasas TALE) y el sistema de CRISPR-Cas recientemente descrito.

Las endonucleasas de asentamiento, también conocidas como meganucleasas, son endonucleasas específicas de secuencia con sitios de escisión grandes (>14 pb) que pueden suministrar roturas bicatenarias de ADN en loci específicos (Thierry y Dujon 1992). Hay un puñado de familias de endonucleasas de asentamiento conocidas que se delimitan basándose en motivos canónicos y las características estructurales que comprenden. Sin embargo, todas
30 comparten la propiedad de reconocer y escindir dianas de ADN largas. Endonucleasas de asentamiento fueron las primeras, y hasta la fecha las únicas, endonucleasas que se producen de manera natural con especificidades a o que se aproximan al 'nivel genómico', lo que significa que tienen secuencias diana supuestas que aparecen muy infrecuentemente, o quizá excepcionalmente, en su genoma huésped. Como propiedad general, las HE tienen un grado moderado de fidelidad a sus secuencias diana de ADN, de manera que la mayoría de las sustituciones de
35 pares de bases en sus secuencias diana de ADN reducen o eliminan la capacidad de la HE de unirse a la misma o escindirla. Las HE son, por tanto, las endonucleasas que se producen de manera natural más específicas aún descubiertas, y de hecho esta propiedad es crítica para el ciclo de vida natural de los elementos genéticos en los que están codificadas.

Los genes de endonucleasas de asentamiento (HEG) se clasifican como un tipo de elemento genético egoísta, ya que su actividad de reconocimiento y escisión de ADN puede conducir a un acontecimiento de reparación del ADN que da como resultado el copiado del HEG en el sitio de escisión. Este mecanismo de transferencia génica horizontal, denominado 'asentamiento' da como resultado un patrón de herencia super-mendeliano. Usando este mecanismo, los HEG y sus productos génicos de endonucleasa pueden propagarse rápidamente dentro de sus poblaciones de especies huésped, y también se han propagado a través de todos los reinos de la vida a lo largo del
45 tiempo evolutivo. Los HEG se encuentran lo más comúnmente en ubicaciones genómicas altamente conservadas que no confieren costes de aptitud a sus organismos huésped, tal como dentro de intrones o como fusiones N- o C-terminales no disruptivas a proteínas huésped.

La familia de endonucleasas de asentamiento LAGLIDADG (LHE) comprende un grupo de nucleasas compactas (< 320 aminoácidos) cuyas propiedades estructurales y mecánicas se han estudiado extensamente debido a sus propiedades atractivas para aplicaciones de modificación por ingeniería genética del genoma. Las LHE funcionan o bien como dímeros o bien como monómeros pseudo-diméricos, apareciendo el sitio activo de escisión del ADN en la cara orientada hacia el ADN de la superficie de contacto de las dos subunidades (en LHE diméricas) o dominios (en LHE monoméricas). Los motivos de consenso de LAGLIDADG para los que se nombran LHE se encuentran en las dos alfa hélices centrales que forman esta superficie de contacto entre las dos subunidades o dominios. En la parte inferior de cada hélice de LAGLIDADG están los residuos que juntos coordinan la reacción de hidrólisis si se cumplen las condiciones apropiadas, tales como si la LHE encuentra y se une a una secuencia diana de ADN apropiada. El sitio activo cubre las bases de ADN 'centrales-4' de la secuencia diana de ADN.

En cualquier lado del sitio activo están los dos dominios de unión a ADN que las LHE usan para reconocer sus

secuencias diana de ADN. Cada dominio comprende una lámina beta antiparalela que envuelve casi un giro completo de ADN y entra en contacto con 9 pares de bases de la secuencia de ADN. Los miembros de la familia de LHE reconocen, por tanto, secuencias diana de ADN de 22 pares de bases (9 pares de bases para cada dominio y 4 pares de bases cubiertas por el sitio activo), que son parcialmente palindrómicas en el caso de LHE diméricas, pero que pueden ser totalmente asimétricas para LHE monoméricas. Surgiendo de cada lámina beta antiparalela están las cadenas laterales de aminoácidos que comprenden la superficie de contacto de reconocimiento de ADN. Aunque existe mucha conservación de aminoácidos a lo largo de todos los residuos que no forman la superficie de contacto con el ADN entre la familia de LHE, las composiciones de aminoácidos de la superficie de contacto de reconocimiento de ADN varían significativamente. Esto se debe a que cada para cada LHE, la superficie de contacto de reconocimiento de ADN comprende una red extensa de contactos de cadena lateral con cadena lateral y cadena lateral con ADN, la mayoría de los cuales son necesariamente únicos para una secuencia diana de ADN de LHE particulares. La composición de aminoácidos de la superficie de contacto de reconocimiento de ADN (y la correspondencia de la misma con una secuencia de ADN particular) es, por tanto, la característica definitiva de cualquier LHE natural o modificada por ingeniería genética. La superficie de contacto de reconocimiento de ADN funciona determinando la identidad de la secuencia diana de ADN que puede alojarse e hidrolizarse y también las propiedades de afinidad y especificidad que definen la calidad de la LHE según las exigencias de la aplicación.

Debido a su pequeño tamaño y propiedades de especificidad exquisitas, las LHE han sido objeto de numerosos esfuerzos para modificar por ingeniería genética sus propiedades de reconocimiento de ADN con el desenlace deseado de escindir y alterar genes de interés en investigación, biotecnología, ciencia de cultivos, salud global y aplicaciones terapéuticas humanas. Sin embargo, la extensión de las redes de residuos que forman la superficie de contacto de reconocimiento de ADN ha impedido generalmente métodos eficaces para redirigir LHE a secuencias diana de ADN de interés. Esto ha conducido a una innovación continuada en el campo de modificación por ingeniería genética de nucleasas específicas de genes, con tres plataformas alternativas de endonucleasas validadas en la actualidad que tienen la capacidad de seleccionar como diana secuencias de ADN con niveles variables (pero generalmente altos) de especificidad, así como métodos nuevos y mejorados para superar los desafíos de la modificación por ingeniería genética de las superficies de contacto de reconocimiento de ADN de LHE.

Las nucleasas de dedos de zinc (ZFN) que se generan fusionando una pluralidad de dominios de unión a ADN basados en dedos de zinc a un dominio catalítico independiente (Kim, Cha *et al.* 1996; Smith, Berg *et al.* 1999; Smith, Bibikova *et al.* 2000) representan otro tipo de nucleasa modificada por ingeniería genética comúnmente usada para estimular la selección como diana de genes y se han usado satisfactoriamente para inducir corrección génica, inserción génica y delección génica en aplicaciones terapéuticas y de investigación. Las ZFN arquetípicas se basan en el dominio catalítico de la enzima de restricción de tipo IIS FokI y dominios de unión a ADN basados en dedos de zinc compuestos por sucesiones de 3 ó 4 dedos de zinc individuales, reconociendo cada uno un triplete de ADN (Pabo, Peisach *et al.* 2001). Dos monómeros de dedo de zinc-FokI tienen que unirse a sus sitios de reconocimiento de ADN de dedos de zinc respectivos sobre hebras opuestas en una orientación invertida con el fin de formar un dímero catalíticamente activo que cataliza la escisión bicatenaria (Bitinaite, Wah *et al.* 1998).

Efectores de tipo activador de la transcripción (TALE) fueron la siguiente plataforma de nucleasas artificiales. Los TALE, derivados de una familia de proteínas usadas en el proceso de infección por patógenos vegetales del género *Xanthomonas* o *Ralstonia*, son proteínas repetitivas caracterizadas por 14-20 repeticiones de 33-35 aminoácidos que difieren esencialmente en dos posiciones. Cada par de bases en la diana de ADN entra en contacto con una única repetición, resultando la especificidad de los dos aminoácidos variantes de la repetición (el denominado dipéptido variable repetido, RVD). La aparente modularidad de estos dominios de unión a ADN se ha confirmado en un cierto grado por el ensamblaje modular de proteína derivada de TALE diseñada con nuevas especificidades (Boch, Scholze *et al.* 2009; Moscou y Bogdanove 2009). Muy similares a ZFN, los TALE se adaptaron fácilmente a nucleasas específicas de sitio alineando repeticiones de TALE con los RVD correspondientes a la secuencia diana de elección y fusionando el alineamiento resultante con un dominio FokI. Como tal, la escisión de ADN mediante una nucleasas TALE requiere dos regiones de reconocimiento de ADN que flanquean a una región central inespecífica. Las nucleasas TALE han proliferado ampliamente desde 2010 debido a su facilidad de producción y eficacia de generación de roturas bicatenarias mejorada.

De estas distintas tecnologías, es importante distinguir las propiedades aventajadas de cada una y determinar modos innovadores para capturar estas propiedades para las aplicaciones de modificación por ingeniería genética del genoma. Una de las aplicaciones más poderosas de la tecnología de nucleasas específicas de sitio es en el campo de la terapéutica humana, que requiere el uso de reactivos de nucleasa altamente eficaces y específicos para editar de manera segura y eficaz la información genética en células o tejidos humanos. Un ejemplo relevante es el campo de la inmunoterapia contra el cáncer, que está a la vanguardia de aplicación de los avances tecnológicos de nucleasas para desarrollar productos terapéuticos novedosos. Estos enfoques se centran en aprovechar las poderosas actividades antitumorales de células T derivadas del paciente (autólogas) o derivadas de donante (alógenas) y hacer uso de este potencial por medio de la modificación por ingeniería genética del genoma de propiedades intrínsecas de la célula tales como proliferación, injerto, migración o longevidad celular. La fabricación satisfactoria y escalable de células T dotadas de actividad anticancerígena potenciada requiere la generación de composiciones de nucleasas altamente eficaces y estrategias de suministro simplificadas, tales como las descritas en algunos aspectos de esta solicitud.

El sistema inmunitario tiene un papel fundacional clave en la detección y prevención del desarrollo de cáncer en seres humanos. La mayoría de las células transformadas son detectadas rápidamente por centinelas inmunitarios y destruidas a través de la activación de células T específicas de antígeno por medio de receptores de células T (TCR) expresados clonalmente. La oncogénesis es, por tanto, un trastorno inmunológico, un fallo del sistema inmunitario en el montaje de la respuesta antitumoral necesaria para suprimir y eliminar duraderamente la enfermedad. Determinadas intervenciones de inmunoterapia desarrolladas a lo largo de las últimas décadas, tales como infusiones de citocinas recombinantes, se han centrado específicamente en la potenciación de la inmunidad de células T, y aunque estas se han asociado con casos esporádicos de remisión de la enfermedad, no han tenido un éxito global sustancial. Terapias recientes con anticuerpos monoclonales que seleccionan como diana moléculas que inhiben la activación de células T, tales como CTLA-4 o PD-1, han mostrado un efecto antitumoral más sustancial, sin embargo estos tratamientos están asociados con toxicidad sustancial debido a activación inmunitaria sistémica. Lo más recientemente, se han explorado y sometido a prueba estrategias terapéuticas que se basan en el aislamiento, la modificación, expansión y reinfusión de células T en ensayos clínicos en fase temprana. Estos tratamientos han mostrado tasas de éxito mixtas, pero varios pacientes han experimentado respuestas objetivas sin precedentes y remisiones duraderas, resaltando el potencial de las inmunoterapias contra el cáncer basadas en células T. En el presente documento se describen estrategias de edición del genoma que están diseñadas para aprovechar este potencial para la implementación extendida satisfactoria de inmunoterapias contra el cáncer de células T.

El reconocimiento satisfactorio de antígenos asociados con células tumorales por células T citolíticas inicia la lisis tumoral dirigida y respalda cualquier enfoque de inmunoterapia contra el cáncer eficaz. Algunos tumores contienen células T infiltrantes de tumores (TIL) que expresan TCR específicamente dirigidos a antígenos asociados a tumores; sin embargo el acceso a números sustanciales de TIL está limitado a sólo unos cuantos cánceres humanos. En respuesta a esta limitación, se han ideado el reconocimiento de antígenos artificiales y transgenes de señalización denominados receptores de antígenos quiméricos (CAR) para ampliar el alcance y la utilidad de la inmunoterapia contra el cáncer basada en células T. Los CAR son proteínas que atraviesan la membrana cuyas porciones extracelulares contienen dominios de reconocimiento de antígenos lo más normalmente derivados de fragmentos variables de cadena sencilla (scFv) de anticuerpos monoclonales, y cuyos dominios intracelulares contienen combinaciones de dominios de señalización para imitar las señales de activación de tipo TCR. Se ha demostrado ampliamente que células T humanas primarias que se hace que expresen CAR pueden responder a y destruir células que llevan el antígeno reconocido por el dominio scFv.

A pesar de los resultados iniciales altamente prometedores con células T transgénicas que expresan CAR, la eficacia, seguridad y escalabilidad de las inmunoterapias de células T basadas en CAR están limitadas por la expresión continua de TCR derivados clonalmente. La expresión de TCR residual puede interferir con la señalización de CAR en células T modificadas por ingeniería genética o pueden iniciarse respuestas patológicas y fuera de diana a auto o aloantígenos. En consecuencia, las células T basadas en CAR sólo se han usado en aplicaciones autólogas. La abolición genética de TCR endógeno a través de edición génica mediada por nucleasas reduciría el riesgo de respuestas colaterales dañinas y disminuiría la posibilidad de enfermedad de injerto contra huésped (EICH) mediada por células T. El principal obstáculo para desarrollar una terapia con células T alogénicas universal es el desarrollo de EICH a través de la activación de TCR de células T de los donantes por el complejo de HLA de los receptores. La eliminación del TCR impediría tales respuestas de injerto contra huésped y permitiría el desarrollo de terapias de células T alogénicas sencillas y ampliamente aplicables. Torikai *et al.*, Blood volumen 119, número 24, páginas 5697 a 5705 (14-06-2012) describen células T modificadas por ingeniería genética que expresan un receptor de antígeno quimérico y que carecen de expresión de TCR endógeno mediante la acción de nucleasas de dedos de zinc.

Además del cáncer, están desarrollándose terapias con células T para una amplia gama de aplicaciones terapéuticas incluyendo infecciones virales crónicas, enfermedad autoinmunitaria y trasplante de células madre. En modelos de enfermedad y modelos clínicos iniciales, han mostrado un papel clave del subconjunto de células T reguladoras (T-reg) en el control del desarrollo y la magnitud de EICH y diversas enfermedades autoinmunitarias. La transferencia de células T reguladoras mejora EICH en pacientes que reciben trasplante de células madre. Además, la transferencia de células T reguladoras mejoró el desenlace de la enfermedad en modelos preclínicos de artritis reumatoide, diabetes tipo 1 y lupus eritematoso sistémico, entre otros. Este enfoque está también sometiéndose a prueba en pacientes con infecciones virales crónicas tales como hepatitis B (VHB). Las células T modificadas por ingeniería genética que contienen CAR específicos de VHB son altamente activas contra células infectadas con VHB. Estos enfoques están sometiéndose a prueba en ensayos clínicos, sin embargo su uso está limitado por los obstáculos de fabricación y escalabilidad asociados con otras terapias autólogas. La combinación de direccionamiento genético de TCR-alfa en células T con CAR que seleccionan como diana dianas virales o de tolerancia representa un modo muy poderoso para desarrollar una terapia de células T alogénicas para el tratamiento de una enfermedad humana.

Sumario de la invención

Una estrategia de modificación por ingeniería genética del genoma para generar productos de células T terapéuticos, tal como en el tratamiento de enfermedades humanas, requiere el uso de reactivos de endonucleasas seguros y eficaces para alterar el gen de TCR-alfa. Se ha caracterizado recientemente que la endonucleasa *I-Onul*,

codificada dentro de un intrón del grupo I en el gen huésped Rps3 de *Ophiostoma novo-ulmi* subsp. *americana*, y sus homólogos estrechamente relacionados, son proteínas monoméricas que presentan las características de las endonucleasas de asentamiento LAGLIDADG y que son suficientemente activas para su uso en la edición del genoma (documento WO2011/156430, (Seturaman, Majer *et al.* 2009; Takeuchi, Lambert *et al.* 2011)).

5 En aspectos particulares, se crearon varias variantes de I-Onul en un intento por seleccionar como diana diferentes secuencias de ADN secuencias en el gen de TCR-alfa. En aspectos adicionales, se proporciona una nueva variante de LHE que selecciona como diana un dominio constante del gen de TCR-alfa. Esta variante de I-Onul particular
10 mostró una alta eficacia en la alteración de la expresión de TCR-alfa en células T. En aspectos adicionales, esta variante particular de la invención se fusionó entonces con algunos dominios de unión a ácido nucleico modificados por ingeniería genética, para formar endonucleasas quiméricas que también mostraban propiedades mejoradas, especialmente aumentos en la especificidad y eficacia que se requieren para obtener reactivos seguros y útiles para
15 tratar células humanas primarias. Estas moléculas han demostrado eficacia para la edición del genoma en el locus de TCR-alfa y serán útiles en numerosos métodos basados en células T para tratar enfermedad humana.

La presente invención proporciona, por tanto, una I-Onul u homólogo de I-Onul, en la que se han introducido mutaciones para obtener una variante que escinde una secuencia de ácido nucleico diana dentro del gen de receptor de células T alfa constante (TRAC); comprendiendo dicha variante de I-Onul u homólogo de I-Onul al
20 menos 10 sustituciones de aminoácido en posiciones seleccionadas del grupo que consiste en: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 138, 159, 168, 178, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 207, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240, en referencia a SEQ ID NO: 2, y que tiene una identidad de secuencia de al menos el 70% con la secuencia proteica de
25 SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10.

Breve descripción de las tablas y las figuras

Para una mejor comprensión de la invención y mostrar cómo la misma puede efectuarse, se mostrarán ahora a modo de ejemplo sólo realizaciones específicas, métodos y procedimientos según la presente invención con
30 referencia a los dibujos adjuntos en los que:

La tabla 1 muestra las posiciones de los residuos de aminoácido en la LHE que selecciona como diana TCRA_S02 que se variaron o se alteraron de otra forma en relación con la secuencia primaria de la proteína I-Onul silvestre durante el proceso de reespecificación. La LHE de TCRA_S02_2E5 contiene variaciones sólo en los 44 residuos que comprenden la superficie de contacto proteína-ADN, no todos de los cuales se reajustaron a aminoácidos diferentes
35 con respecto a la proteína I-Onul silvestre, pero todos de los cuales se variaron en las fases iniciales de la reespecificación. La variante con mejor rendimiento tras el examen por refinamiento tenía seis mutaciones adicionales, 5 de las cuales se ubican dentro de la superficie de contacto proteína-ADN, 1 de las cuales está en otra parte de la proteína.

Figura 1: Representa la ubicación de trece secuencias diana supuestas, indicadas como Onu_S##, en el primer exón del gen de TRAC humano para el que se predicen secuencias de reconocimiento de LHE-ADN superiores. Se indica la ubicación del primer exón del gen de TRAC.
40

La figura 2 muestra esquemática y estructuralmente la ubicación de la superficie de contacto proteína-ADN que define la interacción entre una endonucleasa de asentamiento LAGLIDADG (LHE) y su sustrato de ADN. La ilustración esquemática representa generalmente el concepto de que hay una región continua de la LHE que comprende la interacción con ADN. Las imágenes estructurales demuestran en más detalle la naturaleza de esta interacción, mediante la cual residuos de la superficie de contacto proteína-ADN de la LHE (cuyos átomos de cadenas laterales se muestran como esferas negras) encajan en los surcos mayores de la hélice de ADN. Es la constelación de átomos de cadenas laterales de la superficie de contacto la que determina la complementariedad de una LHE natural o modificada por ingeniería genética con los átomos de los nucleótidos de ADN, que forman por sí mismos patrones específicos.
45

La figura 3 muestra que, de los dos sitios diana que se eligieron para la modificación por ingeniería genética de la superficie de contacto proteína-ADN, sólo uno (TCRA_S02) produjo LHE variantes que podían escindir la secuencia diana completa. Una de estas variantes se muestra en comparación con la enzima I-Onul y su diana relacionada. Los paneles representan análisis de citometría de flujo de hidrólisis de ADN, mediante lo cual la levadura de panadero, *Saccharomyces cerevisiae*, expresa la LHE en la superficie de sus células y se consultan con sustratos de ADN sintéticos marcados con colorantes fluorescentes tal como se ha publicado. En resumen, se tiñen en primer lugar las muestras con un anticuerpo biotinilado frente a un epítopo añadido al extremo N- o C-terminal de la LHE. Durante este procedimiento de tinción, se generan conjugados de estreptavidina marcada con ficoeritrina (eje x) con sustratos de ADN sintéticos marcados con biotina y Alexa fluor-647 (eje y) a una molaridad relativa que conserva algunos sitios de unión a biotina en la estreptavidina. Estos preconjugados se usan entonces para teñir las células de levadura, dando como resultado el perfil de estreptavidina-E/Alexa fluor-647 colineal. Se usan entonces condiciones que inhiben la escisión (Ca²⁺) y que permiten la escisión (Mg²⁺) para determinar si la LHE nativa o modificada por ingeniería genética escinde la diana unida, que, si se escinde, pierde la señal en el eje y debido a pérdida del fluoróforo Alexa fluor-647.
50
55

La figura 4 muestra la eficacia de selección como diana inicial de la variante TCRA_S02_2E5 y las mejoras progresivas en la eficacia de selección como diana logradas mediante el proceso de refinamiento de la actividad. Se midió la eficacia de selección como diana usando un indicador fluorescente de roturas bicatenarias integrado cromosómicamente denominado "indicador de semáforo" (*traffic light reporter*, TLR). Se construyeron fibroblastos de riñón embrionario humano 293T (HEK 293T) para que contuvieran la secuencia de ADN de TCRA_S02 inmediatamente en el sentido de 5' de una proteína fluorescente mCherry fuera de marco (eje y) que, tras uno de tres posibles desenlaces de marco del proceso de reparación del ADN mediante NHEJ, se vuelve fluorescente. El porcentaje de células en el eje y representa, por tanto, aproximadamente 1/3 de todos los acontecimientos de reparación mediados por nucleasa imprecisos. Esta línea celular se transfectó entonces con ARNm transcrito *in vitro* generado sintéticamente (ARNm-IVT) que codifica para las LHE que seleccionan como diana TCRA_S02, con o sin ARNm que codifica para la exonucleasa Trex2. La variante TCRA_S02_2E5 original provocó ineficazmente roturas bicatenarias y, por tanto, produjo sólo pequeños porcentajes de células positivas para mCherry (2ª fila). Dos rondas de cribado por refinamiento de la actividad condujeron a grandes mejoras en la generación de células positivas para mCherry.

La figura 5 muestra una alineación comparativa de la variante TCRA_S02_2E5 y su derivado, TCRA_S02_RD2_8, que se identificó en la segunda ronda de refinamiento de la actividad. Los motivos de hebra-bucle-hebra que comprenden el dominio de unión a ADN se representan por encima de las secuencias alineadas.

La figura 6 muestra un ensayo mediante el cual se usa titulación de la unión a ADN para establecer las propiedades de afinidad de dos LHE diferentes que seleccionan como diana el sitio TCRA_S02. Se incubaron independientemente muestras de levadura que presentan cada variante de LHE con concentraciones crecientes de sustratos de ADN sintéticos marcados con colorantes fluorescentes (eje y). Se incluyó también un anticuerpo frente a un epítipo C-terminal (eje x) de manera que pudo normalizarse la actividad de unión a ADN a la cantidad de proteína LHE expresada sobre la superficie de la levadura, creando el patrón colineal, con una señal superior en el eje y por señal de eje x en donde la afinidad es mayor. Los resultados demuestran que la LHE I-Onul (Onu silvestre) tiene una Kd aproximada de aproximadamente 80 pM y la variante TCRA_S02_2E5 tiene una Kd de aproximadamente 90 pM.

La figura 7 muestra la obtención del perfil de especificidad de la variante de LHE TCRA_S02_2E5_RD2_8 usando análisis de escisión (paneles superiores) y afinidad (paneles inferiores). El análisis de escisión y afinidad se realizó tal como se describió en la figura 3 y 4 respectivamente, sin embargo se sometieron a prueba paneles de sustratos de ADN mediante lo cual cada una de las 22 posiciones a lo largo de la diana se alteró en serie a cada uno de los 3 pares de bases no nativos y se sometió a prueba en condiciones que inhiben la escisión y que permiten la escisión. Los perfiles de escisión y afinidad resultantes para los 67 sustratos diferentes (la diana en el gen de TRAC y los 66 sustratos "puntuales") indican que las variantes de LHE tienen las propiedades de especificidad más ideales y son, por tanto, mejores candidatos para aplicaciones que exigen una especificidad más estricta, tal como en terapéutica humana.

La figura 8 muestra las propiedades de dispersión de citometría de flujo de células T humanas primarias que son altamente susceptibles a roturas de ADN bicatenario que dan como resultado genotoxicidad y muerte celular. Células T transfectadas con especies de ARNm-IVT que codifican para una proteína inocua tal como la proteína fluorescente azul (BFP), muestran el 59% de supervivencia durante el cultivo *in vitro*, un nivel similar a células T sin manipular. La variante TCRA_S02_2E5_RD2_8 da como resultado niveles muy similares de viabilidad de células T, confirmando que esta especificidad de ADN global es de alta calidad.

La figura 9 muestra un plásmido de producción lentiviral de autoinactivación (SIN) a partir del cual se generaron preparaciones de lentivirus y se usaron como vectores para la transducción de líneas celulares y células primarias (A) (SEQ ID NO: 12 y SEQ ID NO: 13) así como un vector no limitativo a modo de ejemplo que contiene el constructo MegaTAL usado o bien en la producción lentiviral o bien para la transcripción *in vitro* para la producción de ARNm-IVT (B) (SEQ ID NO: 12). Las principales características del vector, además de las características lentivirales bien conocidas por los expertos en la técnica tales como las repeticiones terminales largas (LTR), sitio de unión al cebador (PBS), tramo de polipurina central (cPPT) y un promotor de ARN de T7 (T7), son elementos multicistronicos para la expresión de una LHE que selecciona como diana TCRA (tal como se muestra) o MegaTAL unida por medio de un motivo de ligador peptídico T2A a la exonucleasa Trex2 que porta además un sitio interno de entrada al ribosoma (IRES) y proteína fluorescente azul (BFP) para rastrear células transducidas.

La figura 10 muestra la pérdida completa de expresión de la proteína TCR-alfa a partir de la superficie celular de células T humanas primarias transfectadas con ARNm-IVT que codifica para la variante de LHE TCRA_S02_RD2_8 conjuntamente con ARNm-IVT que codifica para la proteína exonucleasa Trex2 auxiliar. Los paneles de citometría de flujo representan el análisis de diferentes puntos de tiempo tras la transfección de ARNm (panel izquierdo: 72 h, panel derecho: 14 días). La presencia de proteína TCR-alfa sobre la superficie celular se detecta por medio de un anticuerpo específico para la molécula de correceptor CD3. La expresión en la superficie del correceptor CD3 requiere la expresión funcional de la proteína TCR-alfa y la ausencia de una señal de CD3 significa la alteración satisfactoria del gen de TRAC.

La figura 11 muestra datos de secuenciación génica en el sitio diana TCRA_S02 a partir de células tratadas con la

variante de LHE TCRA_S02_RD2_8. Se cultivaron células T primarias transfectadas con ARNm-IVT que codifica para la variante de LHE TCRA_S02_RD2_8 conjuntamente con ARNm-IVT que codifica para la exonucleasa Trex2 y se purificaron usando selección por citometría de flujo de células negativas para CD3. Las células negativas para CD3 clasificadas tienen deleciones en el 50% de los alelos de TCRA, consecuente con que el alelo de TRAC no silenciado sea la diana primaria para variantes de LHE TCRA_S02.

La figura 12 representa esquemáticamente una estrategia de tratamiento que podría usarse para generar poblaciones de células T deficientes en TCRA para el tratamiento de cáncer, autoinmunidad o infección viral crónica. En resumen, se aislarían células mononucleares de sangre periférica (PBMC) mediante aféresis, se procesarían para purificar y cultivar células T, se tratarían con agentes de suministro de nucleasas que seleccionan como diana TRAC, se combinarían con receptores de antígenos artificiales (CAR) y se reinfundirían al paciente.

La figura 13 muestra una representación esquemática de la tecnología de la alteración génica de TRAC ultraeficiente basada en la combinación de la arquitectura de MegaTAL y la expresión de Trex2. También se muestra la secuencia diana de TCRA_S02 MegaTAL dentro del ADN del gen de TRAC, con la ubicación del alineamiento de TALE de 11 meros tanto indicada esquemáticamente (las unidades repetitivas no se indican sino que se muestran en el sentido de 5' de la anotación de TCRA_S02) como mostrada su secuencia en negrita.

La figura 14 demuestra la eficacia extremadamente alta de eliminación de proteína TCR-alfa de la superficie celular de células T humanas primarias transfectadas con ARNm-IVT que codifica para la variante de LHE TCRA_S02_MegaTAL conjuntamente con ARNm-IVT que codifica para la proteína exonucleasa Trex2 auxiliar. Los paneles de citometría de flujo demuestran el ensamblaje del complejo de TCR tras la electroporación con un control de proteína fluorescente (BFP) o variante de LHE TCRA_S02 MegaTAL con o sin ARNm-IVT que codifica para Trex2.

La figura 15 muestra la inactivación génica de TCRA eficaz en células T humanas primarias por medio de electroporación con una especie de ARNm que codifica para una proteína de fusión TAL-LHE-Trex2 de tres componentes.

Divulgación detallada de la invención

A menos que se definan específicamente en el presente documento, todos los términos técnicos y científicos usados tienen el mismo significado que entiende comúnmente un experto en la técnica en los campos de terapia génica, bioquímica, genética, biología molecular e inmunología.

Todos los métodos y materiales similares o equivalentes a los descritos en el presente documento pueden usarse en la práctica o las pruebas de la presente invención, describiéndose métodos y materiales adecuados en el presente documento. Todas las publicaciones, solicitudes de patente, patentes y otras referencias mencionadas en el presente documento se incorporan como referencia en su totalidad. En caso de conflicto, la presente memoria descriptiva, incluyendo las definiciones, prevalecerá. Además, los materiales, métodos y ejemplos son ilustrativos sólo y no pretenden ser limitativos, a menos que se especifique lo contrario.

La práctica de la presente invención empleará, a menos que se indique otra cosa, técnicas convencionales de biología celular, cultivo celular, biología molecular, biología transgénica, microbiología, ADN recombinante e inmunología, que están dentro del conocimiento de la técnica. Tales técnicas se explican completamente en la bibliografía. Véanse, por ejemplo, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, EE.UU.); Molecular Cloning: A Laboratory Manual, tercera edición, (Sambrook *et al*, 2001, Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis *et al*. patente estadounidense n.º 4.683.195; Nucleic Acid Hybridización (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); las series, Methods In ENZYMOLOGY (J. Abelson y M. Simon, eds.-in-chief, Academic Press, Inc., Nueva York), específicamente, los volúmenes 154 y 155 (Wu *et al*. eds.) y el volumen 185, "Gene Expresión Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller y M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer y Walquer, eds., Academic Press, Londres, 1987); Handbook Of Experimental Immunology, volúmenes I-IV (D. M. Weir y C. C. Blackwell, eds., 1986); y Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Variantes de I-Onu1 y homólogos de I-Onu1

La presente invención se refiere a endonucleasas de corte poco común que implican variantes de I-Onu1 y variantes de homólogos de I-Onu1 de I-LtrI, I-LtrWI, I-PanMI, I-PanMII, I-PanMIII, I-Gzel, I-GzeMII, I-GzeMIII, I-Gpil, I-GpeMI, I-AabMI, I-AaeMI, I-ApaMI, I-CpaMI, I-CpaMII, I-CpaMIII, I-CpaMIV, I-CpaMV, I-EjeMI, I-CkaMI, I-CraMI, I-MpeMI, I-MveMI, I-NcrMI, I-OheMI, I-OsoMI, I-OsoMII, I-OsoMIII, I-OsoMIV, I-SmaMI, I-SscMI, I-Vdi141I, I-PnoMI, I-ScuMI; (Takeuchi, Lambert *et al*. 2011) en las que se han introducido mutaciones y que pueden seleccionar como diana específicamente una secuencia de ácido nucleico presente en el gen de TRAC.

Las endonucleasas de corte poco común según la presente invención se refieren a enzimas variantes que pueden catalizar la hidrólisis (escisión) de enlaces entre ácidos nucleicos dentro de una molécula de ADN o ARN, preferiblemente una molécula de ADN. Las endonucleasas según la presente invención reconocen y escinden ácidos nucleicos en secuencias de polinucleótido específicas, denominadas adicionalmente “secuencia diana de ácido nucleico”.

Para modificar por ingeniería genética endonucleasas de corte poco común específicas para sitios diana en el gen de TRAC, los inventores construyeron bibliotecas de variantes de *I-Onul* en las que se variaron residuos de aminoácido localizados en la superficie de contacto de reconocimiento de ADN de *I-Onul* natural. Las bibliotecas se examinaron para detectar la actividad de escisión de la diana frente a cada sitio diana de TRAC predicho usando los ensayos de escisión descritos anteriormente (Jarjour, West-Foile *et al.* 2009). La especificidad de la superficie de contacto de reconocimiento de ADN de *I-Onul* se alteró, por tanto, para seleccionar como diana secuencias presentes en el gen de TRAC humano.

Por “variante(s)”, quiere decirse una proteína o un polinucleótido que codifica para la misma que no existe de manera natural en la naturaleza y que se obtiene modificando por ingeniería genética o mediante mutagénesis al azar. Pueden obtenerse, por ejemplo, variantes de *I-Onul* u homólogos de *I-Onul* según la invención mediante delección o sustitución con un aminoácido diferente de al menos un residuo en la secuencia de aminoácidos de sus secuencias silvestres. Pueden introducirse por ejemplo sustitución/sustituciones y delecciones mediante mutagénesis dirigida y/o mediante mutagénesis al azar. En los aspectos de marco de la presente invención, variantes de *I-Onul* u homólogos de *I-Onul* tienen la capacidad de seleccionar como diana el gen de TRAC, lo que significa que pueden interactuar con algunas secuencias de ADN específicas que codifican para dicho gen.

Las variantes u homólogos según la invención comprenden la superficie de contacto de reconocimiento de ADN tal como se describe en el presente documento y tal como se proporciona en la tabla 1.

Una superficie de contacto de reconocimiento de ADN se refiere a los residuos de los dominios proteicos de una endonucleasa de asentamiento o variante de la misma que interactúan con bases diana del ácido nucleico así como los residuos que están adyacentes. Para cada endonucleasa de asentamiento, la superficie de contacto de reconocimiento de ADN comprende una red extensa de contactos de cadena lateral con cadena lateral y cadena lateral con ADN, la mayoría de los cuales son necesariamente únicos para reconocer una secuencia diana de ácido nucleico particular. Por tanto, las composiciones de aminoácidos de la superficie de contacto de reconocimiento de ADN (y la correspondencia de la misma con una secuencia de ácido nucleico particular) varían significativamente y es, por tanto, la característica definitiva de cualquier endonucleasa de asentamiento natural o modificada por ingeniería genética. La superficie de contacto de reconocimiento de ADN determina la identidad de la secuencia diana de ácido nucleico y también las propiedades de afinidad y especificidad que definen la calidad de la endonucleasa de asentamiento según las exigencias de la aplicación.

La variante de *I-Onul* u homólogo de *I-Onul* según la presente invención comprende al menos 10, preferiblemente al menos 15, más preferiblemente al menos 20, incluso más preferiblemente al menos 25 sustituciones de aminoácido en al menos una posición seleccionada del grupo de posiciones que consiste en las posiciones: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 168, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240 de *I-Onul* (SEQ ID NO: 2). En realizaciones particulares, dichas sustituciones son reemplazos de los aminoácidos iniciales, en cada caso por un aminoácido seleccionado del grupo que consiste en: A, D, E, G, H, K, N, P, Q, R, S, T, Y, C, V, L, W, M y I.

Como ejemplos no limitativos, la leucina (L) en la posición 26 puede reemplazarse por/mutarse a isoleucina(I); la arginina en la posición 28 puede reemplazarse por/mutarse a ácido aspártico (D), la asparagina (N) en la posición 32 puede reemplazarse por/mutarse a arginina (R); la lisina (K) en la posición 34 puede reemplazarse por/mutarse a asparagina (N); la serina (S) en la posición 35 puede reemplazarse por/mutarse a ácido glutámico (E); la valina (V) en la posición 37 puede reemplazarse por/mutarse a asparagina (N); la glicina (G) en la posición 38 puede reemplazarse por/mutarse a arginina (R); la serina (S) en la posición 40 puede reemplazarse por/mutarse a arginina (R); el ácido glutámico (E) en la posición 42 puede reemplazarse por/mutarse a serina (S); la glicina (G) en la posición 44 puede reemplazarse por/mutarse a arginina (R) (véase la tabla 1).

La valina (V) en la posición 68 puede reemplazarse por/mutarse a lisina (K); la alanina (A) en la posición 70 puede reemplazarse por/mutarse a treonina (T); la asparagina (N) en la posición 75 puede reemplazarse por/mutarse a arginina (R); la serina (S) en la posición 78 puede reemplazarse por/mutarse a metionina (M); la lisina (K) en la posición 80 puede reemplazarse por/mutarse a arginina (R) (véase la tabla 1).

La leucina (L) en la posición 138 puede reemplazarse por/mutarse a metionina (M); la serina (S) en la posición 159 puede reemplazarse por/mutarse a prolina (P); el ácido glutámico (E) en la posición 178 puede reemplazarse por/mutarse a ácido aspártico (D); la cisteína (C) en la posición 180 puede reemplazarse por/mutarse a tirosina (Y); la fenilalanina (F) en la posición 182 puede reemplazarse por/mutarse a glicina (G); la isoleucina (I) en la posición 186 puede reemplazarse por/mutarse a lisina (K); la serina (S) en la posición 188 puede reemplazarse por/mutarse a valina (V); la serina (S) en la posición 190 puede reemplazarse por/mutarse a glicina (G); la lisina (K) en la posición 191 puede reemplazarse por/mutarse a asparagina (N); la leucina (L) en la posición 192 puede reemplazarse

5 por/mutarse a alanina (A); la glicina (G) en la posición 193 puede reemplazarse por/mutarse a lisina (K); la glutamina (Q) en la posición 195 puede reemplazarse por/mutarse a tirosina (Y); la glutamina (Q) en la posición 197 puede reemplazarse por/mutarse a glicina (G); la valina (V) en la posición 199 puede reemplazarse por/mutarse a arginina (R); la treonina (T) en la posición 203 puede reemplazarse por/mutarse a serina (S); la lisina (K) en la posición 207 puede reemplazarse por/mutarse a arginina (R) (véase la tabla 1).

La tirosina (Y) en la posición 223 puede reemplazarse por/mutarse a serina (S); la lisina (K) en la posición 225 puede reemplazarse por/mutarse a triptófano (W); el ácido aspártico (D) en la posición 236 puede reemplazarse por/mutarse a ácido glutámico (E) (véase la tabla 1).

10 Además, la variante de *I-Onul* u homólogo de *I-Onul* según la presente invención tiene una identidad de secuencia de al menos el 70%, preferiblemente al menos el 75%, más preferiblemente al menos el 80%, más preferiblemente al menos el 85%, más preferiblemente al menos el 90%, más preferiblemente al menos el 95%, más preferiblemente al menos el 97%, más preferiblemente al menos el 99% con la secuencia proteica SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10.

15 En una realización preferida la variante de *I-Onul* u homólogo de *I-Onul* según la presente invención tiene una secuencia proteica seleccionada del grupo que consiste en SEQ ID NO: 6, SEQ ID NO: 8 y SEQ ID NO: 10.

Según una realización preferida de la invención las variantes de *I-Onul* u homólogos de *I-Onul* según la invención escinden una secuencia diana que es diferente de la secuencia diana de la correspondiente endonucleasa silvestre. La escisión en la secuencia diana de ácido nucleico puede corresponder a o bien una rotura bicatenaria o bien una rotura monocatenaria.

20 La presente invención se basa en el hallazgo de que tales endonucleasas variantes con especificidades novedosas pueden usarse para permitir una modificación dirigida eficaz del gen de TRAC.

25 Los presentes inventores han identificado de hecho secuencias diana de *I-Onul* supuestas en el gen de TRAC humano basándose en una serie de características comunes intrínsecas para el grupo de miembros de la subfamilia de LHE de tipo *I-Onul* monoméricos recientemente descritos en (Takeuchi, Lambert *et al.* 2011; Baxter, Lambert *et al.* 2012). Las secuencias diana de LHE supuesta también se identifican basándose en las ubicaciones dentro del gen de TRAC en el que inserciones o deleciones mediadas por endonucleasa pueden provocar alteraciones significativas en la proteína TCR-alfa. Los presentes inventores identificaron dos secuencias diana supuestas en el gen de TRAC humano (SEQ ID NO: 3 a SEQ ID NO: 4) tras lo cual se modificó por ingeniería genética la superficie de contacto de reconocimiento de ADN de las variantes de *I-Onul*. Entre estos dos sitios diana supuestos, una secuencia (TCRA_S02) se ha seleccionado como diana satisfactoriamente mediante las variantes de *I-Onul* resultantes.

35 Por consiguiente, la presente invención se refiere a una endonucleasa de corte poco común que comprende una variante de *I-Onul* u homólogo de *I-Onul* que reconoce una secuencia de ácido nucleico diana presente dentro del gen de TRAC, preferiblemente aquellas presentes en el exón 1 del gen de TRAC, más preferiblemente una secuencia de ácido nucleico diana que comprende la secuencia de ácido nucleico SEQ ID NO: 3.

Endonucleasa química

40 En otro aspecto, la invención se refiere a una endonucleasa de corte poco común bajo la forma de una endonucleasa química que comprende una variante de *I-Onul* u homólogo de *I-Onul* tal como se describió anteriormente, opcionalmente fusionada a al menos un dominio proteico adicional mediante un ligador peptídico. El dominio proteico adicional puede seleccionarse del grupo que consiste en: un dominio de unión a ácido nucleico para permitir una especificidad superior en la secuencia de ácido nucleico diana y evitar un sitio fuera de diana; un dominio catalítico para procesar (por ejemplo polimerizar, despolimerizar, modificar) la secuencia de ácido nucleico diana; y una o más etiquetas de epítopos terminales o proteínas fluorescentes para seguir y visualizar la proteína química.

45 En una realización particular, la variante de *I-Onul* u homólogo de *I-Onul* se fusiona con un dominio de unión a ácido nucleico tal como dominio de unión a ácido nucleico TALE como ejemplo no limitativo para mejorar la selección como diana del gen de TRAC.

50 Dicho efector de tipo activador de la transcripción (TALE) corresponde a un TALE modificado por ingeniería genética que comprende una pluralidad de secuencias repetidas de TALE, comprendiendo cada repetición un RVD específico para cada base de nucleótido de un sitio de reconocimiento de TALE. En la presente invención, cada secuencia repetida de TALE de dicho TALE está constituida por de 30 a 42 aminoácidos, más preferiblemente 33 ó 34 en los que dos aminoácidos críticos (el denominado dipéptido variable repetido, RVD) ubicados en las posiciones 12 y 13 median en el reconocimiento de un nucleótido de dicha secuencia de sitio de unión a TALE; dos aminoácidos críticos equivalentes pueden ubicarse en posiciones distintas de 12 y 13 particularmente en una secuencia repetida de TALE mayor de 33 ó 34 aminoácidos de longitud. Preferiblemente, los RVD asociados con el reconocimiento de los diferentes nucleótidos son HD para reconocer C, NG para reconocer T, NI para reconocer A, NN para reconocer G o A, NS para reconocer A, C, G o T, HG para reconocer T, IG para reconocer T, NK para reconocer G, HA para

reconocer C, ND para reconocer C, HI para reconocer C, HN para reconocer G, NA para reconocer G, SN para reconocer G o A e YG para reconocer T, TL para reconocer A, VT para reconocer A o G y SW para reconocer A. En otra realización, pueden mutarse los aminoácidos críticos 12 y 13 hacia otros residuos de aminoácido con el fin de modular su especificidad hacia los nucleótidos A, T, C y G y en particular para potenciar esta especificidad. Por otros
 5 residuos de aminoácido está previsto cualquiera de los veinte residuos de aminoácido naturales o derivados de aminoácidos no naturales.

En otra realización, dicho TALE de la presente invención comprende entre 5 y 30 secuencias repetidas de TALE. Más preferiblemente, dicho TALE de la presente invención comprende entre 8 y 20 secuencias repetidas de TALE; de nuevo más preferiblemente 10 secuencias repetidas de TALE.

10 En otra realización, dicho TALE comprende una secuencia repetida de TALE truncada individual adicional compuesta por 20 aminoácidos ubicados en el extremo C-terminal de dicho conjunto de secuencias repetidas de TALE, es decir media secuencia repetida de TALE C-terminal adicional. En este caso, dicho TALE de la presente invención comprende entre 5,5 y 30,5 secuencias repetidas de TALE, refiriéndose “,5” a la media secuencia repetida de TALE mencionada anteriormente (o RVD terminal, o mitad de la repetición). Más preferiblemente, dicho TALE de
 15 la presente invención comprende entre 5,5 y 20,5 secuencias repetidas de TALE, de nuevo más preferiblemente, 10,5 secuencias repetidas de TALE. En una realización preferida, dicha media secuencia repetida de TALE está en un contexto de TALE que permite una falta de especificidad de dicha secuencia repetida de la mitad de TALE hacia los nucleótidos A, C, G, T. En una realización más preferida, dicha media secuencia repetida de TALE está ausente. En otra realización, dicho TALE de la presente invención comprende secuencias repetidas de tipo TALE de
 20 diferentes orígenes. En una realización preferida, dicho TALE comprende secuencias repetidas de tipo TALE que se originan de diferentes efectores de TAL que se producen de manera natural. En otra realización preferida, las estructuras internas de algunas secuencias repetidas de tipo TALE del TALE de la presente invención están constituidas por estructuras o secuencias originadas a partir de diferentes efectores de TAL que se producen de manera natural. En otra realización, dicho TALE de la presente invención comprende secuencias repetidas de tipo
 25 TALE. Las secuencias repetidas de tipo TALE tienen una secuencia diferente de las secuencias repetidas de TALE que se producen de manera natural pero tienen la misma función y / o estructura global dentro de dicho armazón central de la presente invención.

La endonucleasa quimérica según la invención puede corresponder, por tanto, a la fusión de una variante de *I-Onul* o variante de homólogo de *I-Onul* tal como se describió anteriormente con un dominio de unión a ácido nucleico modular, tal como un dominio de TALE o uno de dedos de zinc, siendo dicha fusión activa en forma monomérica, como parte de un polipéptido de cadena sencilla.
 30

Según un aspecto adicional de la invención, el dominio proteico fusionado a la variante de *I-Onul* o variante de homólogo de *I-Onul* puede tener al menos una actividad catalítica seleccionada del grupo que consiste en actividad nucleasa, actividad polimerasa, actividad cinasa, actividad fosfatasa, actividad metilasa, actividad topoisomerasa,
 35 actividad integrasa, actividad transposasa, actividad ligasa, actividad helicasa, actividad recombinasa. En una realización preferida, el dominio proteico tiene una actividad endonucleasa, mientras que la variante de *Onu-I* conserva su propia actividad de escisión o conserva únicamente la afinidad de unión a TRAC; en otra realización preferida, dicho dominio proteico es o comprende una actividad exonucleasa. Como ejemplos no limitativos, los dominios catalíticos pueden ser o comprender en parte una de las proteínas seleccionadas del grupo que consiste en: Mmel, colicina-E7 (CEA7_ECOLX), colicina-E9, APFL, EndA, Endo I (END1_ECOLI), Endo G humana (NUCG_HUMAN), Endo G bovina (NUCG_BOVIN), R.HinP11, I-Basl, I-Bmol, I-Hmul, I-TevI, I-TevII, I-TevIII, I-Twol, R.Mspl, R.Mval, NucA, NucM, Vvn, Vvn_CLS, nucleasa de estafilococos (NUC_STAAU), nucleasa de estafilococos (NUC_STAHY), nucleasa de micrococos (NUC_SHIFL), endonucleasa *yncB*, endodesoxirribonucleasa I (ENRN_BPT7), metnasa, Nb.BsrDI, BsrDI A, Nt.BspD6I (subunidad grande de R.BspD6I), ss.BspD6I (subunidad pequeña de R.BspD6I), R.PleI, MlyI, AlwI, Mva1269I, Bsrl, Bsml, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunidad 1, BbvCI subunidad 2, Bpu10I subunidad alfa, Bpu10I subunidad beta, Bmrl, Bfil, I-Crel, hExol (EX01_HUMAN), Exol de levaduras (EX01_YEAST), Exol de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, DNA2 humana, DNA2 de levadura (DNA2_YEAST), TdT y VP16 o un mutante funcional de las mismas.
 40
 45

50 En una realización preferida, el dominio catalítico es una enzima de procesamiento de extremos de ADN. Los ejemplos no limitativos de enzimas de procesamiento de extremos de ADN incluyen exonucleasas 5-3', exonucleasas 3-5', exonucleasas alcalinas 5-3', endonucleasas flap 5', helicadas, fosfatasa, hidrolasas y ADN polimerasas independientes de molde. En una realización más preferida, dicho dominio catalítico tiene una actividad exonucleasa, en particular una actividad exonucleasa 3'-5'. En una realización más preferida, dicho dominio catalítico es TREX2 o una variante funcional del mismo. En otra realización preferida, dicho dominio catalítico está codificado por un polipéptido de TREX2 de cadena sencilla. En una realización particular, dicho dominio catalítico se fusiona al extremo N-terminal o extremo C-terminal de dicha endonucleasa de corte poco común. En una realización más preferida, dicho dominio catalítico se fusiona a dicha endonucleasa de corte poco común mediante un ligador peptídico.
 55

60 En aspectos particulares, los ligadores peptídicos actúan como dispositivo de comunicación/elemento de unión o acoplamiento entre la endonucleasa de corte poco común y un dominio proteico adicional para actuar conjuntamente

para la actividad. Dicho ligador peptídico proporciona una secuencia peptídica que permite la conexión de diferentes monómeros en una proteína de fusión y la adopción de la conformación correcta para la actividad de dicha proteína de fusión, pero no altera la especificidad de cualquiera de los monómeros por sus dianas. Los ligadores peptídicos pueden ser de diversos tamaños, desde 2 aminoácidos hasta 50 aminoácidos como intervalo indicativo no limitativo. Los ligadores peptídicos también pueden estar estructurados o no estructurados.

Alternativamente, la variante de *I-Onul* o variante de homólogo de *I-Onul* según la invención se usa conjuntamente con otra proteína que no está fusionada a la misma, que tiene la misma actividad catalítica que el dominio proteico descrito anteriormente.

Otro aspecto de la invención proporciona polinucleótidos que comprenden una secuencia de ácido nucleico que codifica para las endonucleasas de corte poco común, preferiblemente variantes de *I-Onul*, homólogos o endonucleasa quimérica tal como se describe en el presente documento y vectores que comprenden tales polinucleótidos. El ácido nucleico o los vectores según aspectos adicionales de la presente invención pueden comprender una secuencia de ácido nucleico que codifica para uno o más motivos de localización subcelular, sitios de escisión de proteasa o secuencias de salto ribosómico.

En realizaciones particulares, los ácidos nucleicos de la presente invención pueden comprender al menos un motivo de localización subcelular. Un motivo de localización subcelular se refiere a una secuencia que facilita el transporte o el confinamiento de una proteína a una ubicación subcelular definida que incluye al menos uno del núcleo, citoplasma, membrana plasmática, retículo endoplasmático, aparato de Golgi, endosomas, peroxisomas y mitocondrias. Los motivos de localización subcelular se conocen bien en la técnica. El motivo de localización subcelular requiere una orientación específica, por ejemplo, N- y/o C-terminal con respecto a la proteína. Como ejemplo no limitativo, la señal de localización nuclear (NLS) del antígeno T grande del virus de simios virus 40 puede orientarse al extremo N y/o C-terminal. NLS es una secuencia de aminoácidos que actúa dirigiendo la proteína al núcleo celular a través del complejo de poro nuclear y encaminando una proteína recién sintetizada al núcleo por medio de su reconocimiento por receptores de transporte nuclear citosólicos. Normalmente, un NLS consiste en una o más secuencias cortas de aminoácidos cargados positivamente tales como lisinas o argininas.

Métodos de modificación por ingeniería genética del genoma

Otro aspecto de la invención se refiere al uso de una variante de *I-Onul*, variante de homólogo de *I-Onul* o endonucleasa quimérica derivada de *I-Onul* tal como se describió anteriormente para permitir la sección como diana eficaz del gen de TRAC en una célula. Más particularmente, la invención se refiere a un método para la modificación dirigida del gen de TRAC en una célula que comprende introducir en una célula la endonucleasa de corte poco común o endonucleasa quimérica tal como se describió anteriormente. En una realización particular, la presente invención se refiere a un método para modificar el gen de TRAC en una célula que comprende introducir en la célula la endonucleasa de corte poco común, más particularmente la variante de *I-Onul*, variante de homólogo de *I-Onul* o endonucleasa quimérica, de manera que la endonucleasa de corte poco común escinde una secuencia diana de ácido nucleico en el gen de TRAC.

Según una realización adicional de la invención, la endonucleasa de corte poco común se expresa en una célula con el fin de obtener mutagénesis dirigida al locus de TRAC. Las roturas de la hebra de ácido nucleico provocadas por la endonucleasa de corte poco común se reparan comúnmente a través de distintos mecanismos de recombinación homóloga o unión de extremos no homólogos (NHEJ). Sin embargo, NHEJ es un proceso de reparación imperfecto que da como resultado a menudo cambios en la secuencia de ADN en el sitio de la rotura bicatenaria. Los mecanismos implican volver a unir lo que queda de los dos extremos de ADN a través de religamiento directo (Critchlow y Jackson 1998) o por medio de la denominada unión de extremos mediada por microhomología (Ma, Kim *et al.* 2003). La reparación por medio de unión de extremos no homólogos (NHEJ) da como resultado a menudo pequeñas inserciones o deleciones y puede usarse para la creación de inactivaciones génicas específicas. Dicha modificación puede ser una sustitución, deleción o adición de al menos un nucleótido. Las células en las que se ha producido un acontecimiento de mutagénesis inducida por escisión, es decir un acontecimiento de mutagénesis consecutivo a un acontecimiento de NHEJ, pueden identificarse y/o seleccionarse mediante un método bien conocido en la técnica. Como ejemplo no limitativo, puede generarse un análisis de secuenciación profunda a partir del genoma celular seleccionado como diana alrededor del locus seleccionado como diana. Pueden detectarse, por tanto, acontecimientos de inserción/deleción (acontecimientos de mutagénesis). Como otro ejemplo no limitativo, pueden usarse ensayos basados en endonucleasa de T7 que reconoce ADN no perfectamente apareado, para cuantificar a partir de una PCR específica de locus sobre ADN genómico de las células proporcionadas, apareamientos erróneos entre hebras de ADN que han vuelto a aparearse procedentes de moléculas de ADN escindidas/no escindidas.

En una realización particular de los métodos previstos en el presente documento, la mutagénesis se aumenta introduciendo en la célula un dominio catalítico adicional. En una realización particular, la presente invención proporciona métodos mejorados para garantizar la modificación dirigida en el gen de TRAC y proporciona un método para aumentar la mutagénesis en la secuencia de ácido nucleico de TRAC diana para generar al menos una escisión de ácido nucleico y una pérdida de información genética alrededor de dicha secuencia de ácido nucleico diana impidiendo, por tanto, cualquier religamiento sin cicatrices mediante NHEJ. En una realización más preferida,

dicho dominio catalítico es una enzima de procesamiento de extremos de ADN. Los ejemplos no limitativos de enzimas de procesamiento de extremos de ADN incluyen exonucleasas 5-3', exonucleasas 3-5', exonucleasas alcalinas 5-3', endonucleasas flap 5', helicasas, fosfatasa, hidrolasas y ADN polimerasas independientes de molde. Los ejemplos no limitativos de tal dominio catalítico comprenden al menos un dominio proteico o derivado catalíticamente activo del dominio proteico seleccionado del grupo que consiste en hExol (EXO1_HUMAN), Exol de levaduras (EXO1_YEAST), Exol de *E. coli*, TREX2 humana, TREX1 de ratón, TREX1 humana, TREX1 bovina, TREX1 de rata, TdT (desoxinucleotidil transferasa terminal), DNA2 humana, DNA2 de levaduras (DNA2_YEAST). En una realización más preferida, dicho dominio catalítico tiene una actividad exonucleasa, en particular una actividad exonucleasa 3'-5'. En una realización más preferida, dicho dominio catalítico es TREX2 o variante funcional del mismo. En otra realización preferida, dicho dominio catalítico está codificado por un polipéptido de TREX de cadena sencilla. En una realización particular, dicho dominio catalítico se fusiona al extremo N-terminal o extremo C-terminal de dicha endonucleasa de corte poco común. Se ha encontrado que el acoplamiento de la enzima TREX2 o TREX2 de cadena sencilla con una endonucleasa tal como una meganucleasa garantiza una alta frecuencia de mutagénesis dirigida. Alternativamente, el dominio catalítico anterior puede introducirse por separado en la célula como parte de una proteína independiente.

Se sabe que las roturas endonucleocíticas estimulan la recombinación homóloga. Por tanto, en realizaciones particulares, la presente invención también se refiere a un método para inducir selección como diana de genes homólogos en la secuencia de ácido nucleico diana que comprende además introducir en la célula una matriz donadora que comprende una secuencia homóloga a al menos una porción del gen de TRAC diana, de manera que se produce recombinación homóloga entre la secuencia de ácido nucleico diana y la matriz donadora.

En realizaciones particulares, se logra la selección como diana de genes de TRAC homólogos introduciendo en una célula una endonucleasa de corte poco común tal como se describió anteriormente, para inducir una escisión dentro de o adyacente a una secuencia diana de ácido nucleico, así como una matriz donadora que comprende un transgén para introducir dicho transgén mediante recombinación homóloga. Tras la escisión de la secuencia de ácido nucleico diana, se estimula un acontecimiento de recombinación homóloga entre el genoma que contiene la secuencia de ácido nucleico diana y la matriz donadora. Dicha matriz donadora comprende una secuencia homóloga a al menos una porción de la secuencia de ácido nucleico diana, de manera que se produce recombinación homóloga entre la secuencia de ácido nucleico diana y la matriz donadora. Preferiblemente, se usan secuencias homólogas de al menos 50 pb de longitud, preferiblemente más de 100 pb y más preferiblemente más de 200 pb dentro de dicha matriz donadora. Por tanto, la matriz donadora tiene preferiblemente desde 200 pb hasta 6000 pb de longitud, más preferiblemente desde 1000 pb hasta 2000 pb. En otra realización, dicha matriz donadora comprende dos secuencias homólogas a porciones o porciones adyacentes de dicha secuencia de ácido nucleico diana que flanquean a una secuencia para introducir en la secuencia de ácido nucleico diana. De hecho, están ubicadas homólogas de ADN compartidas en regiones flanqueantes en el sentido de 5' y en el sentido de 3' del sitio de la rotura y la secuencia de ácido nucleico que va a introducirse debe ubicarse entre los dos brazos de homología. En realizaciones particulares, dicha matriz donadora comprende porciones primera y segunda que son homólogas a la región 5' y 3' del ácido nucleico diana, respectivamente. Dicha matriz donadora en estas realizaciones también puede comprender una tercera porción situada entre la porción primera y segunda que comprende poca o ninguna homología con las regiones 5' y 3' del sitio de escisión de ADN. En este caso, dicha matriz donadora permite introducir nuevo material genético en una célula. Dicho nuevo material genético introducido en una célula puede conferir una ventaja selectiva o comercial a dicha célula. En otra realización, dicha matriz donadora permite reemplazar material genético en una célula. En otra realización, dicha matriz donadora permite reparar material genético en una célula.

En realizaciones particulares, dicha matriz donadora puede comprender un marcador de selección positivo entre los dos brazos de homología y en última instancia un marcador de selección negativo en el sentido de 5' del primer brazo de homología o en el sentido de 3' del segundo brazo de homología. El/los marcador(es) permite(n) la selección de células que tienen insertada la secuencia de interés mediante recombinación homóloga en el sitio diana. Dependiendo de la ubicación de la secuencia del genoma seleccionada como diana en la que se ha producido el acontecimiento de escisión, tal matriz donadora puede usarse para inactivar un gen, por ejemplo cuando la matriz donadora está ubicada dentro del marco de lectura abierto de dicho gen, o para introducir nuevas secuencias o genes de interés. Pueden usarse inserciones de secuencias usando tal matriz donadora para modificar un gen existente seleccionado como diana, mediante corrección o reemplazo de dicho gen (intercambio de alelos como ejemplo no limitativo), o para regular por incremento o disminución la expresión del gen seleccionado como diana (intercambio de promotores como ejemplo no limitativo), dicho reemplazo o corrección del gen seleccionado como diana.

Las células en las que se ha producido un acontecimiento de recombinación homóloga pueden seleccionarse mediante métodos bien conocidos en la técnica. Como ejemplo no limitativo, puede realizarse análisis de PCR usando un oligonucleótido que se aparee dentro de la secuencia de ácido nucleico exógena y un oligonucleótido que se aparee con el ácido nucleico genómico de células fuera de dicho ácido nucleico exógeno pero cerca del locus seleccionado como diana. Por tanto, pueden seleccionarse células en las que los métodos de la invención permitieron que se produjera un acontecimiento de mutagénesis o un acontecimiento de recombinación homóloga.

Los diferentes métodos de la invención implican introducir una endonucleasa de corte poco común o endonucleasa

quimérica opcionalmente con una enzima de procesamiento de extremos de ADN o matriz donadora en una célula. Como ejemplo no limitativo, dicha endonucleasa de corte poco común o endonucleasa quimérica opcionalmente con una enzima de procesamiento de extremos de ADN o matriz donadora puede introducirse como transgenes codificados por uno vector o como diferentes vectores plasmídicos. Pueden incluirse diferentes transgenes en un vector que comprende una secuencia de ácido nucleico que codifica para una secuencia de salto ribosómico tal como una secuencia que codifica para un péptido 2A. Los péptidos 2A, que se identificaron en el subgrupo de Aphthovirus de picornavirus, provocan un “salto” ribosómico de un codón al siguiente sin la formación de un enlace peptídico entre los dos aminoácidos codificados por los codones (véase Donnelly *et al.*, J. of General Virology 82: 1013-1025 (2001); Donnelly *et al.*, J. of Gen. Virology 78: 13-21 (1997); Doronina *et al.*, Mol. Y. Cell. Biology 28(13): 4227-4239 (2008); Atkins *et al.*, RNA 13: 803-810 (2007)). Por “codón” quiere decirse tres nucleótidos en un ARNm (o en la hebra sentido de una molécula de ADN) que se traducen mediante un ribosoma en un residuo de aminoácido. Por tanto, pueden sintetizarse dos polipéptidos a partir de un único marco de lectura abierto contiguo dentro de un ARNm cuando los polipéptidos están separados por una secuencia de oligopéptido 2A que está en marco. Tales mecanismos de salto ribosómico se conocen bien en la técnica y se sabe que los usan varios vectores para la expresión de varias proteínas codificadas por un único ARN mensajero. Como ejemplo no limitativo, en la presente invención, se han usado péptidos 2A para expresar en la célula la endonucleasa de corte poco común y una enzima de procesamiento de extremos de ADN. Como ejemplos no limitativos, puede usarse el péptido 2A para expresar en la célula la endonucleasa de corte poco común o la endonucleasa quimérica y un dominio proteico adicional con una actividad catalítica seleccionada del grupo que consiste en actividad nucleasa, actividad polimerasa, actividad cinasa, actividad fosfatasa, actividad metilasa, actividad topoisomerasa, actividad integrasa, actividad transposasa, actividad ligasa, actividad helicasa, actividad recombinasa para procesar una secuencia de ácido nucleico diana. El péptido 2A puede usarse también para expresar en la célula la endonucleasa de corte poco común o la endonucleasa quimérica y una proteína fluorescente.

Dicho vector de plásmido puede contener un marcador de selección que proporciona la identificación y/o selección de células que recibieron dicho vector. Pueden introducirse vectores en una célula mediante una variedad de métodos (por ejemplo, inyección, captación directa, bombardeo con microproyectiles, liposomas, electroporación). Las endonucleasas de corte poco común, endonucleasas quiméricas, enzima de procesamiento de extremos de ADN o matriz donadora según la presente invención pueden expresarse de manera estable o transitoria en células usando vectores de expresión. Los expertos en la técnica conocen bien técnicas de expresión en células eucariotas. (Véase Current Protocols in Human Genetics: capítulo 12 “Vectors For Gene Therapy” y capítulo 13 “Delivery Systems for Gene Therapy”). El polipéptido puede sintetizarse *in situ* en la célula como resultado de la introducción de polinucleótido que codifica para polipéptido dentro de la célula. Dicha expresión proteica puede inducirse en células seleccionadas y dicha endonucleasa de corte poco común o endonucleasa quimérica escinde una secuencia de ácido nucleico diana en células seleccionadas. Alternativamente, el polipéptido podría producirse fuera de la célula y luego introducirse en la misma mediante un método bien conocido de la técnica.

En otro aspecto, la presente invención se refiere a una célula aislada que comprende un gen que codifica para la proteína TCR-alfa inactivada (por ejemplo, con respecto a la biogénesis de proteína TCR-alfa típica y/o expresión en superficie celular de proteína TCR-alfa y/o con respecto a la proteína TCR-alfa que media en el reconocimiento de antígenos y señalización del inmunorreceptor) mediante los métodos descritos anteriormente.

“Célula” o “células” tal como se usa en el presente documento se refiere a cualquier célula viva procarionta o eucariota, líneas celulares derivadas de estos organismos para cultivos *in vitro*, células primarias de origen animal.

“Célula primaria” o “células primarias” tal como se usa en el presente documento se refiere a células tomadas directamente de tejido vivo (es decir, material de biopsia) y establecidas para crecimiento *in vitro*, que han experimentado muy pocas duplicaciones de la población y son, por tanto, más representativas de los componentes funcionales principales y las características de tejidos de los que se derivan, en comparación con líneas celulares tumorigénicas continuas o inmortalizadas artificialmente. Estas células representan, por tanto, un modelo más valioso para el estado *in vivo* al que se refieren.

Más preferiblemente la célula animal es del género *Homo*, *Rattus*, *Mus*, *Sus*, *Bos*, *Danio*, *Canis*, *Felis*, *Equus*, *Salmo*, *Oncorhynchus*, *Gallus*, *Meleagris*, *Drosophila*, *Caenorhabditis*; más preferiblemente, la célula animal es de la especie *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Sus scrofa*, *Bos taurus*, *Danio rerio*, *Canis lupus*, *Felis catus*, *Equus caballus*, *Salmo salar*, *Oncorhynchus mykiss*, *Gallus gallus*, *Meleagris gallopavo*, *Drosophila melanogaster*, *Caenorhabditis elegans*.

En aspectos de la presente invención, la célula puede ser una célula de mamífero, o una línea celular derivada de estos organismos para cultivos *in vitro* o células primarias tomadas directamente de tejido vivo y establecidas para cultivo *in vitro*. Como ejemplos no limitativos pueden seleccionarse líneas celulares del grupo que consiste en células CHO-K1; células HEK293; células Caco2; células U2-OS; células NIH 3T3; células NSO; células SP2; células CHO-S; células DG44; células K-562, células U-937; células MRC5; células IMR90; células Jurkat; células HepG2; células HeLa; células HT-1080; células HCT-116; células Hu-h7; células Huvec; células Molt 4.

En una realización más preferida, dichas células aisladas pueden ser células multipotentes, por ejemplo células madre. Las células madre pueden ser células madre adultas, células madre embrionarias, más particularmente

células madre no humanas, células madre de sangre de cordón umbilical, células progenitoras, células madre de médula ósea, células madre pluripotentes inducidas, células madre totipotentes o células madre hematopoyéticas. Células humanas representativas son células CD34+. En una realización particular de la presente invención, las células son células T, preferiblemente células T humanas.

5 Método para tratar o prevenir cáncer, enfermedad autoinmunitaria o infección viral

Las variantes de *I-Onul*, variante de homólogo de *I-Onul* o endonucleasa quimérica derivada de *I-Onul* según la invención pueden ser útiles como medicamento.

Se describe, por tanto, un método para tratar a un sujeto que tiene cáncer, enfermedad autoinmunitaria o infección viral que comprende introducir en una célula una endonucleasa de corte poco común o endonucleasa quimérica según la invención suficiente para proporcionar mutagénesis o recombinación homóloga en el gen de TRAC, opcionalmente con una matriz donadora y/o enzima de procesamiento de extremos de ADN y administrar las células al sujeto. En aspectos particulares, el método puede combinar la introducción de una endonucleasa de corte poco común o endonucleasa quimérica con la introducción de un receptor de antígeno artificial/quimérico que reconoce un tumor, virus o diana relacionada con autoinmunidad. En una determinada realización, el método puede comprender seleccionar células cultivadas en las que el acontecimiento de recombinación homóloga o mutagénesis se ha producido en el gen de TRAC mediante métodos bien conocidos en la técnica.

Dicho tratamiento puede ser de mejora, curativo o profiláctico. Puede ser o bien parte de un tratamiento autólogo o bien parte de un tratamiento alogénico. Por autólogo, quiere decirse que las células, línea celular o población de células usadas para tratar a los pacientes se originan a partir de dicho paciente. Por alogénico quiere decirse que las células o población de células usadas para tratar a los pacientes no se originan a partir de dicho paciente sino a partir de un donante.

Las células que pueden usarse con los métodos dados a conocer pueden ser células multipotentes, por ejemplo células madre. Las células madre pueden ser células madre adultas, células madre embrionarias, más particularmente células madre no humanas, células madre de sangre del cordón umbilical, células progenitoras, células madre de médula ósea, células madre pluripotentes inducidas, células madre totipotentes o células madre hematopoyéticas. Células humanas representativas son células CD34+ o células T humanas. Antes de la expansión y modificación genética de las células de la invención, puede obtenerse una fuente de células de un sujeto a través de una variedad de métodos no limitativos. Pueden obtenerse células T a partir de varias fuentes no limitativas, incluyendo células mononucleares de sangre periférica, médula ósea, tejido de ganglios linfáticos, sangre de cordón umbilical, tejido del timo, tejido de un sitio de infección, ascitis, efusión pleural, tejido de bazo y tumores. En determinadas realizaciones de la presente invención, puede usarse cualquier número de líneas de células T disponibles y conocidas por los expertos en la técnica.

En otra realización, células aisladas obtenidas mediante los diferentes métodos o línea(s) celular(es) derivadas de dichas células aisladas pueden ser útiles como medicamento. En otra realización, dicho medicamento puede ser útil en el tratamiento de cáncer o enfermedad autoinmunitaria o infección viral en un paciente que lo necesita. En otra realización, dicha célula aislada según la invención o línea celular derivada de dicha célula aislada puede usarse en la fabricación de un medicamento para el tratamiento de un cáncer, enfermedad autoinmunitaria o infección viral en un paciente que lo necesita.

La administración de las células o población de células según la presente invención puede llevarse a cabo de cualquier manera conveniente, incluyendo mediante inhalación de aerosol, inyección, ingestión, transfusión, implantación o trasplante. Las composiciones descritas en el presente documento pueden administrarse a un paciente por vía subcutánea, por vía intradérmica, por vía intratumoral, por vía intraganglionar, por vía intramedular, por vía intramuscular, mediante inyección intravenosa o intralinfática, o por vía intraperitoneal. En una realización, las composiciones de células de la presente invención se administran preferiblemente mediante inyección intravenosa.

En aspectos particulares, la administración de las células o población de células comprende la administración de 10^4 a 10^9 células/kg de peso corporal, preferiblemente de 10^5 a 10^6 células/kg de peso corporal, incluyendo todos los valores números de células dentro de esos intervalos. Las células o población de células pueden administrarse en una o más dosis. En otra realización, dicha cantidad eficaz de células se administran como una única dosis. En otra realización, dicha cantidad eficaz de células se administran como más de una dosis a lo largo de un periodo de tiempo. El momento de administración está dentro del criterio del médico encargado y depende del estado clínico del paciente. Las células o población de células pueden obtenerse de cualquier fuente, tal como un banco de células o un donante. Aunque las necesidades individuales varían, la determinación de intervalos óptimos de cantidades eficaces de un tipo de células dado para una enfermedad particular o condiciones dentro del conocimiento de la técnica. La dosificación administrada dependerá de la edad, salud y peso del receptor, la clase de tratamiento simultáneo, si lo hay, la frecuencia de tratamiento y la naturaleza del efecto deseado.

En otra realización, la administración de células puede combinarse con la administración de un régimen de fármaco inmunosupresor. El régimen inmunosupresor puede incluir, pero sin limitarse a, clases de fármacos citostáticos, de

glucocorticoides y basados en anticuerpos. El régimen inmunosupresor puede administrarse antes, durante o después de la administración de células. El régimen inmunosupresor puede administrarse en una o más dosis. La dosificación, el momento y la composición del régimen de fármaco inmunosupresor están dentro del criterio del médico encargado y dependen del estado clínico del paciente.

- 5 También se describe un método para seleccionar como diana el gen de TCR-alfa en un sujeto, comprendiendo el método administrar a un sujeto un vector que codifica para una endonucleasa de corte poco común según la presente invención.

Definiciones

10 En la descripción anterior, varios términos se usan de manera extensa. Se proporcionan las siguientes definiciones para facilitar la comprensión de las presentes realizaciones.

Tal como se usa en el presente documento, el término “aproximadamente” indica que un valor incluye la variación de error inherente del método que está empleándose para determinar un valor, o la variación que existe entre experimentos.

15 - En el presente documento, se designan residuos de aminoácido en una secuencia de polipéptido según el código de una letra, en el que, por ejemplo, Q significa residuo de Gln o glutamina, R significa residuo de Arg o arginina y D significa residuo de Asp o ácido aspártico.

- Sustitución de aminoácido significa el reemplazo de un residuo de aminoácido por otro, por ejemplo el reemplazo de un residuo de arginina por un residuo de glutamina en una secuencia peptídica es una sustitución de aminoácido.

20 - Los nucleótidos se designan tal como sigue: se usa el código de una letra para designar la base de un nucleósido: a es adenina, t es timina, c es citosina y g es guanina. Para los nucleótidos degenerados, r representa g o a (nucleótidos de purina), k representa g o t, s representa g o c, w representa a o t, m representa a o c, y representa t o c (nucleótidos de pirimidina), d representa g, a o t, v representa g, a o c, b representa g, t o c, h representa a, t o c, y n representa g, a, t o c.

25 - Tal como se usa en el presente documento, “ácido nucleico” o “molécula de ácido nucleico” se refiere a nucleótidos y/o polinucleótidos, tales como ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), oligonucleótidos, fragmentos generados mediante la reacción en cadena de la polimerasa (PCR) y fragmentos generados mediante cualquiera de ligación, escisión, acción de endonucleasas y acción de exonucleasas. Las moléculas de ácidos nucleicos pueden componerse de monómeros que son nucleótidos que se producen de manera natural (tales como ADN y ARN), o análogos de nucleótidos que se producen de manera natural (por ejemplo, formas enantioméricas de nucleótidos que se producen de manera natural), o una combinación de ambos. Los nucleótidos modificados pueden tener alteraciones en restos de azúcar y/o en restos de bases de pirimidina o purina. Las modificaciones de azúcar incluyen, por ejemplo, el reemplazo de uno o más grupos hidroxilo por halógenos, grupos alquilo, aminas y grupos azido, o pueden funcionalizarse azúcares como éteres o ésteres. Además, todo el resto de azúcar puede reemplazarse por estructuras estérica y electrónicamente similares, tales como aza-azúcares y análogos de azúcares carbocíclicos. Los ejemplos de modificaciones en un resto de base incluyen purinas y pirimidinas alquiladas, purinas y pirimidinas aciladas, u otros sustitutos heterocíclicos bien conocidos. Pueden unirse monómeros de ácido nucleico mediante enlaces fosfodiéster o análogos de tales uniones. Los ácidos nucleicos pueden ser o bien mono o bien bicatenarios.

40 - Por “endonucleasa quimérica” quiere decirse una endonucleasa que comprende porciones funcionales de una endonucleasa unidas operativamente a uno o más dominios funcionales de proteína procedentes de otra proteína.

45 - Los términos “proteína de fusión” o “proteína quimérica” indican que la proteína incluye componentes polipeptídicos derivados de más de una proteína o polipéptido original. Normalmente, una proteína de fusión se expresa a partir de un gen de fusión en el que una secuencia de nucleótidos que codifica para una secuencia de polipéptido de una proteína se añade en marco con, y opcionalmente separada mediante un ligador de, una secuencia de nucleótidos que codifica para una secuencia de polipéptido de una proteína diferente. El gen de fusión puede expresarse entonces mediante una célula huésped como una única proteína. Una proteína de fusión puede comprender al menos parte de un polipéptido fusionado a otro polipéptido. En algunas realizaciones, una proteína de fusión puede comprender al menos una parte de un polipéptido fusionado con al menos una parte del mismo polipéptido.

50 - Por “examen” quiere decirse la sección secuencial o simultánea de una o más variantes de meganucleasa que presentan un fenotipo especificado tal como actividad de escisión alterada.

- Por “mutación” está previsto la sustitución, delección, inserción de uno o más nucleótidos/aminoácidos en una secuencia de polinucleótido (ADNc, gen) o una secuencia de polipéptido. Dicha mutación puede afectar a la secuencia codificante de un gen o su secuencia reguladora. También puede afectar a la estructura de la secuencia genómica o la estructura/estabilidad del ARNm codificado.

55 - Por “gen” quiere decirse la unidad básica de la herencia, que consiste en un segmento de ADN dispuesto de una

manera lineal a lo largo de un cromosoma, que codifica para una proteína específica o segmento de proteína. Un gen incluye normalmente un promotor, una región no traducida en 5', una o más secuencias codificantes (exones), opcionalmente intrones, una región no traducida en 3'. El gen puede comprender además un terminador, potenciadores y/o silenciadores.

5 - Tal como se usa en el presente documento, el término "transgen" se refiere a una secuencia que codifica para un polipéptido. Preferiblemente, el polipéptido codificado por el transgén o bien no se expresa o bien se expresa pero no es biológicamente activo en la célula, tejido o individuo en el que se inserta el transgén. Lo más preferiblemente, el transgén codifica para un polipéptido terapéutico útil para el tratamiento de un individuo.

10 - Por "vector de suministro" o "vectores de suministro" está previsto cualquier vector de suministro que puede usarse en la presente invención para poner la célula en contacto (es decir, "puesta en contacto") o suministrar dentro de células o compartimentos subcelulares agentes/productos químicos y moléculas (proteínas o ácidos nucleicos) necesarios en la presente invención. Incluye, pero no se limita a vectores de suministro liposomales, vectores de suministro virales, vectores de suministro farmacológicos, portadores químicos, portadores poliméricos, lipoplejos, poliplejos, dendrímeros, microburbujas (agentes de contraste de ultrasonidos), nanopartículas, emulsiones u otros
15 vectores de transferencia apropiados. Estos vectores de suministro permiten el suministro de moléculas, productos químicos, macromoléculas (genes, proteínas), u otros vectores tales como plásmidos, péptidos desarrollados por Diatos. En estos casos, los vectores de suministro son portadores de moléculas. Por "vector de suministro" o "vectores de suministro" también están previstos métodos de suministro para realizar transfección.

20 - Los términos "vector" o "vectores" se refieren a una molécula de ácido nucleico que puede transportar otro ácido nucleico al que se ha unido. Un "vector" en la presente invención incluye, pero no se limita a, un vector viral, un plásmido, un vector de ARN o una molécula de ADN o ARN lineal o circular que puede consistir en un ácido nucleico cromosómico, no cromosómico, semisintético o sintético. Vectores preferidos son los que pueden dar replicación autónoma (vector episomal) y/o expresión de ácidos nucleicos a los que están unidos (vectores de expresión). Los expertos en la técnica conocen grandes números de vectores adecuados y están disponibles comercialmente.

25 Los vectores virales incluyen retrovirus, adenovirus, parvovirus (por ejemplo virus adenoasociados), coronavirus, virus de ARN de cadena negativa tales como ortomixovirus (por ejemplo, virus de la gripe), rhabdovirus (por ejemplo, virus de la rabia y de la estomatitis vesicular), paramixovirus (por ejemplo, sarampión y Sendai), virus de ARN de cadena positiva tales como picornavirus y alfavirus, y virus de ADN bicatenario incluyendo adenovirus, herpesvirus (por ejemplo, virus del herpes simple tipos 1 y 2, virus de Epstein-Barr, citomegalovirus) y virus de la viruela (por
30 ejemplo, vaccinia, viruela aviar y viruela del canario). Otros virus incluyen virus Norwalk, togavirus, flavivirus, reovirus, papovavirus, hepadnavirus y virus de la hepatitis, por ejemplo. Los ejemplos de retrovirus incluyen: leucosis aviar-sarcoma, virus de tipo C, tipo B, tipo D de mamíferos, grupo de VLTH-VLB, lentivirus, espumavirus (Coffin, J. M., *Retroviridae: The viruses and their replication*, en *Fundamental Virology*, tercera edición, B. N. Campos, *et al.*, Eds., Lippincott-Raven Publishers, Filadelfia, 1996).

35 - Por "vector lentiviral" quiere decirse vectores lentivirales basados en VIH que son muy prometedores para suministro génico debido a su capacidad de empaquetamiento relativamente grande, inmunogenicidad reducida y su capacidad para transducir de manera estable con alta eficacia una gama grande de diferentes tipos de células. Se generan habitualmente vectores lentivirales tras la transfección transitoria de tres o más plásmidos (empaquetamiento, envuelta y transferencia) dentro de células productoras. Como VIH, los vectores lentivirales
40 entran en la célula diana a través de la interacción de glicoproteínas de la superficie viral con receptores en la superficie celular. Al entrar, el ARN viral experimenta transcripción inversa, que está mediada por el complejo de transcriptasa inversa viral. El producto de la transcripción inversa es un ADN viral lineal bicatenario, que es el sustrato para la integración viral en el ADN de células infectadas.

45 - Por "vectores lentivirales integrativos (o VL)", quiere decirse tales vectores como ejemplo no limitativo, que pueden integrarse en el genoma de una célula diana.

En contraposición, por "vectores lentivirales no integrativos" (o VLNI) quiere decirse vectores de suministro génico eficientes que no se integran en el genoma de una célula diana a través de la acción de la integrasa del virus.

50 Un tipo de vector preferido es un episoma, es decir, un ácido nucleico que puede producir replicación extracromosómica. Vectores preferidos son los que pueden producir replicación autónoma y/o expresión de ácidos nucleicos a los que están unidos. Vectores que pueden dirigir la expresión de genes a los que están operativamente unidos se denominan en el presente documento "vectores de expresión". Un vector según la presente invención comprende, pero no se limita a, un YAC (cromosoma artificial de levaduras), un BAC (artificial bacteriano), un vector de baculovirus, un fago, un fagémido, un cósmido, un vector viral, un plásmido, un vector de ARN o una molécula de
55 ADN o ARN lineal o circular que puede consistir en ADN cromosómico, no cromosómico, semisintético o sintético. En general, los vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de "plásmidos" que se refieren generalmente a bucles de ADN bicatenario circular que, en su forma de vector no están unidos al cromosoma. Los expertos en la técnica conocen grandes números de vectores adecuados. Los vectores pueden comprender marcadores seleccionables, por ejemplo: neomicina fosfotransferasa, histidinol deshidrogenasa, dihidrofolato reductasa, higromicina fosfotransferasa, timidina cinasa del virus del herpes simple, adenosina

desaminasa, glutamina sintetasa e hipoxantina-guanina fosforibosil transferasa para el cultivo de células eucariotas; TRP1 para *S. cerevisiae*; resistencia a tetraciclina, rifampicina o ampicilina en *E. coli*. Preferiblemente dichos vectores son vectores de expresión, en los que una secuencia que codifica para un polipéptido de interés se coloca bajo el control de elementos de control traduccional y transcripcional apropiados para permitir la producción o síntesis de dicho polipéptido. Por tanto, dicho polinucleótido está comprendido en un casete de expresión. Más particularmente, el vector comprende un origen de replicación, un promotor operativamente unido a dicho polinucleótido codificante, un sitio de unión al ribosoma, un sitio de corte y empalme del ARN (cuando se usa ADN genómico), un sitio de poliadenilación y un sitio de terminación de la transcripción. También puede comprender elementos potenciadores o silenciadores. La selección del promotor dependerá de la célula en la que se expresa el polipéptido. Los promotores adecuados incluyen promotores inducibles y/o específicos de tejido. Ejemplos de promotores inducibles son: promotor de metalotionina eucariota que se induce mediante niveles aumentados de metales pesados, promotor lacZ procarriota que se induce en respuesta a isopropil- β -D-tiogalacto-piranosido (IPTG) y promotor de choque térmico eucariota que se induce mediante temperatura aumentada. Ejemplos de promotores específicos de tejido son creatinina cinasa de músculo esquelético, antígeno específico de próstata (PSA), α -antitripsina proteasa, proteínas A y B tensioactivas humanas (SP), β -caseína y genes de proteínas de suero lácteo ácidas.

Los vectores de suministro y vectores pueden asociarse o combinarse con cualquier técnica de permeabilización celular tal como sonoporación o electroporación o derivados de estas técnicas.

- El término "endonucleasa" o "nucleasa" se refiere a cualquier enzima variante o silvestre que puede catalizar la hidrólisis (escisión) de enlaces entre ácidos nucleicos dentro de una molécula de ADN o ARN, preferiblemente una molécula de ADN. Las endonucleasas pueden clasificarse como endonucleasas de corte poco común cuando tienen normalmente un reconocimiento de polinucleótido mayor de 12 pares de bases (pb) de longitud, más preferiblemente de 14-45 pb. Las endonucleasas de corte poco común aumentan significativamente HR induciendo roturas bicatenarias de ADN (DSB) en un locus definido (Perrin, Buckle *et al.* 1993; Rouet, Smih *et al.* 1994; Rouet, Smih *et al.* 1994; Choulika, Perrin *et al.* 1995; Pingoud y Silva 2007). Las endonucleasas de corte poco común pueden, por ejemplo, ser una endonucleasa de asentamiento (Paques y Duchateau 2007), una nucleasa de dedos de zinc quimérica (ZFN) (Eisenschmidt, Lanio *et al.* 2005; Arimondo, Thomas *et al.* 2006; Simon, Cannata *et al.* 2008), una nucleasa TALE o una endonucleasa química. En endonucleasas químicas, se conjuga un agente de escisión peptídico o químico o bien a un polímero de ácidos nucleicos o bien a otro ADN que reconoce una secuencia diana específica, dirigiendo de ese modo la actividad de escisión a una secuencia específica. Las endonucleasas químicas también abarcan nucleasas sintéticas como conjugados de ortofenantrolina, una molécula de escisión de ADN y oligonucleótidos que forman tríplex (TFO), que se sabe que se unen a secuencias de ADN específicas (Kalish y Glazer 2005). Tales endonucleasas químicas están comprendidas en el término "endonucleasa" según la presente invención.

- El efector de tipo activador de la transcripción (TALE) es una familia de proteínas usadas en el proceso de infección por patógenos vegetales del género *Xanthomonas* (Boch, Scholze *et al.* 2009; Moscou y Bogdanove 2009; Christian, Cermak *et al.* 2010; Li, Huang *et al.* 2011; Li, Huang *et al.* 2011). El término "nucleasa efectora de TAL" (nucleasa TALE) se refiere a una nucleasa que comprende un dominio efector TAL fusionado a un dominio nucleasa. Estos dominios de unión a ADN pueden modificarse por ingeniería genética para unirse a una diana deseada y fusionarse a un dominio nucleasa, tal como el dominio nucleasa Fok1, para derivar una proteína de fusión de dominio efector TAL-nucleasa.

- El término "nucleasa de dedos de zinc" (ZFN) se refiere a enzimas de restricción artificiales generadas fusionando un dominio de unión a ADN de dedos de zinc a un dominio de escisión de ADN. Pueden modificarse por ingeniería genética dominios de dedos de zinc para que se unan a un sitio diana deseado. En algunas realizaciones, el dominio de escisión comprende el dominio de escisión no específico de FokI (Porteus y Carroll 2005). En otras realizaciones, el dominio de escisión comprende toda o una porción activa de otra nucleasa.

- Por "dominio catalítico" está previsto el dominio proteico o módulo de una enzima que contiene el sitio activo de dicha enzima; por sitio activo está previsto la parte de dicha enzima en la que se produce la catálisis del sustrato. Las enzimas, pero también sus dominios catalíticos, se clasifican y se nombran según la reacción que catalizan. El número de la Comisión de Enzimas (número de EC) es un esquema de clasificación numérico para enzimas, basado en las reacciones químicas que catalizan.

- El término "exonucleasa" se refiere a enzimas que escinden enlaces fosfodiéster en el extremo de una cadena de polinucleótido por medio de una reacción de hidrolización que rompe los enlaces fosfodiéster en el extremo o bien 3' o bien 5'. El polinucleótido puede ser ADN bicatenario (ADNbc), ADN monocatenario (ADNmc), ARN, híbridos bicatenarios de ADN y ARN, y ADN sintético (por ejemplo, que contiene bases distintas de A, C, G y T). El término "exonucleasa 5'" se refiere a exonucleasas que escinden el enlace fosfodiéster en el extremo 5'. El término "exonucleasa 3'" se refiere a exonucleasas que escinden el enlace fosfodiéster en el extremo 3'. Las exonucleasas pueden escindir los enlaces fosfodiéster en el extremo de una cadena de polinucleótido en sitios de corte de endonucleasa o en los extremos generados por otros medios químicos o mecánicos, tales como cizalladura, radiación ionizante, radiación ultravioleta, radicales de oxígeno, hidrólisis química y agentes quimioterápicos. Las exonucleasas pueden escindir los enlaces fosfodiéster en extremos romos o extremos cohesivos. Exonucleasa I y

exonucleasa III de *E. coli* son dos exonucleasas 3' comúnmente usadas que tienen actividad de degradación de cadenas sencillas exonucleolítica 3'. Otros ejemplos de exonucleasas 3' incluyen nucleósido difosfato cinasas (NDK), NDK1 (NM23-H1), NDK5, NDK7 y NDK8, WRN y exonucleasa 2 de reparación tres prima (Trex2). Exonucleasa VII de *E. coli* y gen 6 de exonucleasa de T7 son dos exonucleasas 5'-3' comúnmente usadas que

5 tienen actividad de degradación de cadenas sencillas exonucleolítica del 5%. La exonucleasa puede originarse a partir de procariotas, tales como exonucleasas de *E. coli*, o eucariotas, tales como exonucleasas de levaduras, gusanos, murinas o humanas.

- Por "mutante funcional" está previsto un mutante catalíticamente activo de una proteína o un dominio proteico; tal mutante puede tener la misma actividad en comparación con su proteína o dominio proteico original o propiedades adicionales. Esta definición se aplica a proteínas quiméricas o dominios proteicos que constituyen proteínas quiméricas según la presente invención. También se abarcan en el alcance de esta definición "derivados" de estas proteínas o dominios proteicos que comprenden la totalidad o parte de estas proteínas o dominios proteicos fusionados a otra proteína o partes químicas tales como etiquetas, anticuerpos, polietilenglicol como ejemplos no limitativos.

15 - Por "secuencia homóloga" de ácido nucleico o proteína quiere decirse una secuencia con alto porcentaje de identidad o alto porcentaje de homología con secuencias a niveles nucleotídicos o polipeptídicos. Por alto porcentaje de identidad o alto porcentaje de homología está previsto al menos el 70%, más preferiblemente al menos el 75%, más preferiblemente al menos el 80%, más preferiblemente al menos el 85%, más preferiblemente al menos el 90%, más preferiblemente al menos el 95%, más preferiblemente al menos el 97%, más preferiblemente al menos el 99% o cualquier valor de porcentaje entre el 70% y el 99%.

- "Identidad" se refiere a identidad de secuencia entre dos moléculas de ácido nucleico o polipéptidos. La identidad puede determinarse comparando una posición en cada secuencia que puede alinearse para fines de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base, entonces las moléculas son idénticas en esa posición. Un grado de similitud o identidad entre secuencias de ácido nucleico o aminoácidos es una función del número de nucleótidos idénticos o coincidentes en posiciones compartidas por las secuencias de ácido nucleico. Pueden usarse diversos algoritmos y/o programas de alineación para calcular la identidad entre dos secuencias, incluyendo FASTA o BLAST que están disponibles como parte del paquete de análisis de secuencias GCG (Universidad de Wisconsin, Madison, Wis.), y pueden usarse con, por ejemplo, parámetros por defecto.

25 - El término "escisión" se refiere a la rotura de la estructura principal covalente de un polinucleótido. La escisión puede iniciarse mediante una variedad de métodos incluyendo, pero sin limitarse a, hidrólisis química o enzimática de un enlace fosfodiéster. Son posibles tanto escisión monocatenaria como escisión bicatenaria, y la escisión bicatenaria puede producirse como resultado de dos acontecimientos de escisión monocatenaria distintos. La escisión de ADN bicatenario, ARN o híbridos de ADN/ARN puede dar como resultado la producción de o bien extremos romos o bien extremos escalonados.

30 - Los términos "sitio diana", "secuencia diana", "secuencia de ácido nucleico diana" o "secuencia diana de ácido nucleico" se refieren a una secuencia de ácido nucleico que define una porción de un ácido nucleico al que una molécula de unión se unirá y/o escindirán, siempre que estén presentes condiciones suficientes para la unión y/o escisión.

35 - Un "dominio" de una proteína es cualquier porción de la proteína entera, hasta e incluyendo la proteína completa, pero que comprende normalmente menos que la proteína completa. Un dominio puede, pero no necesariamente, plegarse independientemente del resto de la cadena de proteína y/o correlacionarse con una ubicación o función biológica, bioquímica o estructural particular (por ejemplo, un dominio endonucleasa, un dominio de unión a polinucleótido, tal como un dominio de unión a ADN o un dominio de procesamiento de extremos).

40 - Por receptor de antígeno quimérico (CAR) está previsto moléculas que combinan un dominio de unión contra un componente presente en la célula diana, por ejemplo una especificidad basada en anticuerpo para un antígeno deseado (por ejemplo, antígeno tumoral) con un dominio intracelular de activación de receptores de células T para generar una proteína quimérica que presenta una actividad inmunitaria celular anti-diana. Generalmente, el CAR consiste en un anticuerpo de cadena sencilla extracelular (scFvFc) fusionado al dominio de señalización intracelular de la cadena zeta del complejo de receptor de antígeno de células T (scFvFc:ξ) y tiene la capacidad, cuando se expresa en células T, de redirigir el reconocimiento de antígenos basándose en la especificidad del anticuerpo monoclonal.

45 - El término "sujeto" tal como se usa en el presente documento incluye todos los miembros del reino animal incluyendo primates no humanos y seres humanos.

Ejemplos:

55 Ejemplo 1: Se realizó modificación por ingeniería genética de prototipos de LHE con superficies de contacto de reconocimiento de ADN específicas para dianas en el gen de TRAC humano

En primer lugar, se identificaron secuencias diana de LHE putativas en el gen de TRAC humano para las que el

inventor predijo superficies de contacto de reconocimiento de ADN modificadas por ingeniería genética de alta calidad. Tales predicciones se basan en una serie de características intrínsecas para el armazón de LHE, *I-Onul* (SEQ ID NO: 1), con las cuales iban a modificarse por ingeniería genética las superficies de contacto de reconocimiento de ADN de TRAC. Otras consideraciones, tales como ubicaciones dentro del gen de TRAC que es probable que provoquen alteraciones significativas en la proteína TCR-alfa tras inserciones o deleciones mediadas por endonucleasa, y/o la aparición de codones de terminación TGA, TAG o TAA en el sentido de 3' adyacentes en marcos de lectura alternativos para limitar la producción de péptidos fuera de marco que podrían servir como base para el rechazo inmunológico, también se incorporaron en el proceso de elección de dianas. Véase la figura 1 que ilustra esquemáticamente las ubicaciones de las supuestas secuencias diana.

Se eligieron dos secuencias diana putativas (TCRA_S02 y TCRA_S10; SEQ ID NO: 3 y SEQ ID NO: 4) para las fases iniciales de modificación por ingeniería genética de la superficie de contacto de reconocimiento de ADN. Se construyeron bibliotecas de variantes, mediante lo cual se variaron residuos de aminoácido en subregiones localizadas de la superficie de contacto de reconocimiento de ADN. Véase la figura 2, que muestra diagramas esquemáticos y estructurales de la superficie de contacto de reconocimiento de ADN. Se logró la variación dentro de la superficie de contacto de reconocimiento de ADN de la secuencia de ácido nucleico de *I-Onul* (SEQ ID NO: 1) incorporando codones degenerados en los oligonucleótidos que sirvieron como sustratos para las reacciones PCR para generar bibliotecas de variantes mediante recombinación de huecos en la cepa de levadura *Saccharomyces cerevisiae*. Se examinaron las bibliotecas resultantes para determinar la actividad de escisión de dianas mediante métodos basados en presentación en superficie y citometría de flujo tal como se ha descrito en (Jarjour, West-Foile *et al.* 2009). De esta manera, se alteró la especificidad de la superficie de contacto de reconocimiento de ADN para reconocer dianas en el gen de TRAC humano. En aspectos particulares, se lograron superficies de contacto de reconocimiento de ADN reespecificadas satisfactoriamente para TCRA_S02 (SEQ ID NO: 3) sólo, fallando el proceso para el otro sitio diana supuesto en una fase anterior en el proceso de modificación por ingeniería genética. Véase la figura 3 que ilustra el aislamiento satisfactorio de variantes que escinden la diana TCRA_S02.

Ejemplo 2: Se mostró que LHE con superficies de contacto de reconocimiento de ADN modificadas por ingeniería genética provocaban mutaciones disruptivas en las secuencias diana para las que se modificaron por ingeniería genética para reconocerlas.

Para medir la actividad del LHE que selecciona como diana TCRA, se usó un sistema de indicador fluorescente integrado cromosómicamente que se ha descrito anteriormente. En este sistema, el LHE de interés se transfecta en una línea celular de fibroblastos HEK 293T que se modifica por ingeniería genética para contener la secuencia diana TCRA_S02 en el sentido de 5' de un gen fuera de marco que codifica para la proteína fluorescente mCherry. La escisión de la diana TCRA_S02 incrustada e inserciones o deleciones pequeñas posteriores provocaron la reparación del ADN por medio de la ruta de unión de extremos no homólogos (NHEJ) dando como resultado aproximadamente 1 de tres loci reparados que colocan el gen indicador fluorescente 'en marco'. La fluorescencia en el canal de mCherry en un citómetro de flujo es, por tanto, una lectura de alto rendimiento supuesta de la escisión por LHE de la secuencia diana TCRA_S02 incrustada cromosómicamente.

Los resultados iniciales con la variante TCRA_S02_2E5 (SEQ ID NO: 5 que codifica para SEQ ID NO: 6) mostraron una eficacia muy baja de expresión de mCherry, indicando que esta variante no estaba escindiendo muy activamente su diana en un contexto cromosómico celular. Se realizaron mutagénesis al azar de la variante TCRA_S02_2E5 y el examen basado en presentación en superficie en condiciones de escisión más rigurosas para aislar variantes con actividades catalíticas mejoradas. Dos rondas de mutagénesis y examen condujeron a variantes con tasas de 40 a 50 veces superiores de generación de células que expresan mCherry. Véase la figura 4 que ilustra las lecturas de citometría de flujo del ensayo con indicador para variantes refinadas TCRA_S02_2E5, que incluían TCRA_S02_2E5_RD1_08 (SEQ ID NO: 7 que codifica para SEQ ID NO: 8) y TCRA_S02_2E5_RD2_23 (SEQ ID NO: 9 que codifica para SEQ ID NO: 10). Una variante con un rendimiento superior TCRA_S02_2E5_RD1_08 (SEQ ID NO: 7) contenía seis mutaciones en relación con la variante TCRA_S02_2E5, cuatro de las cuales están ubicadas dentro de la superficie de contacto de reconocimiento de ADN y dos están ubicadas en otra parte en la LHE. Véanse la figura 5 y la tabla 1 que proporcionan las alineaciones relativas de las variantes indicadas así como la información de posición de los residuos que comprende la superficie de contacto de reconocimiento de ADN. Se desconoce en qué grado, si lo hay, las mutaciones individuales identificadas a través de este proceso contribuyen a las características de la LHE que influyen en su actividad de reconocimiento y escisión de ADN. Tomados conjuntamente, conducen a mejoras significativas en la frecuencia de la aparición de mutaciones disruptivas en la secuencia diana TCRA_S02.

Ejemplo 3: Se diferenciaron LHE con superficies de contacto de reconocimiento de ADN que tienen alta afinidad, alta especificidad y baja toxicidad.

La LHE que contiene las superficies de contacto de reconocimiento de ADN modificadas por ingeniería genética para la diana TCRA_S02 (SEQ ID NO: 3) se sometió a prueba para determinar las características de afinidad, especificidad y toxicidad. Se sometió a prueba a la afinidad incubando independientemente levadura que presenta la variante TCRA_S02_2E5_RD1_08, (SEQ ID NO: 7, que codifica para SEQ ID NO: 8) con sustratos de ADN que contienen sus secuencias diana a diversas concentraciones. Véase la figura 6 que muestra las propiedades de afinidad de esta variante en relación con la proteína *I-Onul* silvestre. Estos datos demuestran que la variante

TCRA_S02_2E5_RD1_08 se une a su diana de ADN con afinidades comparables o superiores a las de la interacción entre la LHE I-Onul nativa y su secuencia diana (SEQ ID NO: 11). Se sometió a prueba la especificidad analizando la afinidad relativa y la capacidad de escisión de ADN de la variante de LHE hacia la secuencia diana que contenía cada uno de los tres pares de bases de ADN alternativos en cada posición a lo largo del sustrato. Véase la figura 7 que ilustra el perfil de especificidad de la variante TCRA_S02_2E5_RD1_08 (SEQ ID NO: 7). Estos datos demuestran que la variante de LHE TCRA_S02_2E5_RD1_08 tiene una alta especificidad global, ya que escinde/se une exclusivamente a su par de bases de sustrato específico en la mayoría de las posiciones a lo largo de su diana, sustituciones no tolerantes. Se analizó la toxicidad transcribiendo *in vitro* cada LHE para dar ARNm y transfectando células T humanas primarias mediante electroporación, seguido por análisis de citometría de flujo de la supervivencia de las células en relación con la transfección con un ARNm de control que codifica para una proteína fluorescente azul (BFP). Véase la figura 8 que muestra el análisis de citometría de flujo de células T humanas primarias tras la electroporación con ARNm-IVT que codifica para la variante TCRA_S02_2E5_RD1_08 LHE. Estos datos muestran que las variantes de LHE TCRA_S02 tienen toxicidad mínima en células humanas primarias.

Ejemplo 4: Se mostró que la expresión transitoria de LHE que selecciona como diana TRAC provoca pérdida de proteína TCR-alfa de la superficie celular y conduce a mutaciones disruptiva en el gen de TRAC.

A continuación, se examinó la LHE que selecciona como diana TRAC para determinar si: i) escindía eficazmente el sitio diana TCRA_S02 en el gen de TRAC en células humanas (SEQ ID NO: 3); y ii) las alteraciones mediadas por NHEJ resultantes daban como resultado la pérdida de la proteína TRAC de la superficie celular. Una motivación primaria para lograr altas eficacias es en el desarrollo de intervenciones terapéuticas humanas basadas en nucleasas que alteran TRAC. En una aplicación de este tipo, el uso de vectores virales que suministran de manera permanente (tales como para vectores retrovirales, lentivirales o de virus espumosos) o de manera transitoria (tales como vectores adenovirales o de virus adenoasociados) reactivos de nucleasa es laborioso, costoso y requiere muchos recursos, escasamente escalable y supone un reto abordarlo desde una perspectiva regulatoria. Un procedimiento y reactivo terapéutico más atractivo implicaría reemplazar el vector biológico por un reactivo de expresión sintético, tal como ARNm transcrito *in vitro* (ARNm-IVT). Se transfetaron células T humanas primarias con ARNm transcrito *in vitro* (ARNm-IVT) que codifica para o bien BFP o bien la LHE TCRA_S02_2E5, o variantes tal como se describió anteriormente usando métodos bien conocidos por los expertos en la técnica. La electroporación de ARNm-IVT produce un estallido transitorio de expresión de proteínas que dura de 4 a 12 horas. La duración y el grado de expresión de proteínas dependen de la estructura de ARNm-IVT. Un ejemplo del plásmido usado para la producción de ARNm-IVT se muestra en la figura 9 (SEQ ID NO: 12). Se añaden factores de estabilidad de ARNm secundario en el transcurso de la producción de ARNm-IVT. La caperuza de ARNm 5' (m7G) regula la expresión uniéndose a factores de iniciación eucariotas (eIF). La adición de una cola poliadenilada (poli(A)) retrasa la degradación de ARNm-IVT y aumenta la expresión de proteína LHE.

Tres días tras la electroporación de células T humanas primarias con ARNm-IVT que codifica para Trex2 y LHE que selecciona como diana TRAC, más del 15% de las células habían perdido su expresión en superficie de la proteína TCR-alfa. Véase la figura 10 que demuestra el análisis por citometría de flujo la alteración de TRAC y la estabilidad de la alteración de TRAC a lo largo de un periodo de cultivo de dos semanas. Se clasificaron las células negativas para TCR-alfa usando citometría de flujo y se secuenció el sitio diana TCRA_S02 en el gen de TRAC para analizar la alteración genética del gen de TRAC y confirmar y caracterizar el espectro de mutaciones inducidas por LHE. El gen de TRAC se expresa en un modo monoalélico, debido al silenciamiento del alelo no productivo durante el desarrollo de células T. Véase la figura 11 para resultados de secuenciación que muestran que el 50% de todos los alelos de TRAC contenían acontecimientos de alteración de NHEJ en el sitio diana TCR_S02. La conclusión crítica es que las variantes TCRA_S02_LHE seleccionan como diana eficazmente el alelo de TRAC transcrito de manera activa para bloquear la expresión de TCR-alfa.

Ejemplo 5: La LHE que selecciona como diana TRAC se mejoró mediante fusión con dominios TALE, que permiten una alteración génica de TRAC más eficaz con métodos de suministro sintéticos transitorios.

Los resultados de los ejemplos proporcionados anteriormente demuestran que los reactivos de nucleasa descritos en el presente documento pueden generar eficazmente células primarias que carecen de expresión de TCR-alfa. El objetivo clave de la enzima LHE y variantes descritas es la generación de células T deficientes en TCR-alfa para el tratamiento de cáncer, autoinmunidad e infección viral. Véase la figura 12 que ilustra esquemáticamente la estrategia terapéutica propuesta mediante la cual se combina la LHE que selecciona como diana TRAC con un reactivo de CAR secundario para producir células T para el tratamiento de cáncer y otras enfermedades. Los estudios iniciales con el suministro de Trex2 y LHE que selecciona como diana TCRA_S02 en la forma de ARNm-IVT mostraron tasas globales prometedoras pero subóptimas de alteración génica de TRAC. Aunque sin limitarse, tasas superiores de alteración de TCR-alfa simplificarían la fabricación de productos y darían como resultado costes de desarrollo reducidos.

Se creó una arquitectura de endonucleasa quimérica que mejoraría la eficacia de la LHE que selecciona como diana TRAC. Tal como se describe en el presente documento, las proteínas TALE ofrecen un modo modular único de reconocimiento de ADN. Los inventores razonaron, por tanto, que un alineamiento de repeticiones de TALE que reconocían una secuencia diana adyacente a la diana TCRA_S02 se fusionaría a la LHE que selecciona como diana

TCRA_S02 para potenciar eficazmente la colocalización de la nucleasa y su sustrato. Véase la figura 13 que ilustra esquemáticamente la endonucleasa quimérica y su secuencia de reconocimiento (SEQ ID NO: 3). Esta endonucleasa quimérica (SEQ ID NO: 14 codificada por SEQ ID NO: 13), una arquitectura denominada 'MegaTAL', se convirtió entonces en ARNm mediante métodos de transcripción *in vitro* descritos anteriormente. Entonces se suministraron especies de ARNm-IVT que codifican para la exonucleasa Trex2 y la MegaTAL que selecciona como diana TCRA_S02 mediante electroporación en células T humanas primarias. Este método de expresión transitoria de estos reactivos de nucleasa dio como resultado una eliminación extremadamente eficaz de la proteína TCR-alfa de la superficie celular. Véase la figura 14 que demuestra la eficacia de este enfoque usando análisis por citometría de flujo de la expresión de TCR-alfa en células T primarias tratadas con especies de ARNm-IVT que codifican para la exonucleasa Trex2 y megaTAL, ya que >65% de las células carecían de expresión TCR-alfa.

Ejemplo 6: Fusiones de TALE-LHE, que seleccionan como diana TCRA, se mejoraron mediante fusión con Trex2, que permite la alteración génica de TCRA ultraeficaz con una proteína de fusión de tres componentes expresada a partir de una única especie de ARNm.

A continuación, se evaluó la alteración génica de TCRA eficaz para el logro de mediante el suministro de una única especie de ARNm que expresa una proteína de fusión que comprende un alineamiento de TAL, la LHE que selecciona como diana TCRA.S02 y Trex2. Se colocó la proteína de fusión de tres componentes (SEQ ID NO: 16 codificada por SEQ ID NO: 15) en un vector que contenía un promotor de T7 para facilitar la transcripción *in vitro* y posterior poliadenilación y ocupación de extremos. Se suministró el ARNm resultante a células T humanas primarias mediante electroporación y se evaluó la expresión del complejo de CD3 sobre la superficie celular 72 horas más tarde mediante citometría de flujo (figura 16). Las muestras de control incluían células T humanas primarias no transfectadas, células T transfectadas con la MegaTAL que selecciona como diana TCRA.S02 y una muestra en donde se cotransfectó la MegaTAL que selecciona como diana TCRA.S02 con una especie de ARNm sintetizada independientemente que codifica para Trex2. Las muestras que reciben Trex2 o bien independientemente o bien como una fusión directa con la MegaTAL que selecciona como diana TCRA.S02 mostraron un porcentaje aumentado de células negativas para CD3, indicando tasas de alteración génica de TCRA potenciadas en estas muestras.

Bibliografía:

- Arimondo, P. B., C. J. Thomas, *et al.* (2006). "Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates." *Mol Cell Biol* 26(1): 324-33.
- Baxter, S., A. R. Lambert, *et al.* (2012). "Engineering domain fusion chimeras from I-Onul family LAGLIDADG homing endonucleases." *Nucleic Acids Res.*
- Bitinaite, J., D. A. Wah, *et al.* (1998). "FokI dimerization is required for DNA cleavage." *Proc Natl Acad Sci U S A* 95(18): 10570-5.
- Boch, J., H. Scholze, *et al.* (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." *Science* 326(5959): 1509-12.
- Choulika, A., A. Perrin, *et al.* (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*." *Mol Cell Biol* 15(4): 1968-73.
- Christian, M., T. Cermak, *et al.* (2010). "Targeting DNA double-strand breaks with TAL effector nucleases." *Genetics* 186(2): 757-61.
- Cong, L., F. A. Ran, *et al.* (2013). "Multiplex genome engineering using CRISPR/Cas systems." *Science* 339(6121): 819-23.
- Critchlow, S. E. y S. P. Jackson (1998). "DNA end-joining: from yeast to man." *Trends Biochem Sci* 23(10): 394-8.
- Eisenschmidt, K., T. Lanio, *et al.* (2005). "Developing a programmed restriction endonuclease for highly specific DNA cleavage." *Nucleic Acids Res* 33(22): 7039-47.
- Jarjour, J., H. West-Foyle, *et al.* (2009). "High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display." *Nucleic Acids Res* 37(20): 6871-80.
- Jinek, M., K. Chylinski, *et al.* (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." *Science* 337(6096): 816-21.
- Kalish, J. M. y P. M. Glazer (2005). "Targeted genome modification via triple helix formation." *Ann N Y Acad Sci* 1058: 151-61.
- Kim, Y. G., J. Cha, *et al.* (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." *Proc Natl Acad Sci U S A* 93(3): 1156-60.

- Li, T., S. Huang, *et al.* (2011). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNACleavage domain." *Nucleic Acids Res* 39(1): 359-72.
- Li, T., S. Huang, *et al.* (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes." *Nucleic Acids Res* 39(14): 6315-25.
- 5 Ma, J. L., E. M. Kim, *et al.* (2003). "Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences." *Mol Cell Biol* 23(23): 8820-8.
- Mali, P., L. Yang, *et al.* (2013). "RNA-guided human genome engineering via Cas9." *Science* 339(6121): 823-6. Moore, I., M.
- Samalova, *et al.* (2006). "Transactivated and chemically inducible gene expression in plants." *Plant J* 45(4): 651-83.
- 10 Moscou, M. J. y A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors." *Science* 326(5959): 1501.
- Pabo, C. O., E. Peisach, *et al.* (2001). "Design and selection of novel Cys2His2 zinc finger proteins." *Annu Rev Biochem* 70: 313-40.
- 15 Padidam, M., M. Gore, *et al.* (2003). "Chemical-inducible, ecdysone receptor-based gene expression system for plants." *Transgenic Res* 12(1): 101-9.
- Paques, F. y P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy." *Curr Gene Ther* 7(1): 49-66.
- Perrin, A., M. Buckle, *et al.* (1993). "Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions." *Embo J* 12(7): 2939-47.
- 20 Pingoud, A. y G. H. Silva (2007). "Precision genome surgery." *Nat Biotechnol* 25(7): 743-4.
- Porteus, M. H. y D. Carroll (2005). "Gene targeting using zinc finger nucleases." *Nat Biotechnol* 23(8): 967-73.
- Rouet, P., F. Smih, *et al.* (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." *Proc Natl Acad Sci U S A* 91(13): 6064-8.
- 25 Rouet, P., F. Smih, *et al.* (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease." *Mol Cell Biol* 14(12): 8096-106.
- Sethuraman, J., A. Majer, *et al.* (2009). "Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa." *Mol Biol Evol* 26(10): 2299-315.
- 30 Simon, P., F. Cannata, *et al.* (2008). "Sequence-specific DNA cleavage mediated by bipyridine polyamide conjugates." *Nucleic Acids Res* 36(11): 3531-8.
- Smith, J., J. M. Berg, *et al.* (1999). "A detailed study of the substrate specificity of a chimeric restriction enzyme." *Nucleic Acids Res* 27(2): 674-81.
- Smith, J., M. Bibikova, *et al.* (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains." *Nucleic Acids Res* 28(17): 3361-9.
- 35 Takeuchi, R., A. R. Lambert, *et al.* (2011). "Tapping natural reservoirs of homing endonucleases for targeted gene modification." *Proc Natl Acad Sci U S A* 108(32): 13077-82.
- Thierry, A. y B. Dujon (1992). "Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes." *Nucleic Acids Res* 20(21): 5625-31.
- 40 Wang, R., X. Zhou, *et al.* (2003). "Chemically regulated expression systems and their applications in transgenic plants." *Transgenic Res* 12(5): 529-40.
- Zuo, J. y N. H. Chua (2000). "Chemical-inducible systems for regulated expression of plant genes." *Curr Opin Biotechnol* 11(2): 146-51.

Lista de secuencias

45 <110> CELLECTIS
PREGENEN INC.

<120> Endonucleasa de asentamiento LAGLIDADG que escinde el gen de receptor de células T alfa y usos de la

ES 2 716 867 T3

misma

<130> P81303518PCT00

5 <150> PA201370303
<151> 31-05-2013

<160> 16

10 <170> PatentIn versión 3.5

<210> 1

<211> 909

<212> ADN

15 <213> *Ophiostoma novo-ulmi*

<220>

<223> I-Onul silvestre

20 <400> 1

| | | | | | | |
|-------------|------------|------------|------------|------------|------------|-----|
| atggcataca | tgtcgcgcag | agagtccatc | aacccatgga | ttctgactgg | tttcgctgat | 60 |
| gccgaaggat | ccttcttgct | gagaatccga | aacaataaca | agagctccgt | gggttactct | 120 |
| accgagttgg | gctttcaaat | cactctgcac | aacaaggaca | aatcgattct | ggagaatatc | 180 |
| cagtcgactt | ggaaggtcgg | cgtgattgct | aactcaggcg | acaatgccgt | cagtctgaaa | 240 |
| gttacgcggt | tcgaagattt | gaaagtgatt | atcgaccact | tcgagaaata | tccgctgatt | 300 |
| accagaaat | tgggcgatta | caagttgttt | aaacaggcat | tcagcgtcat | ggagaacaaa | 360 |
| gaacatctta | aggagaatgg | gattaaggag | ctcgtacgaa | tcaaagctaa | gatgaattgg | 420 |
| ggtctcactg | acgaattgaa | aaaagcattt | ccagagaaca | ttagcaaaga | gcgccccctt | 480 |
| atcaataaga | acattccgaa | tttcaaatgg | ctggctggat | tcacatctgg | tgaaggctgc | 540 |
| ttctttgtga | acttgatcaa | gtccaaatct | aagctgggtg | tacaggttca | attggtcttc | 600 |
| agcattactc | agcacatcag | agacaagaac | ctgatgaatt | cattgataac | atacctaggc | 660 |
| tgtggttaca | tcaaagagaa | gaacaagtcc | gagttcagtt | ggctcgactt | tgtggttacc | 720 |
| aaattcagcg | atatcaacga | caagatcatt | ccggtattcc | aggaaaatac | tctgattggc | 780 |
| gtcaaaactcg | aggactttga | agattggtgc | aaggttgcca | aattgatcga | agagaagaaa | 840 |
| cacctgaccg | aatccggttt | ggatgagatt | aagaaaatca | agctgaacat | gaacaaaggt | 900 |
| cgtgtcttc | | | | | | 909 |

<210> 2

<211> 303

25 <212> PRT

<213> *Ophiostoma novo-ulmi*

<220>

<223> I-Onul silvestre

30

<400> 2

Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn
20 25 30

ES 2 716 867 T3

Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Gln Ile Thr
 35 40 45
 Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60
 Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys
 65 70 75 80
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
 85 90 95
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
 100 105 110
 Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
 115 120 125
 Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
 130 135 140
 Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
 145 150 155 160
 Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser
 165 170 175
 Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu
 180 185 190
 Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Arg Asp
 195 200 205
 Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile
 210 215 220
 Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr
 225 230 235 240
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp

ES 2 716 867 T3

275

280

285

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe
 290 295 300

5 <210> 3
 <211> 22
 <212> ADN
 <213> secuencia artificial

10 <220>
 <223> TCRA_S02 diana

<400> 3
 tgtctgccta ttcaccgatt tt 22

15 <210> 4
 <211> 22
 <212> ADN
 <213> secuencia artificial

20 <220>
 <223> TCRA_S010 diana

<400> 4
 ctagcacagt tttgtctgtg at 22

25 <210> 5
 <211> 909
 <212> ADN
 <213> secuencia artificial

30 <220>
 <223> LHE CCR5_S02_2E5

<400> 5
 atggcataca tgtcgcgcag agagtccatc aacccatgga ttctgactgg tttcgtgat 60
 gccgaaggat cattcatact agacatccgc aaccgaaaca acgaaagcaa cagataccga 120
 acttcgctga gattccagat caccctgcac aacaaggaca aatcgattct ggagaatatac 180
 cagtcgactt ggaaggtcgg caagatcaca aacagcggcg acagagccgt catgctgagg 240
 gtcacccggt tcgaagattt gaaagtgatt atcgaccact tcgagaaata tccgctgatt 300
 acccagaaat tgggcgatta caagtgttt aacagggcat tcagcgtcat ggagaacaaa 360
 gaacatctta aggagaatgg gattaaggag ctctgacgaa tcaaagctaa gatgaattgg 420
 ggtctcaatg acgaattgaa aaaagcattt ccagagaaca ttagcaaaga ggcccccctt 480
 atcaataaga acattccgaa tttcaaattg ctggctggat tcacatctgg tgaaggctac 540
 ttccggcgtga acctaaaaaa ggtaaagggc aacgcaaagg tatacgtggg actgagattc 600
 tcaatcacac agcacatcag agacaagaac ctgatgaatt cattgataac atacctaggc 660
 tgtggttcca tcagggagaa gaacaagtct gagttcagtt ggctcgagtt cgtcgtaac 720
 aaattcagcg atatcaacga caagatcatt ccggtattcc aggaaaatac tctgattggc 780
 gtcaaactcg aggactttga agattggtgc aaggttgcca aattgatcga agagaagaaa 840
 cacctgaccg aatccggttt ggatgagatt aagaaaatca agctgaacat gaacaaaggt 900
 cgtgtcttc 909

35 <210> 6
 <211> 303
 <212> PRT
 <213> secuencia artificial

40 <220>
 <223> LHE CCR5_S02_2E5

<400> 6

ES 2 716 867 T3

Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg
20 25 30

Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr
35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
50 55 60

Lys Val Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg
65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
100 105 110

Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser
165 170 175

Gly Glu Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala
180 185 190

Lys Val Tyr Val Gly Leu Arg Phe Ser Ile Thr Gln His Ile Arg Asp
195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile
210 215 220

ES 2 716 867 T3

Arg Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr
 225 230 235 240

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255

Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270

Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
 275 280 285

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe
 290 295 300

<210> 7
 <211> 909
 5 <212> ADN
 <213> secuencia artificial

<220>
 10 <223> LHE TCRA_S02_2E5_RD1_08

<400> 7
 atggcataca tgtcgcgcag agagtccatc aacccatgga ttctgactgg tttcgcctgat 60
 gccgaaggat cattcatact agacatccgc aaccgaaaca acgaaagcaa cagataaccga 120
 acttcgctga gattccagat caccctgcac aacaaggaca aatcgattct ggagaatatc 180
 cagtcgactt ggaaggtcgg caagatcaca aacagcggcg acagagccgt catgctgagg 240
 gtcacccggtt tcgaagattt gaaagtgatt atcgaccact tcgagaaata tccgctgatt 300
 acccagaaat tgggcgatta caagttgttt aaacaggcat tcagcgtcat ggagaacaaa 360
 gaacatctta aggagaatgg gattaaggag ctcgtacgaa tcaaagctaa gatgaattgg 420
 ggtctcaacg acgaattgaa aaaagcattt ccagagaaca ttagcaaaga gcgccccctt 480
 atcaataaga acattccgaa tttcaaatgg ctggctggat tcacatctgg tgagggctac 540
 ttcggcgtga atctaaaaaa ggtaaagggc aacgcaaagg tatacgtggg actgagattc 600
 tcaatctcac agcacatcag agacaagaac ctgatgaatt cattgataac atacctaggc 660
 tgtggttcca tctgggagaa gaacaagtct gagttcagtt ggctcgagtt cgtcgttaacc 720
 aaattcagcg atatcaacga caagatcatt ccggtattcc aggaaaatac tctgattggc 780
 gtcaaaactcg aggactttga agattggtgc aaggttgcca aattgatcga agagaagaaa 840
 cacctgaccg aatccggttt ggatgagatt aagaaaatca agctgaacat gaacaaaggt 900
 cgtgtcttc 909

<210> 8
 15 <211> 303
 <212> PRT
 <213> secuencia artificial

<220>
 20 <223> LHE TCRA_S02_2E5_RD1_08

<400> 8
 Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15

ES 2 716 867 T3

Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg
 20 25 30

Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr
 35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60

Lys Val Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg
 65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
 85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
 100 105 110

Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
 115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
 130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
 145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser
 165 170 175

Gly Glu Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala
 180 185 190

Lys Val Tyr Val Gly Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp
 195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile
 210 215 220

Trp Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr
 225 230 235 240

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
 245 250 255

Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
 260 265 270

ES 2 716 867 T3

Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
 275 280 285

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe
 290 295 300

<210> 9
 <211> 909
 <212> ADN
 <213> secuencia artificial

<220>
 <223> LHE TCRA_S02_2E5_RD2_23

<400> 9
 atggcataca tgtcgcgcag agagtccatc aaccatgga ttctgactgg tttcgctgat 60
 gccgaaggat cattcatact agacatccgc aaccgaaaca acgaaagcaa cagataccga 120
 acttcgctga gattccagat caccctgcac aacaaggaca aatcgattct ggagaatatac 180
 cagtcgactt ggaaggtcgg caagatcaca aacagcagtg acagagccgt catgctgagg 240
 gtcacccggt tcgaagattt gaaagtgatt atcgaccact tcgagaaata tccgctgatt 300
 acccagaaat tgggcgatta caagttgttt aacagggcat tcagcgtcat ggagaataaa 360
 gaacatctta aggagaatgg gattaaggag ctctgacgaa tcaaagctaa gatgaattgg 420
 ggtctcaatg acgaattgaa aaaagcattt ccagagaaca ttagcaaaga gcgccccctt 480
 atcaataaga acattccgaa tttcaaatgg ctggctggat tcacagctgg tgaaggctac 540
 ttcggcgtga atctaaaaaa ggtaaagggc accgcaaagg tatacgtggg actgagattc 600
 tcaatctcac agcacatcag agacaagaac ctgatgaatt cattgataac atacctaggc 660
 tgtggttcca tctgggagaa gaacaagtct gagttcagat ggctcgagtt cgtcgttaacc 720
 aaattcagcg atatcaacga caagatcatt ccggtattcc aggaaaatac tctgattggc 780
 gtcaaactcg aggactttga agattggtgc aaggttgcca aattgatcga agagaagaaa 840
 cacctgaccg aatccggttt ggatgagatt aagaaaatca agctgaacat gaacaaaggt 900
 cgtgtcttc 909

<210> 10
 <211> 303
 <212> PRT
 <213> secuencia artificial

<220>
 <223> LHE TCRA_S02_2E5_RD2_23

<400> 10
 Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr
 1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg
 20 25 30

Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr
 35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp
 50 55 60

ES 2 716 867 T3

Lys Val Gly Lys Ile Thr Asn Ser Ser Asp Arg Ala Val Met Leu Arg
65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys
85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln
100 105 110

Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile
115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp
130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu
145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ala
165 170 175

Gly Glu Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Thr Ala
180 185 190

Lys Val Tyr Val Gly Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp
195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile
210 215 220

Trp Glu Lys Asn Lys Ser Glu Phe Arg Trp Leu Glu Phe Val Val Thr
225 230 235 240

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn
245 250 255

Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val
260 265 270

Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp
275 280 285

Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe
290 295 300

<210> 11
<211> 22
<212> ADN

ES 2 716 867 T3

<213> secuencia artificial

<220>

<223> Onu-I diana silvestre

5 <400> 11
tttccactta ttcaaccttt ta 22

10 <210> 12
<211> 7842
<212> ADN
<213> secuencia artificial

15 <220>
<223> Vector lentiviral de LHE TCRA_S02

<400> 12

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|------|
| gacgtcaatg | tagtcttatg | caatactctt | gtagtcttgc | aacatggtaa | cgatgagtta | 60 |
| gcaacatgcc | ttacaaggag | agaaaaagca | ccgtgcatgc | cgattggtgg | aagtaaggtg | 120 |
| gtacgatcgt | gccttattag | gaaggcaaca | gacgggtctg | acatggattg | gacgaaccac | 180 |
| tgaattgccg | cattgcagag | atattgtatt | taagtgccta | gctcgataca | taaacgggtc | 240 |
| tctctggtta | gaccagatct | gagcctggga | gctctctggc | taactagggg | accactgct | 300 |
| taagcctcaa | taaagcttgc | cttgagtgc | tcaagtagtg | tgtgcccgtc | tgttgtgtga | 360 |
| ctctggtaac | tagagatccc | tcagaccctt | ttagtcagtg | tggaaaatct | ctagcagtgg | 420 |
| cgcccgaaca | gggacttgaa | agcgaagg | aaaccagagg | agctctctcg | acgcaggact | 480 |
| cggcttgctg | aagcgcgcac | ggcaagaggc | gagggcggc | gactggtgag | tacgccaaaa | 540 |
| atthtgacta | gcggaggcta | gaaggagaga | gatgggtgcg | agagcgtcag | tattaagcgg | 600 |
| gggagaatta | gatcgcgatg | ggaaaaaatt | cggttaaggc | cagggggaaa | gaaaaaatat | 660 |
| aaattaaaac | atatagtatg | ggcaagcagg | gagctagaac | gattcgcagt | taatcctggc | 720 |
| ctgtagaaa | catcagaagg | ctgtagacaa | atactgggac | agctacaacc | atcccttcag | 780 |
| acaggatcag | aagaacttag | atcattatat | aatacagtag | caaccctcta | ttgtgtgcat | 840 |
| caaaggatag | agataaaaaga | caccaaggaa | gctttagaca | agatagagga | agagcaaaac | 900 |
| aaaagtaaga | ccaccgcaca | gcaagcggcc | ctgatcttca | gacctggagg | aggagatatg | 960 |
| agggacaatt | ggagaagtga | attatataaa | tataaagtag | taaaaattga | accattagga | 1020 |
| gtagcaccca | ccaaggcaaa | gagaagagtg | gtgcagagag | aaaaagagc | agtgggaata | 1080 |
| ggagctttgt | tccctgggtt | cttgggagca | gcaggaagca | ctatgggcgc | agcgtcaatg | 1140 |
| acgctgacgg | tacaggccag | acaattatg | tctggtatag | tgcagcagca | gaacaatttg | 1200 |
| ctgagggcta | ttgaggcgca | acagcatctg | ttgcaactca | cagtctgggg | catcaagcag | 1260 |
| ctccaggcaa | gaatcctggc | tgtggaaaga | tacctaaagg | atcaacagct | cctggggatt | 1320 |
| tggggttgct | ctggaaaact | catttgcacc | actgctgtgc | cttggaatgc | tagttggagt | 1380 |
| aatgaatctc | tggaacagat | ttggaatcac | acgacctgga | tggagtggga | cagagaaatt | 1440 |
| aacaattaca | caagcttaat | acactcctta | attgaagaat | cgaaaacca | gcaagaaaag | 1500 |
| aatgaacaag | aattattgga | attagataaa | tgggcaagtt | tgtggaattg | gtttaaata | 1560 |
| acaaattggc | tgtggtatat | aaaattattc | ataatgatag | taggaggctt | ggtaggttta | 1620 |
| agaatagttt | ttgctgtact | ttctatagtg | aatagagtta | ggcagggata | ttcaccatta | 1680 |
| tcgtttcaga | cccacctccc | aaccccaggg | ggaccgcaca | ggcccgaagg | aatagaagaa | 1740 |
| gaaggtggag | agagagacag | agacagatcc | attcgattag | tgaacggatc | tcgacgggat | 1800 |
| cggttaactt | ttaaaagaaa | aggggggatt | gggggtaca | gtgcagggga | aagaatagta | 1860 |
| gacataatag | caacagacat | acaaaactaaa | gaattacaaa | aacaaattac | aaaaattcaa | 1920 |
| aatthtatcg | attacgcgtc | acgtgctagc | tgcagtaacg | ccatthtgca | aggcatggaa | 1980 |
| aaataccaaa | ccaagaatag | agaagttcag | atcaagggcg | ggtacatgaa | aatagctaac | 2040 |
| gttgggcca | acaggatata | tgcggtgagc | agtttcggcc | ccggcccggg | gccaagaaca | 2100 |
| gatggtcacc | gcagtttcgg | ccccggcccc | aggccaagaa | cagatggtcc | ccagatatgg | 2160 |
| ccaaccctc | agcagtttct | taagaccct | cagatgtttc | caggctcccc | caaggacctg | 2220 |
| aatgaccct | gcgccttatt | tgaattaacc | aatcagcctg | cttctcgctt | ctgttcgcgc | 2280 |
| gcttctgctt | cccagactct | ataaaagagc | tcacaacccc | tactcggcg | cgccagctct | 2340 |
| ccgacagact | gagtcgcccg | ctcgagccgc | caccatggga | tatccatag | atgtcccaga | 2400 |
| ttatgcgcca | cctaagaaga | aacgcaaagt | cacgcgatcg | cgagagagtg | ccatcaacc | 2460 |
| atggattctg | actggtttcg | ctgatgccga | aggatcattc | atactagaca | tccgcaaccg | 2520 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| aaacaacgaa | agcaacagat | accgaacttc | gctgagattc | cagatcacc | tgcaacaaca | 2580 |
| ggacaaatcg | attctggaga | atatccagtc | gacttggaa | gtcggcaaga | tcacaaacag | 2640 |
| cagtgcacaga | gccgtcatgc | tgagggtcac | ccgtttcgaa | gatttgaaag | tgattatcga | 2700 |
| ccacttcgag | aaatatccgc | tgattaccca | gaaattgggc | gattacaagt | tgtttaaaaca | 2760 |
| ggcattcagc | gtcatggaga | ataaagaaca | tcttaaggag | aatgggatta | aggagctcgt | 2820 |
| acgaatcaaa | gctaagatga | attggggtct | caatgacgaa | ttgaaaaaag | catttccaga | 2880 |
| gaacattagc | aaagagcgcc | cccttatcaa | taagaacatt | ccgaatttca | aatggctggc | 2940 |
| tgatttcaca | gctggtgaag | gctacttcgg | cgtgaatcta | aaaaaggtaa | agggcaccgc | 3000 |
| aaaggtatac | gtgggactga | gatttctcaat | ctcacagcac | atcagagaca | agaacctgat | 3060 |
| gaattcattg | ataacatacc | taggctgtgg | ttccatctgg | gagaagaaca | agtctgagtt | 3120 |
| cagatggctc | gagttcgtcg | taaccaaatt | cagcgatatac | aacgacaaga | tcattccggg | 3180 |
| attccaggaa | aatactctga | ttggcgtcaa | actcgaggac | tttgaagatt | gggcaaggt | 3240 |
| tgccaaattg | atcgaagaga | agaaacacct | gaccgaatcc | ggtttggatg | agattaagaa | 3300 |
| aatcaagctg | aacatgaaca | aaggtcgtgt | cttcagcggc | cgctcctgat | agtaatgacc | 3360 |
| tgcaagtcga | cctatgcatt | agggcggcca | attccgcccc | tctccccccc | ccccctccc | 3420 |
| ctccccccc | ctaactgta | ctggcogaag | ccgcttgaa | taaggccggt | gtgctgtgt | 3480 |
| ctatatgtta | ttttccacca | tattgccgtc | ttttggcaat | gtgaggccc | ggaaacctgg | 3540 |
| ccctgtcttc | ttgacgagca | ttcctagggg | tctttcccct | ctcgccaaag | gaatgcaagg | 3600 |
| tctgttgaat | gtcgtgaagg | aagcagttcc | tctggaagct | tcttgaagac | aaacaacgct | 3660 |
| tgtagcgacc | ctttgcaggc | agcggaaacc | cccacctggc | gacaggtgcc | tctgcccga | 3720 |
| aaagccacgt | gtataagata | cacctgcaaa | ggcggcacia | ccccagtgcc | acgttgtgag | 3780 |
| ttggatagtt | gtggaaagag | tcaaatggct | ctcctcaagc | gtattcaaca | aggggctgaa | 3840 |
| ggatgccag | aaggtacccc | attgtatggg | atctgatctg | ggcctcggg | gcacatgctt | 3900 |
| tacatgtggt | tagtcgaggt | taaaaaaacg | tctaggcccc | ccgaaccacg | gggacgtggt | 3960 |
| tttcccttga | aaaacacgat | gataagcttg | ccacaaccct | taccggtcgc | caccatgagc | 4020 |
| gagctgatta | aggagaacat | gcacatgaag | ctgtacatgg | agggcaccgt | ggacaacct | 4080 |
| cacttcaagt | gcacatccga | gggcgaaggc | aagccctacg | agggcaccga | gaccatgaga | 4140 |
| atcaaggtgg | tcgagggcgg | ccctctcccc | ttcgccttcg | acatcctggc | tactagcttc | 4200 |
| ctctacggca | gcaagacctt | catcaaccac | accaggggca | tccccgactt | cttcaagcag | 4260 |
| tccttccctg | agggettcc | atgggagaga | gtcaccacat | acgaagacgg | gggctgctg | 4320 |
| accgttacc | aggacaccag | cctccaggac | ggctgcctca | tctacaacgt | caagatcaga | 4380 |
| ggggtgaact | tcacatccaa | cgccctgtg | atgcgaaga | aaacactcgg | ctgggaggcc | 4440 |
| ttcaccgaga | cgctgtaccc | cgctgacggc | ggcctggaag | gcagaaacga | catggccctg | 4500 |
| aagctcgtgg | gcgggagcca | tctgatcgca | aacatcaaga | ccacatatag | atccaagaaa | 4560 |
| cccgctaaga | acctcaagat | gcctggcgtc | tactatgtgg | actacagact | ggaaagaatc | 4620 |
| aaggaggcca | acaacgagac | ctacgtcgag | cagcacgagg | tggcagtggc | cagatactgc | 4680 |
| gacctcccta | gcaactggg | gcacaagctt | aattgattct | agagtcgacc | gagcatctta | 4740 |
| ccgccattta | taccatatt | tgttctgttt | ttcttgattt | gggtatacat | ttaaatgtta | 4800 |
| atagaacaaa | aatcatttac | atctttagg | atctttagg | atatgtaatt | actagttcag | 4860 |
| gtgtattgcc | acaagcaaaa | catgttaaga | aactttccc | ttatttacgc | tctgttcctg | 4920 |
| ttaatcaacc | tctggattac | aaaatttgtg | aaagattgac | tgatattctt | aactatgtttg | 4980 |
| ctccttttac | gctgtgtgga | tatgctgctt | tatagcctct | gtatctagct | attgcttccc | 5040 |
| gtacggcttt | cgttttctcc | tccttgata | aatcctgggt | gctgtctctt | ttagaggagt | 5100 |
| tgtggcccgt | tgccgtcaa | cggtggcgtg | tgtgctctgt | gtttgcctgac | gcaaccccc | 5160 |
| ctggctgggg | cattgccacc | acctgtcaac | tcctttctgg | gactttcgct | ttccccctcc | 5220 |
| cgatcgccac | ggcagaactc | atcgccgctt | gccttgccc | ctgctggaca | ggggctaggt | 5280 |
| tgctggggc | tgataattcc | gtgggtgtgt | catcggtacc | tttttaaaag | aaaagggggg | 5340 |
| actggaagg | ctaattcact | cccaacgaag | acaagatatac | ataacttcgt | atagcataca | 5400 |
| ttatacgaag | ttataattta | tttgtgaaat | ttgtgatgct | attgctttat | ttgtaacct | 5460 |
| atgtttat | gtgaaatttg | tgatgctatt | gctttat | taaccattgc | tttttgcttg | 5520 |
| tactgggtct | ctctgggttag | accagatctg | agcctgggag | ctctctggct | aactagggaa | 5580 |
| cccactgctt | aagcctcaat | aaagcttgcc | tcgaccagcc | tcgactgtgc | cttctagttg | 5640 |
| ccagccact | gttgtttgcc | cctccccctg | gccttccttg | accctggaag | gtgccactcc | 5700 |
| cactgtcctt | tcctaataaa | atgaggaaat | tgcatcgcat | tgctctgagta | gggtctcttc | 5760 |
| tattctgggg | gttgggtgg | ggcaggacag | caagggggag | gattgggaag | acaatagcag | 5820 |
| gcatgctggg | gatcgggtgg | gctctatggc | ctgcagctgc | attaatgaat | cggccaacgc | 5880 |
| gcggggagag | gcggtttgcg | tattggggcg | tcttccgctt | cctcgctcac | tgactcgctg | 5940 |
| cgctcggtcg | ttcggtcgcg | gcgagcggta | tcagctcact | caaaggcggg | aatacggtta | 6000 |
| tccacagaat | caggggataa | cgcaggaaag | aacatgtgag | caaaaggcca | gcaaaaggcc | 6060 |
| aggaaccgta | aaaaggccgc | gttgcctggc | ttttccata | ggctccgccc | ccctgacgag | 6120 |
| catcacaana | atcgacgctc | aagtcagagg | tgggcaaac | cgacaggact | ataaagatac | 6180 |
| caggcgcttc | ccccggaag | ctccctcgtg | cgctctcctg | ttccgaccct | gcccgttacc | 6240 |
| ggatacctgt | ccgcctttct | cccttcggga | agcgtggcgc | tttctcatag | ctcacgctgt | 6300 |

ES 2 716 867 T3

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|------|
| aggatctca | gttcggtgta | ggtcgttcgc | tccaagctgg | gctgtgtgca | cgaaccccc | 6360 |
| gttcagcccg | accgctgcmc | cttatccggt | aactatcgtc | ttgagtccaa | cccggtaaga | 6420 |
| cacgacttat | cgccactggc | agcagccact | ggtaacagga | ttagcagagc | gaggtatgta | 6480 |
| ggcgggtgcta | cagagttctt | gaagtgggtg | cctaactacg | gctacactag | aaggacagta | 6540 |
| tttggtatct | gcgctctgct | gaagccagtt | accttcggaa | aaagagttgg | tagctcttga | 6600 |
| tccggcaaac | aaaccaccgc | tggtagcgg | ggtttttttg | tttgcaagca | gcagattacg | 6660 |
| cgcagaaaaa | aaggatctca | agaagatcct | ttgatctttt | ctacggggtc | tgacgctcag | 6720 |
| tggaacgaaa | actcacgtta | agggatthttg | gtcatgagat | tatcaaaaag | gatcttcacc | 6780 |
| tagatccttt | taaattaa | atgaagtttt | aatcaatct | aaagtatata | tgagtaaact | 6840 |
| tggctgaca | gttaccaatg | cttaatcagt | gaggcaccta | tctcagcgat | ctgtctatth | 6900 |
| cgttcatcca | tagttgcctg | actccccgtc | gtgtagataa | ctacgatacg | ggagggtta | 6960 |
| ccatctggcc | ccagtgcctg | aatgataccg | cgagaccac | gctcaccggc | tccagattta | 7020 |
| tcagcaataa | accagccagc | cggaagggcc | gagcgcagaa | gtggctctgc | aactttatcc | 7080 |
| gcctccatcc | agtctattaa | ttgttgccgg | gaagctagag | taagtagttc | gccagttaat | 7140 |
| agtttgcgca | acgttggtgc | cattgctaca | ggcatcgtgg | tgtcacgctc | gtcgtttggt | 7200 |
| atggcttcat | tcagctccgg | ttcccaacga | tcaaggcgag | ttacatgatc | ccccatggtg | 7260 |
| tgcaaaaaag | cggtagctc | cttcggtcct | ccgatcgttg | tcagaagtaa | gttggccgca | 7320 |
| gtgttatcac | tcatggttat | ggcagcactg | cataattctc | ttactgtcat | gccatccgta | 7380 |
| agatgctttt | ctgtgactgg | tgagtactca | accaagtcat | tctgagaata | gtgtatgagg | 7440 |
| cgaccgagtt | gctcttgccc | ggcgtcaata | cgggataata | ccgcgccaca | tagcagaact | 7500 |
| ttaaaagtgc | tcatcattgg | aaaacgthct | tccgggcgaa | aactctcaag | gatcttaccg | 7560 |
| ctgttgagat | ccagttcgat | gtaaccact | cgtgcacca | actgatcttc | agcatctth | 7620 |
| actttcacca | gcgthctg | gtgagcaaaa | acaggaaggc | aaaatgccgc | aaaaaaggga | 7680 |
| ataaggcgca | cacggaaatg | ttgaatactc | atactcttc | ttttcaata | ttattgaagc | 7740 |
| atthtaccag | gthattgtct | catgagcgga | tacatathtg | aatgtattta | gaaaaataaa | 7800 |
| caaatagggg | ttccgcgcac | atthccccga | aaagtgccac | ct | | 7842 |

<210> 13
 <211> 2661
 5 <212> ADN
 <213> secuencia artificial

<220>
 <223> TCRA_S02 de MegaTal

10

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| <400> 13 | | | | | | |
| atgggatcct | gcaggtatcc | atatgatgtc | ccagattatg | cgccaccta | gaagaaacgc | 60 |
| aaagtgcgtg | atctacgcac | gctcggctac | agtcagcagc | agcaagagaa | gatcaaaccg | 120 |
| aaggtgcgtt | cgacagtggc | gcagcaccac | gaggcactgg | tgggccatgg | gtttacacac | 180 |
| gcgcacatcg | ttgcgctcag | ccaacacccc | gcagccttag | ggaccgtcgc | tgtcacgtat | 240 |
| cagcacataa | tcacggcgtt | gccagaggcg | acacacgaag | acatcgttgg | cgtcggcaaa | 300 |
| cagtgggtccg | gcgcagcgc | cctggaggcc | ttgctcacgg | atgcggggga | gthgagaggt | 360 |
| ccgcccgttac | agthggacac | aggccaactt | gtgaagattg | caaaacgtgg | cggcgtgacc | 420 |
| gcaatggagg | cagtgcacgc | atcgcgcaat | gcactgacgg | gtgccccct | gaacctgacc | 480 |
| ccggaccaag | tggthggctat | cgccagcaac | atthggcgca | agcaagcgtc | cgaaaccgtg | 540 |
| cagcggctgt | tgcccgtgct | gtgccagac | catggcctga | ccccggacca | agthggtggct | 600 |
| atcggccagca | acattggcgg | caagcaagcg | ctcgaaacgg | tcagcggct | gthgcccgtg | 660 |
| ctgtgccagg | accatggcct | gaccccggac | caagthggtg | ctatcgccag | caacattggc | 720 |
| ggcaagcaag | cgctcgaaac | ggtgcagcgg | ctgthgcccg | tgctgtgcca | ggaccatggc | 780 |
| ctgacccccg | accaagtgg | ggctatcgcc | agcaaccgtg | gcggcaagca | agcgtcga | 840 |
| acggtgcagc | ggctgthgccc | ggtgctgtgc | caggaccatg | gcctgactcc | ggaccaagtg | 900 |
| gtggctatcg | ccagccacga | tggcggcaag | caagcgtcgc | aaacgggtga | gcggctgthg | 960 |
| ccggtgctgt | gccaggacca | tggcctgact | ccggaccaag | tggthggctat | cgccagccac | 1020 |
| gatggcggca | agcaagcgtc | cgaaacggtg | cagcggctgt | tgcccgtgct | gtgccaggac | 1080 |
| catggccctga | ccccggacca | agthggtggct | atcggcagca | acattggcgg | caagcaagcg | 1140 |
| ctcgaaacgg | tgcaagcggct | gthgcccgtg | ctgtgccagg | accatggcct | gaccccggac | 1200 |
| caagthggtg | ctatcgccag | caacaatggc | ggcaagcaag | cgctcgaac | ggtgcagcgg | 1260 |
| ctgthgcccg | tgtgtgcca | ggaccatggc | ctgacccccg | accaagtgg | ggctatcgcc | 1320 |
| agcaaccgtg | gcggcaagca | agcgtcga | acggtgcagc | ggctgthgccc | ggtgctgtgc | 1380 |
| caggaccatg | gcctgacccc | ggaccaagtg | gtggctatcg | ccagcaaca | tggcggcaag | 1440 |
| caagcgtcgc | aaacgggtga | gcggctgthg | ccggtgctgt | gccaggacca | tggcctgacc | 1500 |

ES 2 716 867 T3

ccggaccaag tgggtggctat cgccagcaac attggcggca agcaagcgct cgaaagcatt 1560
 gtggcccgagc tgagccggcc tgatccggcg ttggccgcgt tgaccaacga ccacctcgtc 1620
 gccttggcct gcctcggcgg acgtcctgcc atggatgcag tgaaaaaggg attgccgcac 1680
 gcgcccgaat tgatcagaag agtcaatcgc cgtattggcg aacgcacgtc ccatcgcgtt 1740
 gcgatatcta gagtgggagg aagctctcgc agagagtcca tcaacccatg gattctgact 1800
 ggtttcgctg atgccgaagg atcattcata ctagacatcc gcaaccgaaa caacgaaagc 1860
 aacagatacc gaacttcgct gagattccag atcacctgc acaacaagga caaatcgatt 1920
 ctggagaata tccagtcgac ttggaaggct ggcaagatca caaacagcag tgacagagcc 1980
 gtcattgctga gggtcacccg tttcgaagat ttgaaagtga ttatcgacca cttcgagaaa 2040
 tatccgctga ttaccagaa attgggcgat tacaagttgt ttaaacaggc attcagcgtc 2100
 atggagaata aagaacatct taaggagaat gggattaagg agctcgtacg aatcaaagct 2160
 aagatgaatt ggggtctcaa tgacgaattg aaaaaagcat ttccagagaa cattagcaaa 2220
 gagcgccccc ttatcaataa gaacattccg aatttcaaat ggctggctgg attcacagct 2280
 ggtgaaggct acttcggcgt gaatctaaaa aaggtaaagg gcaccgcaa ggtatacgtg 2340
 ggactgagat tctcaatctc acagcacatc agagacaaga acctgatgaa ttcattgata 2400
 acatacctag gctgtggttc catctgggag aagaacaagt ctgagttcag atggctcgag 2460
 ttcgtcgtaa ccaaattcag cgatatcaac gacaagatca ttccggtatt ccaggaaaat 2520
 actctgattg gcgtcaaact cgaggacttt gaagattggt gcaaggttgc caaattgatc 2580
 gaagagaaga aacacctgac cgaatccggt ttggatgaga ttaagaaaat caagctgaac 2640
 atgaacaaaag gtcgtgtctt c 2661

<210> 14
 <211> 886
 5 <212> PRT
 <213> secuencia artificial

<220>
 <223> TCRA_S02 de MegaTal

10

<400> 14
 Met Gly Ser Cys Arg Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Pro Pro
 1 5 10 15

 Lys Lys Lys Arg Lys Val Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln
 20 25 30

 Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln
 35 40 45

 His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His Ile Val
 50 55 60

 Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr
 65 70 75 80

 Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val
 85 90 95

 Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu
 100 105 110

 Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly
 115 120 125

ES 2 716 867 T3

Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala
 130 135 140

Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr
 145 150 155 160

Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
 165 170 175

Leu Glu Thr Val Gln Arg Leu Leu Val Leu Cys Gln Asp His Gly Leu
 180 185 190

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 195 200 205

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His
 210 215 220

Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 225 230 235 240

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 245 250 255

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 260 265 270

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 275 280 285

Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser
 290 295 300

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 305 310 315 320

Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 325 330 335

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 340 345 350

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val
 355 360 365

Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln

ES 2 716 867 T3

Ser Leu Arg Phe Gln Ile Thr Leu His Asn Lys Asp Lys Ser Ile Leu
625 630 635 640

Glu Asn Ile Gln Ser Thr Trp Lys Val Gly Lys Ile Thr Asn Ser Gly
645 650 655

Asp Arg Ala Val Met Leu Arg Val Thr Arg Phe Glu Asp Leu Lys Val
660 665 670

Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys Leu Gly
675 680 685

Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn Lys Glu
690 695 700

His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys Ala Lys
705 710 715 720

Met Asn Trp Gly Leu Thr Asp Glu Leu Lys Lys Ala Phe Pro Glu Asn
725 730 735

Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn Phe Lys
740 745 750

Trp Leu Ala Gly Phe Thr Ser Gly Asp Gly Tyr Phe Gly Val Asn Leu
755 760 765

Lys Lys Val Lys Gly Asn Ala Lys Val Tyr Val Gly Leu Arg Phe Ser
770 775 780

Ile Ser Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu Ile Thr
785 790 795 800

Tyr Leu Gly Cys Gly Ser Ile Trp Glu Lys Asn Lys Ser Glu Phe Ser
805 810 815

Trp Leu Glu Phe Val Val Thr Lys Phe Ser Asp Ile Asn Asp Lys Ile
820 825 830

Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu Glu Asp
835 840 845

Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Glu Glu Lys Lys His
850 855 860

Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu Asn Met
865 870 875 880

Asn Lys Gly Arg Val Phe

ES 2 716 867 T3

<210> 15
 <211> 3378
 <212> ADN
 <213> secuencia artificial

5

<220>
 <223> TCRA_S02_Trex2 de MegaTal

10

<400> 15
 atgggatcct gcaggtatcc atatgatgtc ccagattatg cgccacctaa gaagaaacgc 60
 aaagtcgtag atctacgcac gctcggctac agtcagcagc agcaagagaa gatcaaaccg 120
 aaggtgctgt cgacagtgcc gcagcaccac gaggcactgg tgggcatatg gtttacacac 180
 gcgcacatcg ttgcgctcag ccaacacccc gcagcgttag ggaccgtcgc tgtcacgtat 240
 cagcacataa tcacggcgtt gccagaggcg acacacgaag acatcgttgg cgtcggcaaa 300
 cagtggctcc gcgcacgcgc cctggaggcc ttgctcacgg atgcggggga gttgagaggt 360
 ccgccgttac agttggacac aggccaaact gtgaagattg caaaacgtgg cggcgtgacc 420
 gcaatggagg cagtgcacgc atcgcgcaat gcactgacgg gtgccccctt gaacctgacc 480
 ccggaccaag tgggtggctat cgccagcaac attggcggca agcaagcgtc cgaacgggtg 540
 cagcggctgt tgccgggtgt gtgccaggac catggcctga ccccgaccac agtggtggct 600
 atcgccagca acattggcgg caagcaagcg ctccgaaacg tgcagcggct gttgccgggtg 660
 ctgtgccagg acctatggct gaccccgacc caagtgggtg ctatcgccag caacattggc 720
 ggcaagcaag cgctcgaaac ggtgcagcgg ctgttgccgg tgctgtgcca ggaccatggc 780
 ctgacccccg accaagtggg ggtatcgcc agcaacgggt gcggcaagca agcgtcgaac 840
 acggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gcctgactcc ggaccaagtg 900
 gtggctatcg ccagccacga tggcggcaag caagcgtcgc aaacgggtga cgggctgttg 960
 ccggtgctgt gccaggacca tggcctgact ccggaccaag tgggtggctat cgccagccac 1020
 gatggcggga agcaagcgtc cgaacgggtg cagcggctgt tgccgggtgt gtgccaggac 1080
 catggcctga ccccgaccac agtggtggct atcgccagca acattggcgg caagcaagcg 1140
 ctcgaaacgg tgcagcggct gttgcgggtg ctgtgccagg acctggcctt gaccccgacc 1200
 caagtgggtg ctatcgccag caacaatggc ggcaagcaag cgctcgaaac ggtgcagcgg 1260
 ctgttgccgg tgctgtgcca ggacctggc ctgacccccg accaagtggg ggtatcgcc 1320
 agcaacgggt gcggaagca agcgtcgaac acggtgcagc ggctgttgcc ggtgctgtgc 1380
 caggaccatg gcctgacccc ggaccaagtg gtggctatcg ccagcaaaa tggcggcaag 1440
 caagcgtcgc aaacgggtga gcggctgttg ccggtgctgt gccaggacca tggcctgacc 1500
 ccggaccaag tgggtggctat cgccagcaac attggcggca agcaagcgtc cgaagcatt 1560
 gtggcccagc tgagcccggc tgatccggcg ttggccgctt tgaccaacga ccacctcgtc 1620
 gccttggcct gcctcggcgg acgtcctgcc atggatgcag tgaaaaaggg attgccgcac 1680
 gcgccggaat tgatcagaag agtcaatcgc cgtattggcg aacgcacgtc ccacgcgtt 1740
 gcgatataa gagtgggagg aagctctcgc agagagtcca tcaacctatg gattctgact 1800
 ggtttcgtct atgcccgaag atcattcata ctagacatcc gcaaccgaaa caacgaaagc 1860
 aacagatacc gaacttcgct gagattccag atcacctgac acaacaagga caaatcgatt 1920
 ctggagaata tccagtcac ttggaaggtc ggcaagatca caaacagcag tgacagagcc 1980
 gtcattgctg gggtcacccg tttcgaagat ttgaaagtga ttatcgacca cttcgagaaa 2040
 tatccgctga ttaccagaaa attggcggat tacaagttgt ttaaacaggc attcagcgtc 2100
 atggagaata aagaacatct taaggagaat gggattaagg agctcgtacg aatcaaagct 2160
 aagatgaatt ggggtctcaa tgacgaattg aaaaaagcat ttccagagaa cattagcaaa 2220
 gagcgcctcc ttatcaataa gaacattccg aatttcaaat ggctggctgg attcacagct 2280
 ggtgaaggct acttcggcgt gaatctaaaa aaggtaaagg gcaccgcaaa ggtatacgtg 2340
 ggactgagat tctcaatctc acagcacatc agagacaaga acctgatgaa ttcattgata 2400
 acatacctag gctgtggttc catctgggag aagaacaagt ctgagttcag atggctcag 2460
 ttctgctgaa ccaaatcag cgatatcaac gacaagatca ttccgggtatt ccaggaaaat 2520
 actcagattg gcgtcaaact cgaggacttt gaagattggg gcaaggttgc caaattgatc 2580
 aaagagaaga aacacctgac cgaatccggg ttggtatgaga ttaagaaaat caagtgaaac 2640
 atgaacaaaag gtcgtgtctt cgctagcacc ggttctgagc cacctcgggc tgagacctt 2700
 gtattcctgg acctagaagc cactgggctc ccaaactagg accctgagat tgacagagata 2760
 tccctttttg ctgttcacccg ctcttccctg gagaaccag aacgggatga ttctggttcc 2820
 ttggtgctgc cccgtgttct ggacaagctc aactgtgca tgtgcccgga gcgcccttt 2880

ES 2 716 867 T3

```

actgccaaagg ccagtgagat tactggtttg agcagcgaaa gcctgatgca ctgcgggaag 2940
gctggtttca atggcgctgt ggtaaggaca ctgcagggct tcctaagccg ccaggagggc 3000
cccatctgcc ttgtggccca caatggcttc gattatgact tcccactgct gtgcacggag 3060
ctacaacgtc tgggtgccca tctgccccaa gacactgtct gcctggacac actgcctgca 3120
ttgcggggcc tggaccgtgc tcacagccac ggcaccaggg ctcaaggccg caaaagctac 3180
agcctggcca gtctcttcca ccgctacttc caggctgaac ccagtgctgc ccattcagca 3240
gaaggtgatg tgcacaccct gcttctgatc ttctctgatc gtgctcctga gctgctcgcc 3300
tgggcagatg agcaggcccc cagctgggct catattgagc ccatgtacgt gccacctgat 3360
ggtccaagcc tcgaagcc 3378

```

<210> 16
 <211> 1126
 5 <212> PRT
 <213> secuencia artificial

<220>
 <223> TCRA_S02_Trex2 de MegaTal

10

<400> 16

Met Gly Ser Cys Arg Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Pro Pro
 1 5 10 15

Lys Lys Lys Arg Lys Val Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln
 20 25 30

Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln
 35 40 45

His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His Ile Val
 50 55 60

Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr
 65 70 75 80

Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val
 85 90 95

Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu
 100 105 110

Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly
 115 120 125

Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala
 130 135 140

Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr
 145 150 155 160

Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
 165 170 175

ES 2 716 867 T3

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 180 185 190

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys
 195 200 205

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 210 215 220

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly
 225 230 235 240

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 245 250 255

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 260 265 270

Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 275 280 285

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 290 295 300

Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 305 310 315 320

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 325 330 335

Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 340 345 350

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 355 360 365

Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val
 370 375 380

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 385 390 395 400

Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu
 405 410 415

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr

ES 2 716 867 T3

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
| | 420 | | 425 | | 430 | | | | | | | | | | | | | | |
| Pro | Asp | Gln | Val | Val | Ala | Ile | Ala | Ser | Asn | Gly | Gly | Gly | Lys | Gln | Ala | | | | |
| | 435 | | | | | | 440 | | | | | 445 | | | | | | | |
| Leu | Glu | Thr | Val | Gln | Arg | Leu | Leu | Pro | Val | Leu | Cys | Gln | Asp | His | Gly | | | | |
| | 450 | | | | | 455 | | | | | 460 | | | | | | | | |
| Leu | Thr | Pro | Asp | Gln | Val | Val | Ala | Ile | Ala | Ser | Asn | Asn | Gly | Gly | Lys | | | | |
| 465 | | | | | 470 | | | | | 475 | | | | | 480 | | | | |
| Gln | Ala | Leu | Glu | Thr | Val | Gln | Arg | Leu | Leu | Pro | Val | Leu | Cys | Gln | Asp | | | | |
| | | | | 485 | | | | | 490 | | | | | 495 | | | | | |
| His | Gly | Leu | Thr | Pro | Asp | Gln | Val | Val | Ala | Ile | Ala | Ser | Asn | Ile | Gly | | | | |
| | | | 500 | | | | | 505 | | | | | 510 | | | | | | |
| Gly | Lys | Gln | Ala | Leu | Glu | Ser | Ile | Val | Ala | Gln | Leu | Ser | Arg | Pro | Asp | | | | |
| | | 515 | | | | | 520 | | | | | 525 | | | | | | | |
| Pro | Ala | Leu | Ala | Ala | Leu | Thr | Asn | Asp | His | Leu | Val | Ala | Leu | Ala | Cys | | | | |
| | 530 | | | | | 535 | | | | | 540 | | | | | | | | |
| Leu | Gly | Gly | Arg | Pro | Ala | Met | Asp | Ala | Val | Lys | Lys | Gly | Leu | Pro | His | | | | |
| 545 | | | | | 550 | | | | | 555 | | | | | 560 | | | | |
| Ala | Pro | Glu | Leu | Ile | Arg | Arg | Val | Asn | Arg | Arg | Ile | Gly | Glu | Arg | Thr | | | | |
| | | | | 565 | | | | | 570 | | | | | 575 | | | | | |
| Ser | His | Arg | Val | Ala | Ile | Ser | Arg | Val | Gly | Gly | Ser | Ser | Arg | Arg | Glu | | | | |
| | | | 580 | | | | | 585 | | | | | 590 | | | | | | |
| Ser | Ile | Asn | Pro | Trp | Ile | Leu | Thr | Gly | Phe | Ala | Asp | Ala | Glu | Gly | Ser | | | | |
| | | 595 | | | | | 600 | | | | | 605 | | | | | | | |
| Phe | Ile | Leu | Asp | Ile | Arg | Asn | Arg | Asn | Asn | Glu | Ser | Asn | Arg | Tyr | Arg | | | | |
| | 610 | | | | | 615 | | | | | 620 | | | | | | | | |
| Thr | Ser | Leu | Arg | Phe | Gln | Ile | Thr | Leu | His | Asn | Lys | Asp | Lys | Ser | Ile | | | | |
| 625 | | | | | 630 | | | | | 635 | | | | | 640 | | | | |
| Leu | Glu | Asn | Ile | Gln | Ser | Thr | Trp | Lys | Val | Gly | Lys | Ile | Thr | Asn | Ser | | | | |
| | | | | 645 | | | | | 650 | | | | | 655 | | | | | |
| Ser | Asp | Arg | Ala | Val | Met | Leu | Arg | Val | Thr | Arg | Phe | Glu | Asp | Leu | Lys | | | | |
| | | | 660 | | | | | 665 | | | | | 670 | | | | | | |

ES 2 716 867 T3

Val Ile Ile Asp His Phe Glu Lys Tyr Pro Leu Ile Thr Gln Lys Leu
675 680 685

Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe Ser Val Met Glu Asn Lys
690 695 700

Glu His Leu Lys Glu Asn Gly Ile Lys Glu Leu Val Arg Ile Lys Ala
705 710 715 720

Lys Met Asn Trp Gly Leu Asn Asp Glu Leu Lys Lys Ala Phe Pro Glu
725 730 735

Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn Lys Asn Ile Pro Asn Phe
740 745 750

Lys Trp Leu Ala Gly Phe Thr Ala Gly Glu Gly Tyr Phe Gly Val Asn
755 760 765

Leu Lys Lys Val Lys Gly Thr Ala Lys Val Tyr Val Gly Leu Arg Phe
770 775 780

Ser Ile Ser Gln His Ile Arg Asp Lys Asn Leu Met Asn Ser Leu Ile
785 790 795 800

Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu Lys Asn Lys Ser Glu Phe
805 810 815

Arg Trp Leu Glu Phe Val Val Thr Lys Phe Ser Asp Ile Asn Asp Lys
820 825 830

Ile Ile Pro Val Phe Gln Glu Asn Thr Leu Ile Gly Val Lys Leu Glu
835 840 845

Asp Phe Glu Asp Trp Cys Lys Val Ala Lys Leu Ile Lys Glu Lys Lys
850 855 860

His Leu Thr Glu Ser Gly Leu Asp Glu Ile Lys Lys Ile Lys Leu Asn
865 870 875 880

Met Asn Lys Gly Arg Val Phe Ala Ser Thr Gly Ser Glu Pro Pro Arg
885 890 895

Ala Glu Thr Phe Val Phe Leu Asp Leu Glu Ala Thr Gly Leu Pro Asn
900 905 910

Met Asp Pro Glu Ile Ala Glu Ile Ser Leu Phe Ala Val His Arg Ser
915 920 925

ES 2 716 867 T3

Ser Leu Glu Asn Pro Glu Arg Asp Asp Ser Gly Ser Leu Val Leu Pro
 930 935 940

Arg Val Leu Asp Lys Leu Thr Leu Cys Met Cys Pro Glu Arg Pro Phe
 945 950 955 960

Thr Ala Lys Ala Ser Glu Ile Thr Gly Leu Ser Ser Glu Ser Leu Met
 965 970 975

His Cys Gly Lys Ala Gly Phe Asn Gly Ala Val Val Arg Thr Leu Gln
 980 985 990

Gly Phe Leu Ser Arg Gln Glu Gly Pro Ile Cys Leu Val Ala His Asn
 995 1000 1005

Gly Phe Asp Tyr Asp Phe Pro Leu Leu Cys Thr Glu Leu Gln Arg Leu
 1010 1015 1020

Gly Ala His Leu Pro Gln Asp Thr Val Cys Leu Asp Thr Leu Pro Ala
 1025 1030 1035 1040

Leu Arg Gly Leu Asp Arg Ala His Ser His Gly Thr Arg Ala Gln Gly
 1045 1050 1055

Arg Lys Ser Tyr Ser Leu Ala Ser Leu Phe His Arg Tyr Phe Gln Ala
 1060 1065 1070

Glu Pro Ser Ala Ala His Ser Ala Glu Gly Asp Val His Thr Leu Leu
 1075 1080 1085

Leu Ile Phe Leu His Arg Ala Pro Glu Leu Leu Ala Trp Ala Asp Glu
 1090 1095 1100

Gln Ala Arg Ser Trp Ala His Ile Glu Pro Met Tyr Val Pro Pro Asp
 1105 1110 1115 1120

Gly Pro Ser Leu Glu Ala
 1125

REIVINDICACIONES

1. *I-Onul* u homólogo de *I-Onul*, en la que se han introducido mutaciones para obtener una variante que escinde una secuencia de ácido nucleico diana dentro del gen de receptor de células T alfa constante (TRAC); comprendiendo dicha *I-Onul* o variante homóloga de *I-Onul* al menos 10 sustituciones de aminoácido en posiciones seleccionadas del grupo que consiste en: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 138, 159, 168, 178, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 207, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240, en referencia a SEQ ID NO: 2, y que tiene una identidad de secuencia de al menos el 70% con la secuencia proteica de SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10.
2. Variante según la reivindicación 1, en la que escinde una secuencia de ácido nucleico diana SEQ ID NO: 3.
3. Variante según la reivindicación 1 ó 2, que comprende al menos 15, preferiblemente al menos 20, más preferiblemente al menos 25 sustituciones de aminoácido en posiciones seleccionadas del grupo que consiste en: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 138, 159, 168, 178, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 207, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240, en referencia a SEQ ID NO: 2.
4. Variante según una cualquiera de las reivindicaciones 1 a 3, que tiene una identidad de secuencia de al menos el 75%, preferiblemente al menos el 80%, más preferiblemente al menos el 85%, más preferiblemente al menos el 90%, más preferiblemente al menos el 95%, más preferiblemente al menos el 97% y más preferiblemente el 99% con la secuencia proteica de SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10.
5. Variante según una cualquiera de las reivindicaciones 1 a 4, que tiene la secuencia proteica seleccionada del grupo que consiste en SEQ ID NO: 6, SEQ ID NO: 8 y SEQ ID NO: 10.
6. Endonucleasa quimérica que comprende a *I-Onul* o variante homóloga de *I-Onul* según una cualquiera de las reivindicaciones 1 a 5, fusionada a al menos un dominio proteico adicional seleccionado del grupo que consiste en: dominio de unión a ácido nucleico, dominio catalítico, etiquetas de epítipo terminal y proteínas fluorescentes.
7. Endonucleasa quimérica según la reivindicación 6, en la que dicho dominio proteico adicional es un dominio de unión a ácido nucleico seleccionado del grupo que consiste en TALE y dominio de dedos de zinc.
8. Endonucleasa quimérica según la reivindicación 7, en la que dicha endonucleasa quimérica es la secuencia proteica MegaTAL TCRA_S02 de SEQ ID NO: 14 y SEQ ID NO: 16.
9. Endonucleasa quimérica según la reivindicación 6 u 8, en la que el dominio proteico adicional tiene una actividad catalítica seleccionada del grupo que consiste en: actividad nucleasa, actividad polimerasa, actividad cinasa, actividad fosfatasa, actividad metilasa, actividad topoisomerasa, actividad integrasa, actividad transposasa, actividad ligasa, actividad helicasa, actividad recombinasa.
10. Endonucleasa quimérica según la reivindicación 9, en la que dicho dominio catalítico es una exonucleasa 5'-3', más preferiblemente Trex2 y más preferiblemente Trex2 de cadena sencilla.
11. Endonucleasa quimérica según una cualquiera de las reivindicaciones 6 a 10, en la que dicho dominio proteico adicional se fusiona a la variante de *I-Onul* mediante un ligador peptídico.
12. Polinucleótido que codifica para una variante de *I-Onul* según una cualquiera de las reivindicaciones 1 a 5 o endonucleasa quimérica según una cualquiera de las reivindicaciones 6 a 11.
13. Vector que comprende un polinucleótido según la reivindicación 12.
14. Método *in vitro* para modificar un gen de TRAC en un célula que comprende introducir una variante de *I-Onul* según una cualquiera de las reivindicaciones 1 a 5 o la endonucleasa quimérica según una cualquiera de las reivindicaciones 6 a 11 en dicha célula.
15. Método *in vitro* según la reivindicación 14, en el que la mutagénesis se aumenta introduciendo en la célula una enzima de procesamiento de extremos de ADN.
16. Método *in vitro* según la reivindicación 15, en el que dicha enzima de procesamiento de extremos de ADN se introduce en la célula con la variante de *I-Onul* o la endonucleasa quimérica como transgenes codificados por el mismo vector que comprende una secuencia de ácido nucleico que codifica para secuencia de salto ribosómico.
17. Método *in vitro* según la reivindicación 15 o 16, en el que dicha enzima de procesamiento de extremos de ADN tiene actividad exonucleasa 3'-5'.

18. Método *in vitro* según la reivindicación 17, en el que dicha enzima de procesamiento de extremos de ADN es TREX2.
19. Método *in vitro* según la reivindicación 17, en el que dicha enzima de procesamiento de extremos de ADN es un polipéptido de TREX2 de cadena sencilla.
- 5 20. Método *in vitro* según la reivindicación 14, que comprende introducir una matriz donadora que comprende una secuencia que va a introducirse flanqueada por al menos una secuencia que comparte homologías con al menos una región del gen de TRAC que rodea al sitio de escisión de ácido nucleico de dicha endonucleasa de corte poco común.

| Posición | Nativo | ^r | 2E5 | Refinado |
|----------|--------|--------------|-----|----------|
| 24 | Ser | | Ser | |
| 26 | Leu | | Ile | |
| 28 | Arg | | Asp | |
| 30 | Arg | | Arg | |
| 32 | Asn | | Arg | |
| 34 | Lys | | Asn | |
| 35 | Ser | | Glu | |
| 36 | Ser | | Ser | |
| 37 | Val | | Asn | |
| 38 | Gly | | Arg | |
| 40 | Ser | | Arg | |
| 42 | Glu | | Ser | |
| 44 | Gly | | Arg | |
| 46 | Gln | | Glu | |
| 48 | Thr | | Thr | |
| 68 | Val | | Lys | |
| 70 | Ala | | Thr | |
| 72 | Ser | | Ser | |
| 73 | Gly | | | Ser |
| 75 | Asn | | Arg | |
| 76 | Ala | | Ala | |
| 78 | Ser | | Met | |
| 80 | Lys | | Arg | |
| 82 | Thr | | Thr | |
| 176 | Ser | | | Ala |
| 180 | Cys | | Tyr | |
| 182 | Phe | | Gly | |
| 184 | Asn | | Asn | |
| 186 | Ile | | Lys | |
| 188 | Ser | | Val | |
| 189 | Lys | | Lys | |
| 190 | Ser | | Gly | |
| 191 | Lys | | Asn | Thr |
| 192 | Leu | | Ala | |
| 193 | Gly | | Lys | |
| 195 | Gln | | Tyr | |
| 197 | Gln | | Gly | |
| 199 | Val | | Arg | |
| 201 | Ser | | Ser | |
| 203 | Thr | | Thr | Ser |
| 223 | Tyr | | Ser | |
| 225 | Lys | | Arg | Trp |
| 227 | Lys | | Lys | |
| 229 | Lys | | Lys | |
| 232 | Phe | | Phe | |
| 233 | Ser | | | Arg |
| 234 | Trp | | Trp | |
| 236 | Asp | | Glu | |
| 238 | Val | | Val | |
| 240 | Thr | | Thr | |

Tabla 1

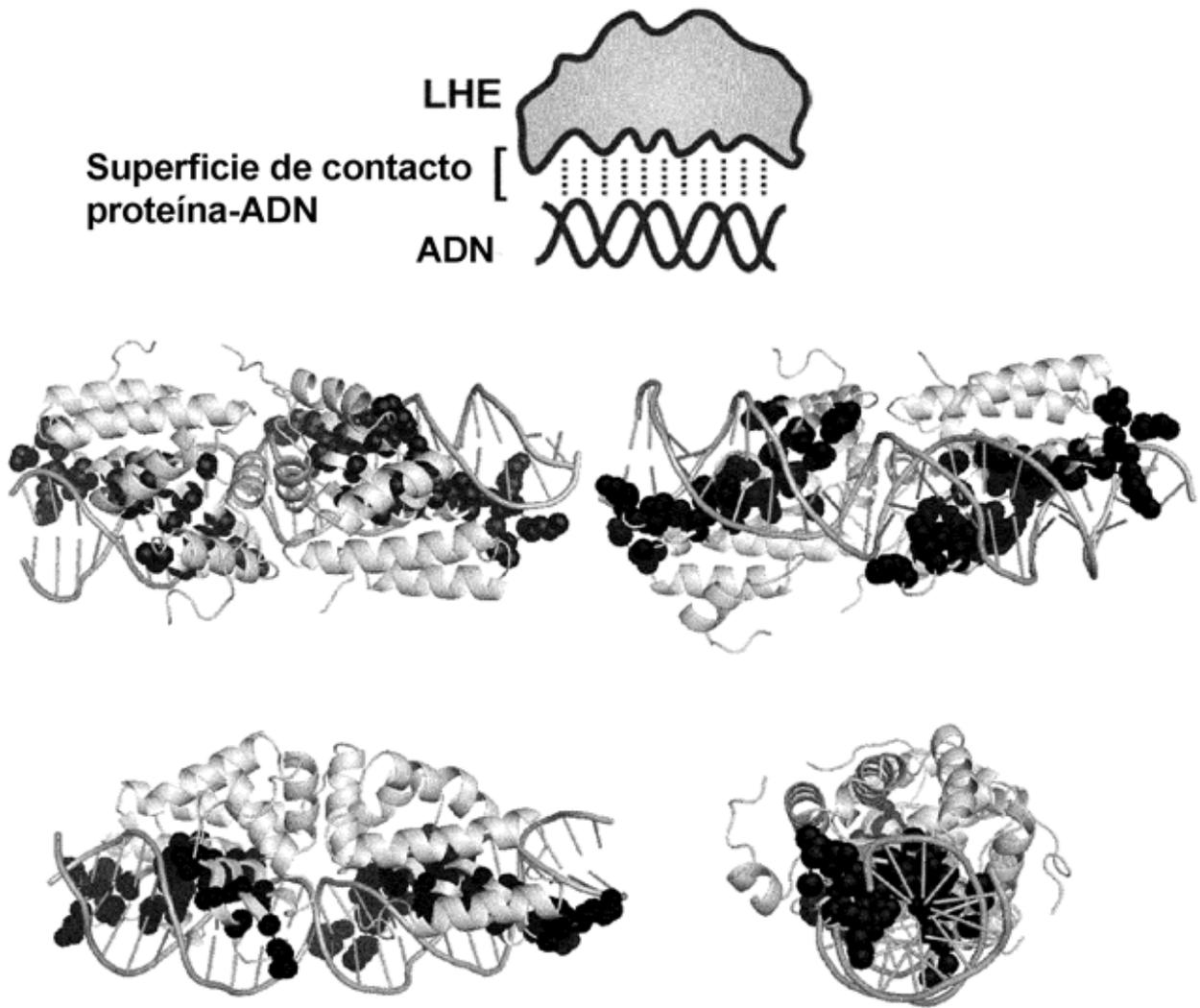


Figura 2

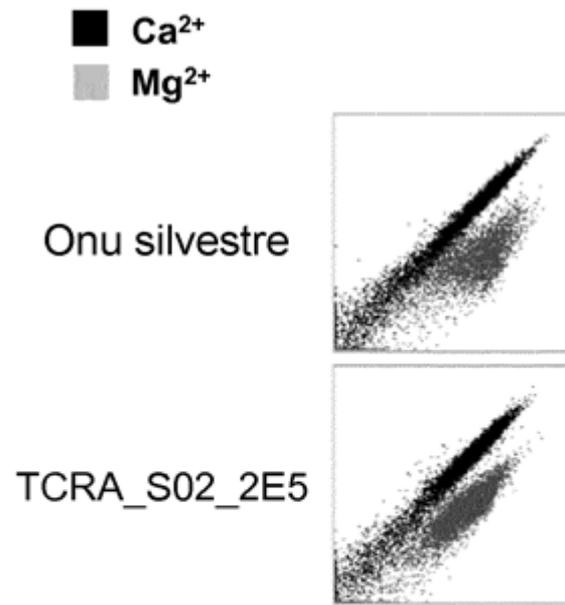


Figura 3

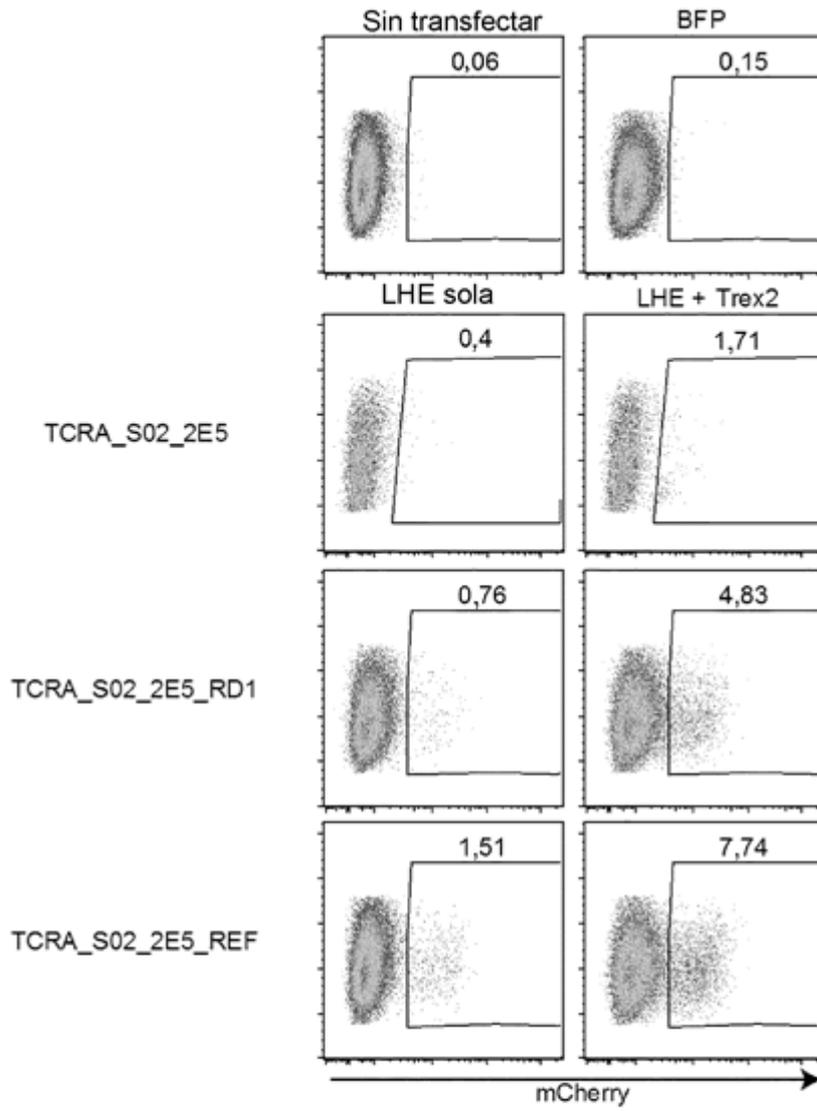


Figura 4

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 1 MAYMSRRESINPWILTGFADEGS **F**ILD **I**RRN**R**NN**R**YR**T**S**L**R**F**QITLHNKD 50
 10 MAYMSRRESINPWILTGFADEGS **F**ILD **I**RRN**R**NN**R**YR**T**S**L**R**F**QITLHNKD 50

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 60 KSILENIQSTWKV**G**KITNSG**D**RAV**M**LRVTRFEDLKV **I**DHFEKYPLITQKLG 100
 70 KSILENIQSTWKV**G**KITNS**S**DRAV**M**LRVTRFEDLKV **I**DHFEKYPLITQKLG 100

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 110 YKLFKQAFSVMENKEHLKENG **I**KELV**R**IKAKMNWGLNDELKKAFFPENISKERP 150
 120 YKLFKQAFSVMENKEHLKENG **I**KELV**R**IKAKMNWGLNDELKKAFFPENISKERP 150

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 160 LINKNIPNFKWLAGFT**A**GEG**Y**FGVNL**K**KK**M**GN**A**K**V**Y**V**GL**R**FSITQHIRDKNLM 210
 170 LINKNIPNFKWLAGFT**A**GEG**Y**FGVNL**K**KK**M**GN**A**K**V**Y**V**GL**R**FSITQHIRDKNLM 210

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 220 NSLITYLGCG**S**I**R**EK**N**K**S**E**F**S**W**L**F**V**V**T**K**F**S**D**I**N**D**K**I**I**P**V**F**Q**E**N**T**L**I**G**V**K**L**E**D** 250
 230 NSLITYLGCG**S**I**W**EK**N**K**S**E**F**R**W**L**F**V**V**T**K**F**S**D**I**N**D**K**I**I**P**V**F**Q**E**N**T**L**I**G**V**K**L**E**D** 250

TCRA_S02_2E5
 TCRA_2E5_RD2_23
 270 FEDWCKVAKLIEEK**K**H**L**TES**G**L**D**E**I**K**K**IK**L**N**M**N**K**G**R**V**F** 300
 280 FEDWCKVAKLIEEK**K**H**L**TES**G**L**D**E**I**K**K**IK**L**N**M**N**K**G**R**V**F** 300

Figura 5

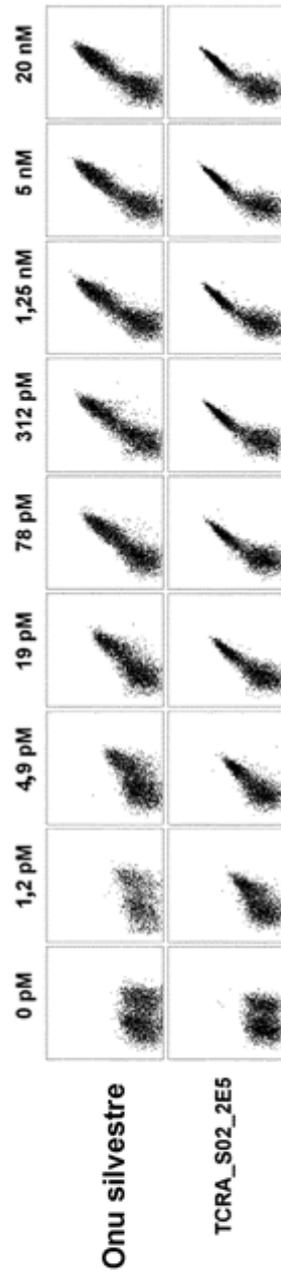


Figura 6

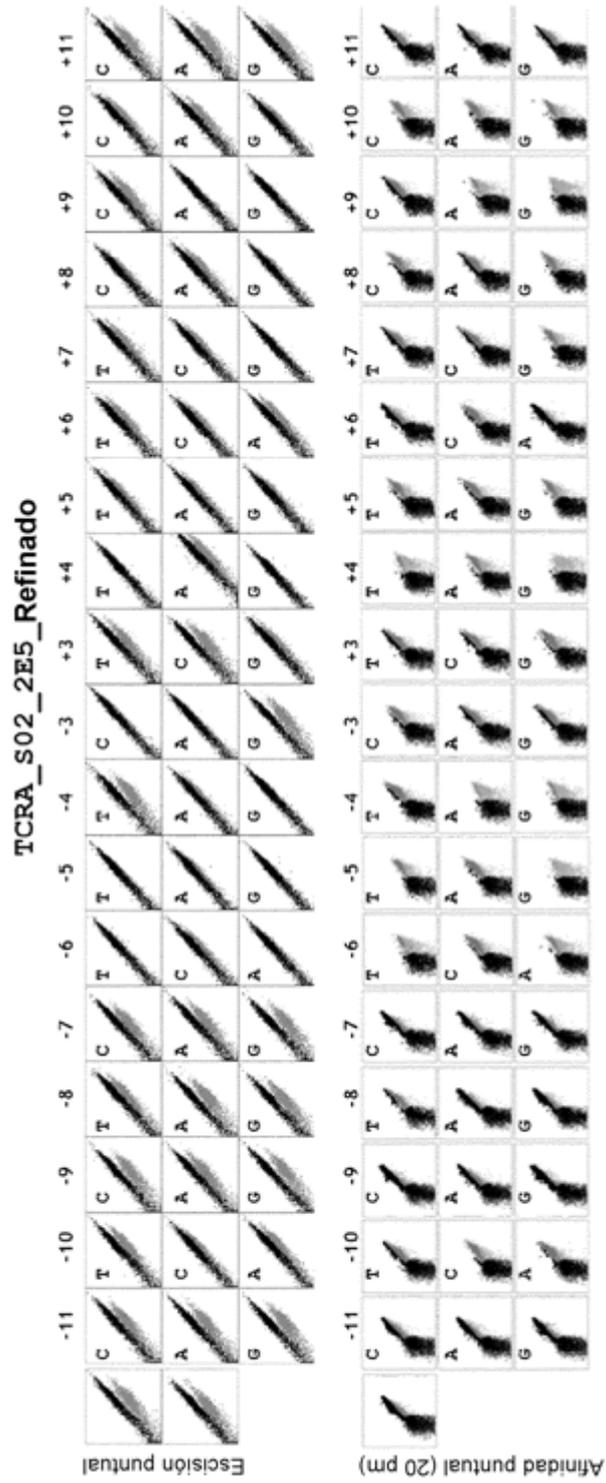


Figura 7

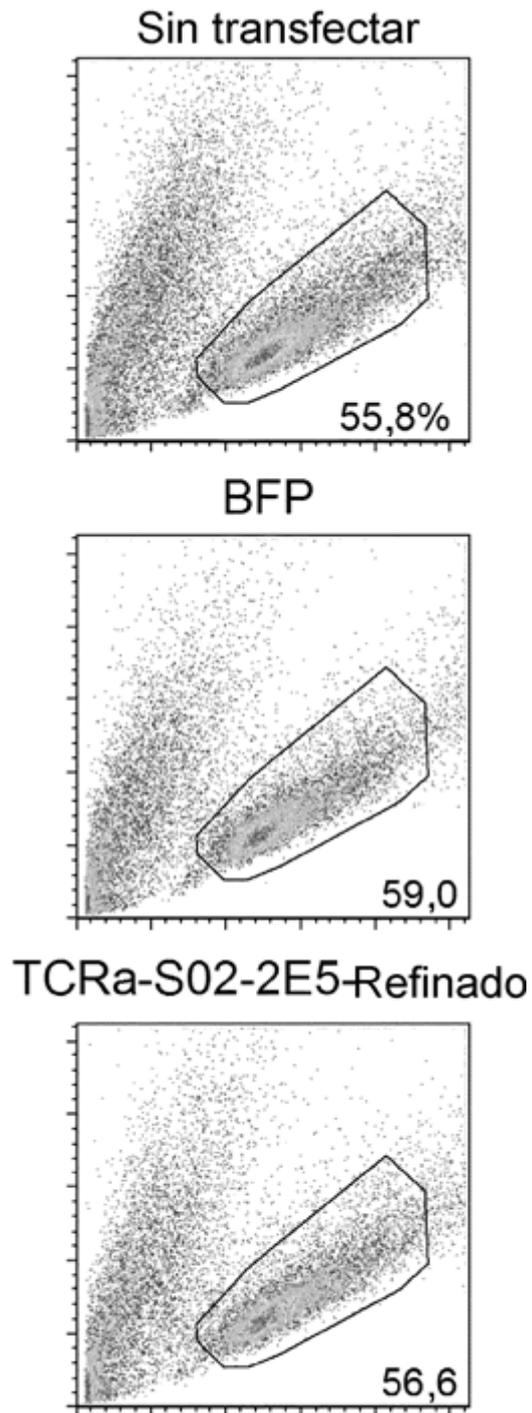


Figura 8

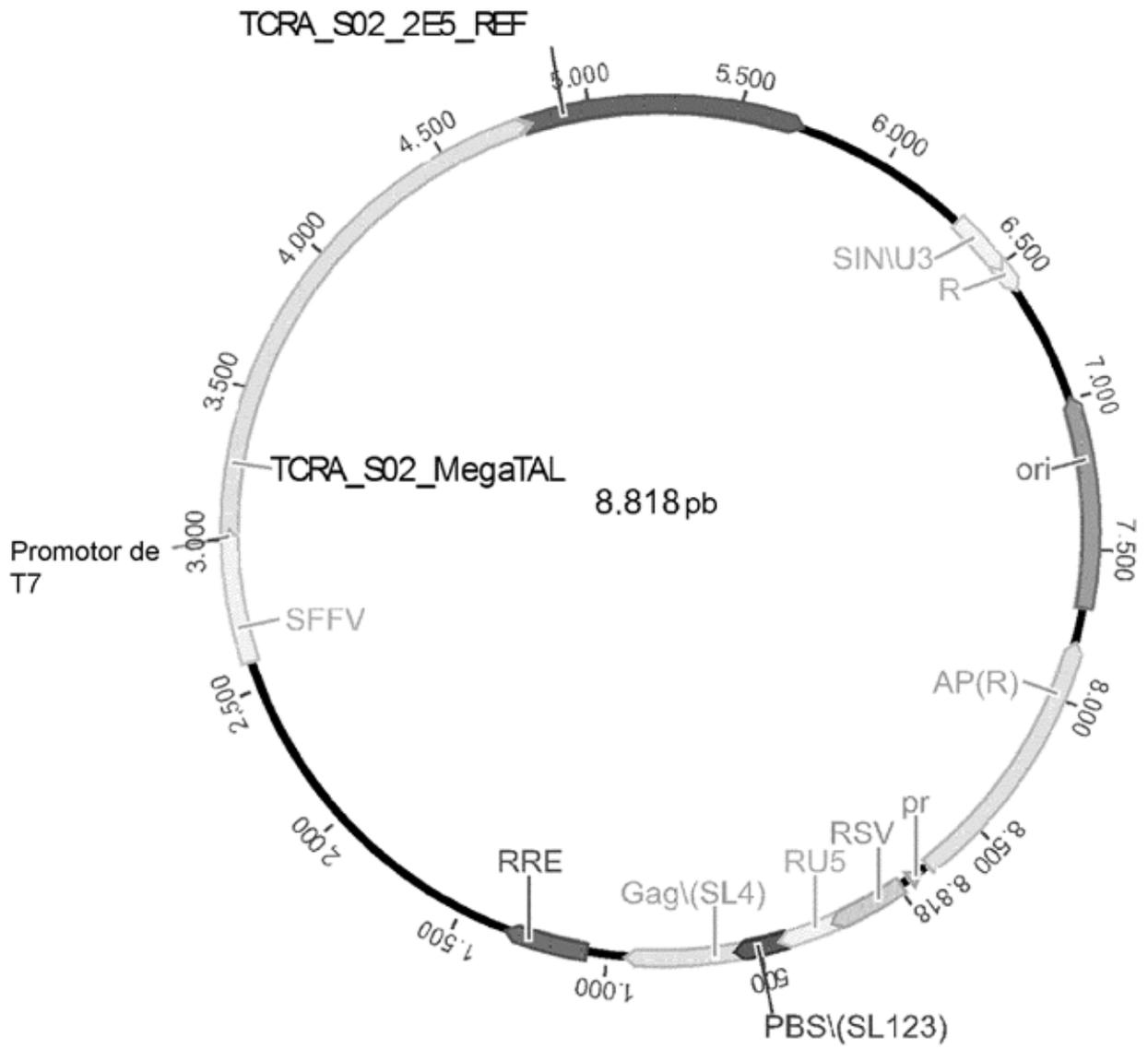


Figura 9

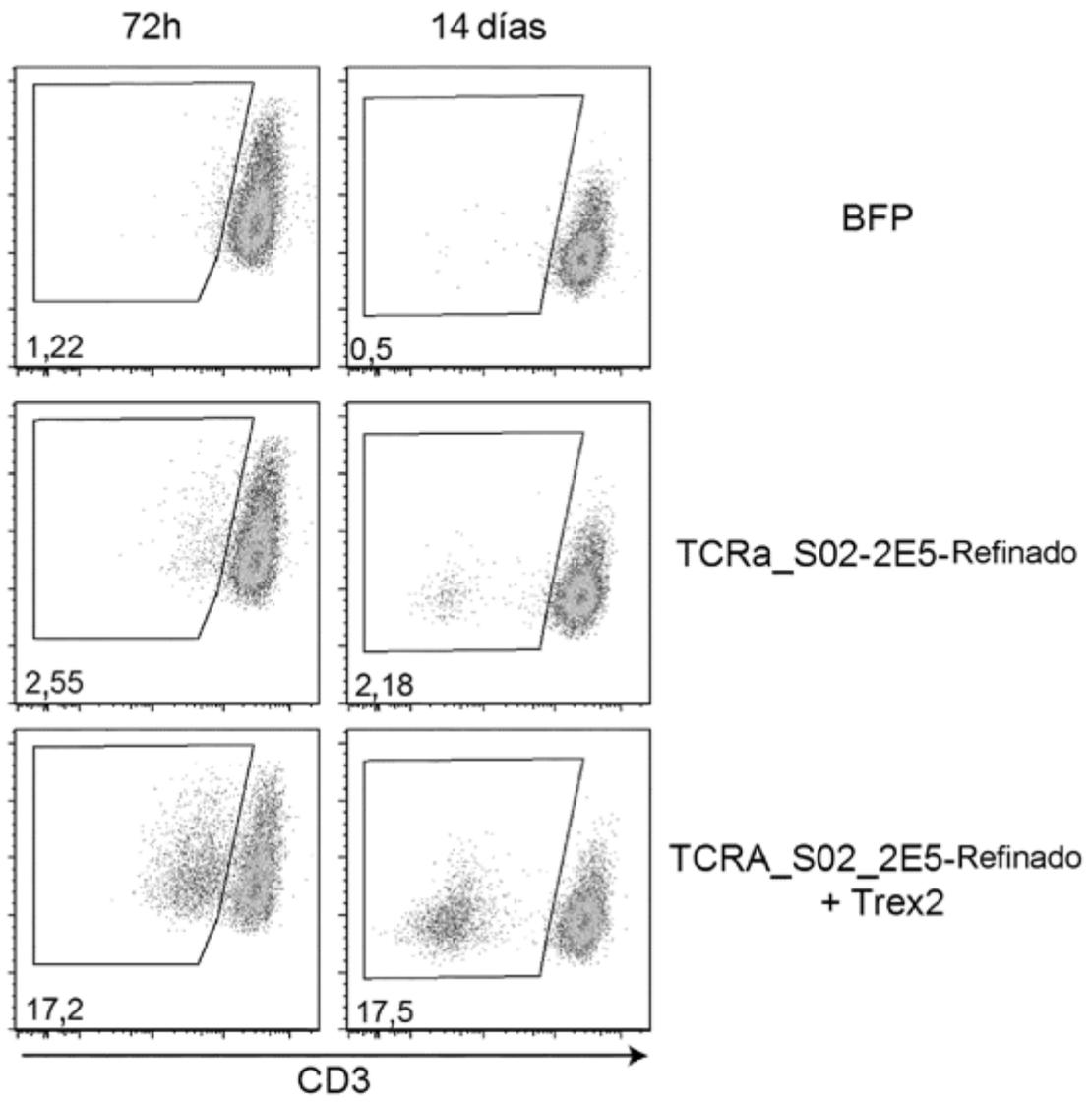


Figura 10

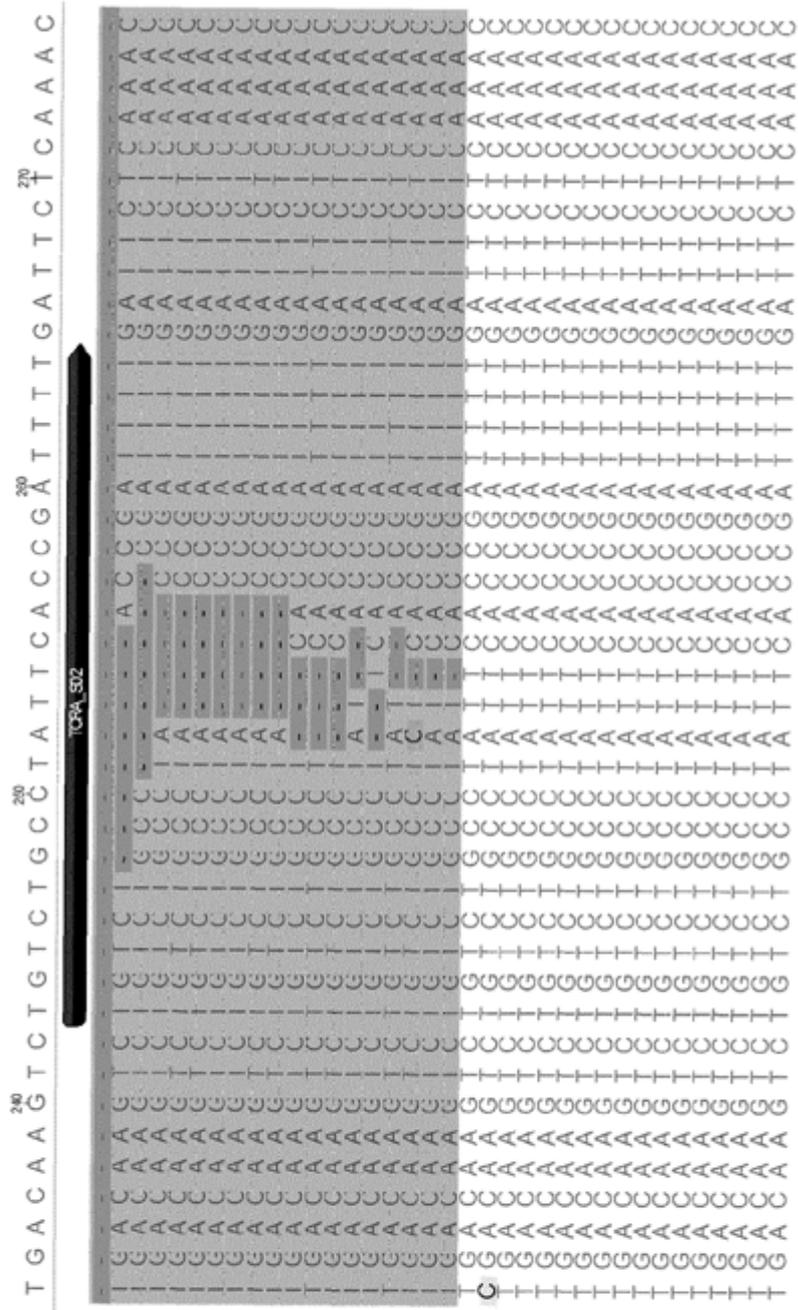


Figura 11

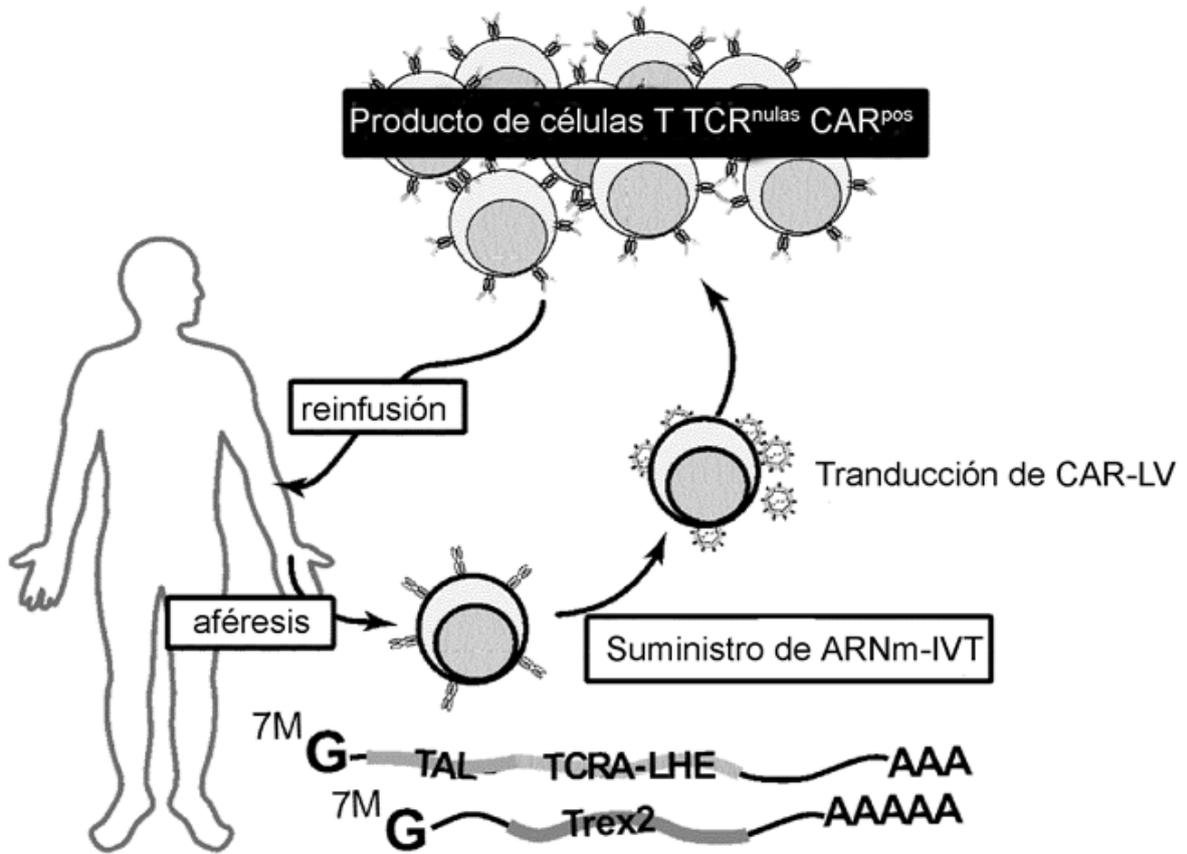


Figura 12

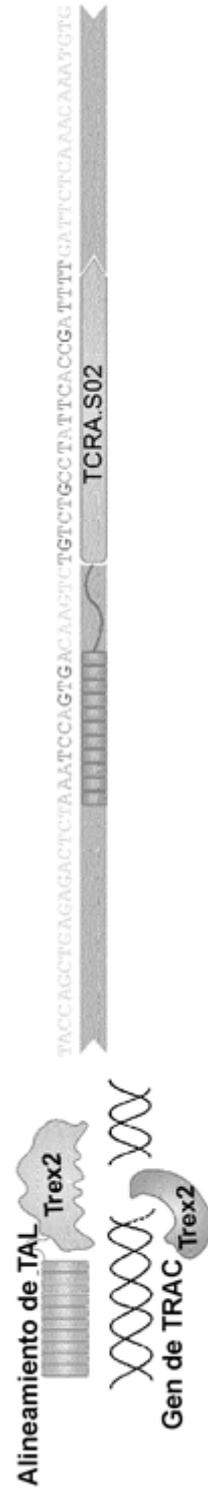


Figura 13

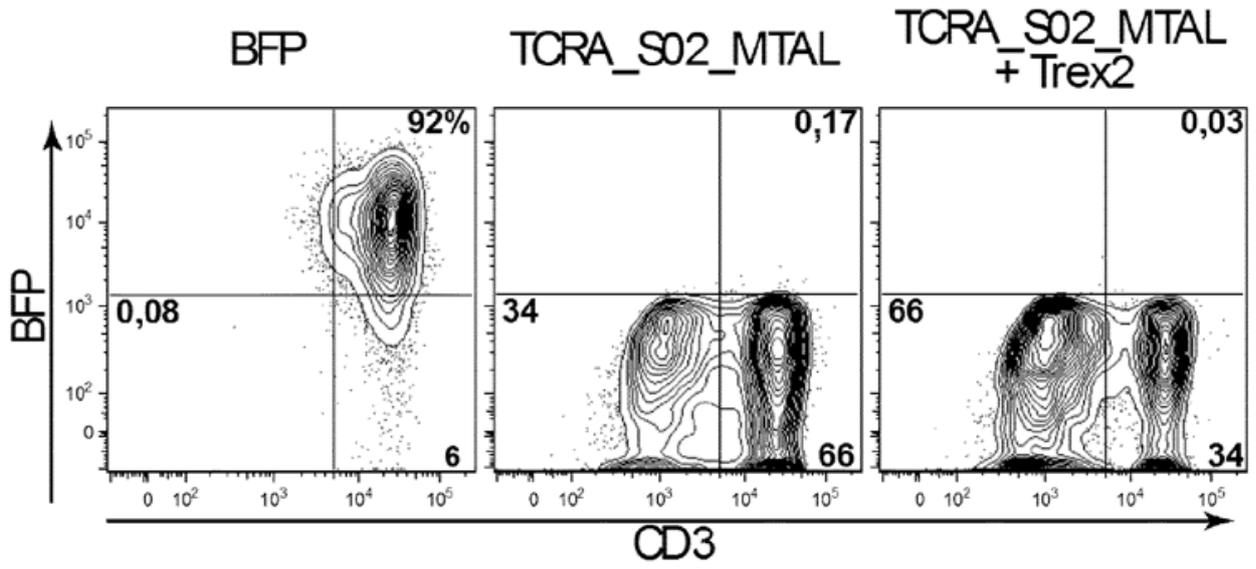


Figura 14

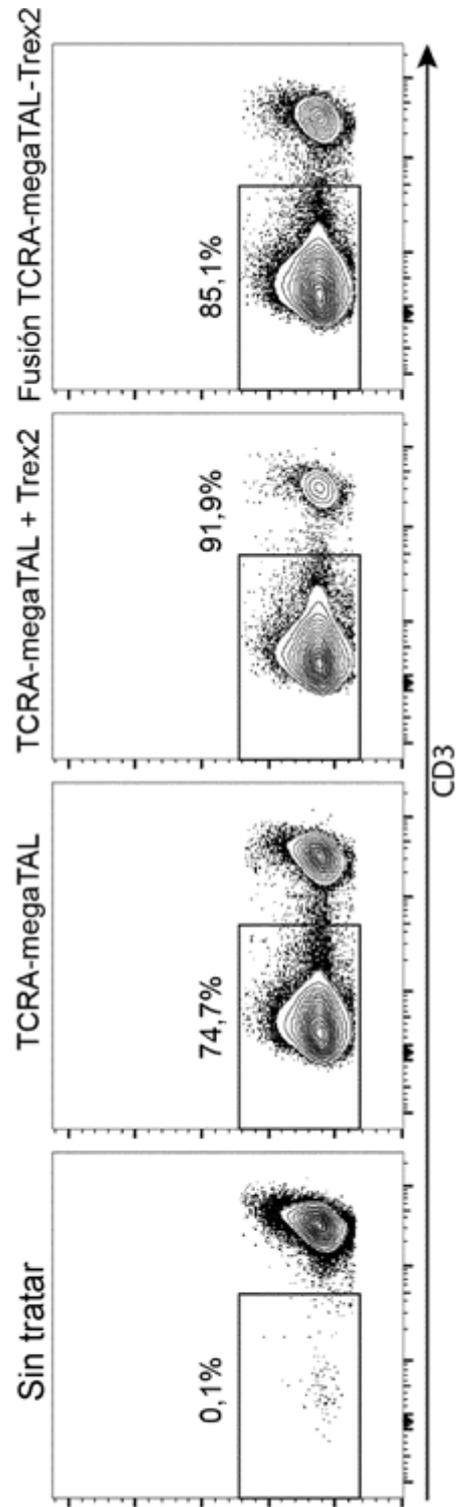


Figura 15