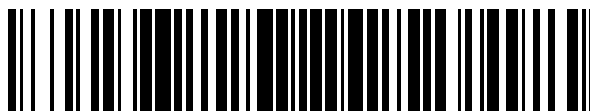


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 777 534**

51 Int. Cl.:

C12N 9/22 (2006.01)
C07K 14/005 (2006.01)
C12N 15/62 (2006.01)
A61K 48/00 (2006.01)
C12N 15/867 (2006.01)
C07K 19/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.02.2016** **PCT/EP2016/053008**

87 Fecha y número de publicación internacional: **18.08.2016** **WO16128549**

96 Fecha de presentación y número de la solicitud europea: **12.02.2016** **E 16704213 (4)**

97 Fecha y número de publicación de la concesión europea: **12.02.2020** **EP 3256578**

54 Título: **Polipéptidos para la ingeniería de proteínas integrasas quiméricas y su uso en terapia génica**

30 Prioridad:

13.02.2015 EP 15305217

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.08.2020

73 Titular/es:

**INSERM - INSTITUT NATIONAL DE LA SANTÉ ET
DE LA RECHERCHE MÉDICALE (20.0%)
101, rue de Tolbiac
75013 Paris, FR;
CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (C.N.R.S.) (20.0%);
UNIVERSITÉ DE PARIS (20.0%);
COMMISSARIAT À L'ÉNERGIE ATOMIQUE ET
AUX ÉNERGIES ALTERNATIVES (20.0%) y
CNAM CONSERVATOIRE NATIONAL DES ARTS
DES MÉTIERS (20.0%)**

72 Inventor/es:

**BRIDIER-NAHMIAS, ANTOINE;
WERNER, MICHEL y
LESAGE, PASCALE**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 777 534 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

Polipéptidos para la ingeniería de proteínas integrasas quiméricas y su uso en terapia génica

Campo:

La presente descripción se relaciona con polipéptidos para la ingeniería de proteínas integrasas quiméricas y su uso en terapia génica.

Antecedentes:

La capacidad de introducir una secuencia génica extraña o nativa particular en una célula y de controlar la expresión de ese gen tiene valor en los campos de la medicina y de la investigación biológica. Se puede facilitar la introducción de un gen extraño o nativo particular en una célula huésped introduciendo una secuencia génica en un vector de ácido nucleico adecuado. Se ha desarrollado una variedad de métodos que permiten la introducción de dicho vector recombinante en una célula huésped deseada. El uso de vectores víricos puede dar lugar a la rápida introducción de la molécula recombinante en una amplia variedad de células huésped. Por ejemplo, los retrovirus son virus de ARN que se replican a través de un intermediario provírico de ADN que normalmente se integra en el genoma de la célula huésped infectada. Todos los retrovirus conocidos comparten características del ciclo de replicación, incluyendo el empaquetamiento del ARN vírico en viriones, la entrada en las células diana, la transcripción inversa del ARN vírico para formar el provirus de ADN y la integración estable del provirus en el genoma de la célula diana. Los provirus simples competentes en cuanto a replicación comprenden típicamente repeticiones terminales largas reguladoras (LTR) y los genes gag, pro, pol y env, que codifican proteínas nucleares (gag), una proteasa (pro), transcriptasa inversa (pol), ARNasa H (pol), integrasa (pol) y glicoproteínas de la envuelta (env). Los retrovirus complejos también comprenden típicamente genes accesorios adicionales. Los vectores retrovíricos son una herramienta común para administración de genes en cuanto a que la capacidad de los vectores retrovíricos para administrar un gen de una sola copia no reorganizado a una amplia gama de células los hace muy adecuados para transferir genes a una célula. Aunque los vectores retrovíricos recombinantes permiten la integración de un transgén en el genoma de una célula huésped, la mayoría de los retrovirus sólo pueden transducir células en división. Esto puede limitar su uso para la transferencia génica *in vivo* a células no proliferantes, tales como hepatocitos, miofibros, células madre hematopoyéticas (HSC) y neuronas. Las células que no se dividen son el tipo celular de larga vida predominante en el organismo, y representan los objetivos más deseables de transferencia génica, incluyendo hígado, músculo y cerebro. Los lentivirus son un subgrupo de retrovirus que son capaces de infectar a células que no se dividen. Estos virus incluyen, aunque sin limitación, VIH-1, VIH-2, SIV, EIAV1 y FIV. Los lentivirus poseen genes gag, pol y env además de otros genes accesorios que están flanqueados por dos secuencias de repeticiones terminales largas (LTR). Un reto clave para la transferencia génica basada en el uso de vectores retrovíricos es conseguir una expresión estable de transgenes minimizando al mismo tiempo la mutagénesis de inserción y la inducción de una respuesta de daños en el ADN debido a la presencia de ADN de doble hélice. Un enfoque para evitar la mutagénesis de inserción es dirigir la integración del transgén a una localización específica en el genoma.

Ya se han descrito algunos mutantes de la Ty1 integrasa (Kenna, M. A., Brachmann, C. B., Devine, S. E., y Boeke, J. D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition *in vivo*. Molecular and Cellular Biology, 18(2), 1115-1124).

Además, Delelis O. *et al.* describen la integrasa del VIH (Delelis, O., Carayon, K., Saïb, A., Deprez, E., y Mouscadet, J. F. (2008). Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology, 5(1), 114).

Compendio de la presente invención:

Como se define en las reivindicaciones, la presente invención se relaciona con:

- un polipéptido que consiste en la secuencia de aminoácidos seleccionada del grupo consistente en la SEQ ID NO:2 a la SEQ ID NO:12,
- una integrasa quimérica que comprende el polipéptido de la presente invención y que es seleccionada del grupo consistente en integrasas del virus de la inmunodeficiencia felina (FIV), integrasas del virus Espumoso (FV), integrasas del virus de la leucemia murina (MLV) e integrasas del lentivirus,
- una molécula de ácido nucleico que codifica el polipéptido de la presente invención o la integrasa quimérica de la presente invención,
- un vector retrovírico que comprende la integrasa quimérica de la presente invención,
- un método *in vitro* para expresar un transgén en el genoma de una pluralidad de células que comprende infectar la pluralidad de células con el vector retrovírico de la presente invención, y
- el vector retrovírico de la presente invención para uso en terapia génica.

Descripción detallada:

Los elementos genéticos móviles son ubicuos. Su sitio de integración influye en la estabilidad del genoma y la expresión génica. El retrotransposón Ty1 de la levadura *Saccharomyces cerevisiae* se integra secuenciada arriba de los genes transcritos con ARN polimerasa III (Pol III), pero el determinante primario de especificidad diana ha permanecido evasivo. Aquí, los inventores describen una interacción entre la integrasa Ty1 y la subunidad AC40 de Pol III, y demuestran que AC40 es el determinante predominante de la dirección de la integración de Ty1 secuenciada arriba de los genes transcritos con Pol III. La falta de una interacción integrasa/AC40 altera de manera dramática la elección del sitio diana, para dar lugar a una redistribución de inserciones de Ty1 en el genoma, principalmente en los extremos cromosómicos. El mecanismo de especificidad diana para genes individualmente no esenciales permite a Ty1 proliferar y aún minimiza el daño genético a su huésped. Por consiguiente, el dominio de Ty1 responsable de la interacción con la ARN polimerasa III podría, por lo tanto, ser adecuado para la ingeniería de la integrasa de retrovirus para dirigir la integración dirigida de un transgén en un genoma eucariótico.

Tal como se usa aquí, el término "integrasa Ty1" o "Ty1 IN" tiene su significado general en la técnica y se refiere a la integrasa codificada del retroelemento Ty1. Ty1, el retrotransposón de LTR más activo y abundante de *S. cerevisiae*, preferiblemente se integra en una ventana de 1 kb secuenciada arriba de los genes transcritos con Pol III. Ty1 IN tiene dos regiones filogenéticamente conservadas (un dominio de unión a zinc N-terminal y un núcleo catalítico con la unidad DDX35E) y un extremo C menos conservado que contiene una señal de localización nuclear bipartita necesaria para la localización nuclear de IN y una retrotransposición eficiente (Figura 1A) (M. Kenna, C. Brachmann, S. Devine, J. Boeke, Mol Cell Biol 18, 1115-1124 (1998).; 9. S. P. Moore, L. A. Rinckel, D. J. Garfinkel, Mol Cell Biol 18, 1105-1114 (1998)). Se representa una secuencia ejemplar de aminoácidos de integrasa Ty1 mediante la SEQ ID NO:1:

SEQ ID NO:1: (Número de acceso Genbank: CAA88158-1)

```
NVHTSESTRKYPYPFIHRLAHANAQTIRYSLKNNTITYFNESDVDWSSA 50
IDYQCPDCLIGKSTKHRHIKGSRLKYQNSYEPFQYLHTDIFGPVHNLPS 100
APSYFISFTDETTFKFRWVYPLHDRREDSILDVFTTILAFIKNQFQASVLV 150
IQMDRGSEYTNRTLHKFLEKNGITPCYTTTADSRAGVAERLNRTLDDC 200
RTQLQCSGLPNHLWFS AIEFSTIVRNSLASPKSKKSARQHAGLAGLDIST 250
LLPFGQPVIVNDHNPNISKIHPRGIPGYALHPSRNSYGYIIYLP SLKKTVD 300
TTNYVILQGKESRLDQFN YDALTFDEDLNRLTASYHSFIASNEIQESNDL 350
NIESDHDFQSDIELHPEQPRNVLSKAVSPTDSTPPSTHTEDSKRVSKTNI 400
RAPREVDPNISESNILPSKKRSSTPQISNIESTGSGGMHKLNVLLAPMS 450
QSNTHES SHASKSKDFRHSDSYSENETNHTNVPISSTGGTNNKTV PQISD 500
QETEKRIIHRSPSIDASPPENNSSHNIVPIKTPTTVSEQNTEESIIADLP 550
LPDLPPESPTEFPDPFKELPPINSRQTNSSLGGIGDSNAYTTINSKKRSL 600
EDNETEIKVSRDWTNT KNMRSLEPPRSKKRIHLIA 635
```

Un primer objeto de la presente descripción se relaciona, por lo tanto, con un polipéptido que comprende la secuencia de aminoácidos que va del residuo de aminoácido en la posición 617 al residuo de aminoácido en la posición 622 en la SEQ ID NO:1 (KNMRSLE = SEQ ID NO:2) o una función conservadora del mismo.

Tal como se usa aquí, el término "variantes de función conservadora" corresponde a aquellas en las que se ha cambiado un residuo de aminoácido dado en el polipéptido sin alterar la conformación y función globales (es decir, interacción con AC40) del polipéptido, incluyendo, aunque sin limitación, el reemplazo de un aminoácido por uno que tiene propiedades similares (sustitución conservadora) o el reemplazo de un aminoácido por uno que tiene propiedades diferentes (sustitución no conservadora). Los aminoácidos distintos de los indicados como conservados pueden diferir en el polipéptido, de tal forma que el porcentaje de identidad entre cualesquiera dos polipéptidos de función similar puede variar, y puede ser, por ejemplo, del 70% al 99%, según se determina de acuerdo con un esquema de alineación, tal como mediante el Método de Grupos, en donde la similitud se basa en el algoritmo MEGALIGN. Una "variante de función conservadora" también incluye un polipéptido que tiene una secuencia de aminoácidos que tiene al menos un 50% de identidad con la secuencia que va desde el residuo de aminoácido en la posición 617 hasta el residuo de aminoácido en la posición 622 en la SEQ ID NO: 1. En particular, la variante de función conservadora comprende una secuencia de aminoácidos que tiene un 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% de identidad con la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 617 hasta el residuo de aminoácido en la posición 622 en la SEQ ID NO:1.

Tal como se usa aquí, un "residuo de aminoácido" se refiere a cualquier aminoácido natural, cualquier derivado de aminoácido o cualquier imitador de aminoácido conocido en la técnica. En particular, los residuos de la proteína o del péptido son secuenciales, sin que ningún aminoácido interrumpa la secuencia de residuos de aminoácido. En otras realizaciones, la secuencia puede comprender uno o más restos no aminoácido. En realizaciones particulares, la secuencia de residuos de la proteína o péptido puede estar interrumpida por uno o más restos no aminoácido. Por consiguiente, el término proteína o péptido abarca secuencias de aminoácidos que comprenden al menos uno de los 20 aminoácidos comunes que se encuentran en las proteínas naturales, o al menos un aminoácido modificado o inusual, incluyendo, aunque sin limitación, ácido 2-Aminoadípico (Aad), N-Etilasparragina (EtAsn), ácido 3-Aminoadípico (Baad), Hidroxilisina (Hil), β alanina, ácido β Aminopropiónico (Bala), alo Hidroxilisina (AHil), ácido 2-Aminobutírico (Abu), 3-Hidroxiprolina (3Hip), ácido 4-Aminobutírico (4Abu), 4-Hidroxiprolina (4Hip), ácido 6-Aminocaproico (Acp), Isodesmosina (Ide), ácido 2-Aminoheptanoico (Ahe), alo Isoleucina (Alle), ácido 2-Aminoisobutírico (Aib), N-Metilglicina (MeGly), ácido 3-Aminoisobutírico (Baib), N-Metilisoleucina (Melle), ácido 2-Aminopimélico (Apm), 6-N-Metilisina (MeLys), ácido 2,4-Diaminobutírico (Dbu), N-Metilvalina (MeVal), Desmosina (Des), Norvalina (Nva), ácido 2,2'-Diaminopimélico (Dpm), Norleucina (Nle), ácido 2,3-Diaminopropiónico (Dpr), Ornitina (Orn) o N-Etilglicina (EtGly).

En particular, el residuo de aminoácido en la posición 617, 619, 621 ó 622 está sustituido de manera conservadora. En el contexto de la presente descripción, una "sustitución conservadora" se define por sustituciones en las clases de aminoácidos reflejadas como sigue:

Residuos alifáticos I, L, V y M

Residuos asociados a cicloalqueno F, H, W e Y

Residuos hidrofóbicos A, C, F, G, H, I, L, M, R, T, V, W e Y

Residuos cargados negativamente D y E

Residuos polares C, D, E, H, K, N, Q, R, S y T

Residuos cargados positivamente H, K y R

Residuos pequeños A, C, D, G, N, P, S, T y V

Residuos muy pequeños A, G y S

Residuos implicados en el giro A, C, D, E, G, H, K, N, Q, R, S, P y formación T

Residuos flexibles Q, T, K, S, G, P, D, E y R

Como agrupaciones de sustituciones más conservadoras, se incluyen: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparragina-glutamina. También se retiene sustancialmente conservación en términos de propiedades hidropáticas/hidrofílicas y peso/tamaño de residuos en el polipéptido de la presente descripción en comparación con la secuencia nativa. La importancia del índice de aminoácidos hidropáticos al conferir una función biológica interactiva a una proteína generalmente se entiende en la técnica. Se acepta que el carácter hidropático relativo del aminoácido contribuye a la estructura secundaria de la proteína resultante, la cual a su vez define la interacción de la proteína con otras moléculas, por ejemplo, enzimas, sustratos, receptores, ADN, anticuerpos, antígenos y similares. Se ha asignado a cada aminoácido un índice hidropático en base a su hidrofobicidad y características de carga; éstos son: isoleucina (+4,5), valina (+4,2), leucina (+3,8), fenilalanina (+2,8), cisteína/cistina (+2,5), metionina (+1,9), alanina (+1,8), glicina (-0,4), treonina (-0,7), serina (-0,8), triptófano (-0,9), tirosina (-1,3), prolina (-1,6), histidina (-3,2), glutamato (-3,5), glutamina (-3,5), aspartato (-3,5), asparragina (-3,5), lisina (-3,9) y arginina (-4,5). También se puede medir, o medir de manera alternativa, la retención de residuos similares mediante una puntuación de similitud, según se determina mediante el uso de un programa BLAST (por ej., BLAST 2,2,8, que se puede adquirir a través del NCBI usando ajustes estándar BLOSUM62, Hueco Abierto= 11 y Hueco Ampliado= 1).

En particular, el aminoácido en la posición 618 ó 620 está sustituido por cualquier aminoácido (es decir, una sustitución no conservadora). En particular, el residuo de aminoácido en la posición 618 está sustituido por una alanina (A). En particular, el residuo de aminoácido en la posición 620 está sustituido por un residuo de ácido aspártico (D).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:3 (KEMDSL).

En particular, el polipéptido de la presente descripción comprende 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179,

180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199 ó 200 residuos de aminoácido. En particular, el polipéptido de la presente descripción comprende 7, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 ó 60 residuos de aminoácido consecutivos en la SEQ ID NO:1.

- 5 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 617 hasta el residuo de aminoácido en la posición 623 en la SEQ ID NO:1 (KNMRSLE = SEQ ID NO: 4).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:5 (KAMRSLE).

- 10 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:6 (KNMRSLEA).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 617 hasta el residuo de aminoácido en la posición 625 (KNMRSLEPP = SEQ ID NO:7).

- 15 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 614 hasta el residuo de aminoácido en la posición 622 en la SEQ ID NO:1 (WNTKNMRSLE = SEQ ID NO: 17).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 609 hasta el residuo de aminoácido en la posición 623 (VSRDTWNTKNMRSLE = SEQ ID NO:8).

- 20 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 600 al residuo de aminoácido en la posición 625 (SLEDNETEIKVSRDTWNTKNMRSLEPP = SEQ ID NO: 9).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 595 hasta el residuo de aminoácido en la posición 630 (SKKRSLEDNETEIKVSRDTWNTKNMRSLEPPRSKKR = SEQ ID NO:10).

- 25 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 595 hasta el residuo de aminoácido en la posición 635 (SKKRSLEDNETEIKVSRDTWNTKNMRSLEPPRSKKRIHLIA = SEQ ID NO: 11).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 578 hasta el residuo de aminoácido en la posición 635 (NSSLGGIGDSNAYTTINSK KRSLEDNETEIKVSRDTWNTKNMRSLEPPRSKKRIHLIA = SEQ ID NO:12).

- 30 En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos que va desde el residuo de aminoácido en la posición 578 hasta el residuo de aminoácido en la posición 635 (NSSLGGIGDSNAYTTINSK KRSLEDNETEIKVSRDTWNTKNMRSLEPPRSKKRIHLIA = SEQ ID NO:12).

En particular, el polipéptido de la presente descripción comprende la secuencia de aminoácidos seleccionada del grupo consistente en la SEQ ID NO:2 a la SEQ ID NO:12 y secuencias de aminoácidos que tienen al menos un 50% de identidad con la SEQ ID NO:2 a la SEQ ID NO:12.

- 35 En particular, el polipéptido de la presente descripción está fusionado a un polipéptido heterólogo (es decir, un polipéptido que no deriva de la secuencia de aminoácidos nativa de la integrasa Ty1) para formar una proteína de fusión. Tal como se usa aquí, una "proteína de fusión" comprende todo o parte (típicamente biológicamente activa) de un polipéptido de la presente descripción operativamente unido a un polipéptido heterólogo (es decir, un polipéptido distinto del mismo polipéptido). En la proteína de fusión, el término "operativamente unido" pretende indicar que el polipéptido de la presente descripción y el polipéptido heterólogo están fusionados en marco entre sí. El polipéptido heterólogo puede fusionarse al extremo N o al extremo C del polipéptido de la presente descripción. En alguna realización, el polipéptido heterólogo se fusiona al extremo C-terminal del polipéptido de la presente descripción. En particular, el polipéptido de la presente descripción y el polipéptido heterólogo se fusionan entre sí directamente (es decir, sin el uso de un conector) o mediante un conector. El conector es típicamente un péptido conector y, en la presente descripción, será seleccionado de manera que permita la unión del polipéptido al polipéptido heterólogo. Los conectores adecuados estarán claros para el experto en la técnica en base a la descripción que aquí se da, opcionalmente tras algún grado limitado de experimentación rutinaria. Se describen aquí conectores adecuados, y pueden - por ejemplo y sin limitación - comprender una secuencia de aminoácidos, la cual secuencia de aminoácidos preferiblemente tiene una longitud de 2 o más aminoácidos. Típicamente, el conector tiene 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 ó 30 aminoácidos. Sin embargo, el límite superior no es crítico, pero se selecciona por razones de conveniencia en cuanto a, por ej., la producción biofarmacéutica de dichas proteínas de fusión. La secuencia del conector puede ser una secuencia natural o una secuencia no natural. Si se usa con fines terapéuticos, el ligante es preferiblemente no inmunogénico en el sujeto al que se administra la proteína de fusión de la presente descripción. Un grupo útil de secuencias conectoras son los ligantes derivados de la región de bisagra de anticuerpos de cadenas pesadas, como se describe en WO 96/34103 y WO 94/04678. Otros

ejemplos son secuencias conectoras de polialanina, tales como Ala-Ala-Ala. Otros ejemplos preferidos de secuencias ligantes son conectores Gly/Ser de diferente longitud, incluyendo (gly4ser)³, (gly4ser)⁴, (gly4ser), (gly3ser), gly3 y (gly3ser)²₃.

Otro aspecto de la presente descripción se relaciona con una integrasa quimérica, en donde una integrasa retroviral comprende un polipéptido de la presente descripción. Según la descripción, la integrasa quimérica de la presente descripción es particularmente adecuada para dirigir la integración lejos de los sitios genómicos que pueden resultar perjudiciales para la célula o el organismo que alberga la célula. En particular, se somete así a ingeniería una integrasa recombinante (IN) para que contenga un polipéptido de la presente descripción.

Tal como se usa aquí, el término "integrasa retroviral" tiene su significado general en la técnica y se refiere a una enzima producida por un retrovirus (tal como el VIH) que permite que su material genético se integre en el ADN de la célula infectada. La principal función de IN es insertar el ADN vírico en el ADN cromosómico del huésped, una etapa que es esencial para la replicación del VIH. La integración es un punto de no retorno para la célula, que se convierte en un portador permanente del genoma vírico (provirus). Todas las proteínas IN retrovirales tienen tres dominios físicamente distintos. (1) Un dominio N-terminal que incluye tres hélices y una unidad de unión a zinc. Se ha implicado a este dominio en la dimerización y en la unión de los extremos LTR. (2) El dominio central que contiene la tríada catalítica conservada DDE. (3) El dominio C-terminal que contribuye a la oligomerización y tiene actividad de unión a ADN no específica. El término "retrovirus" incluye, aunque sin limitación, los miembros de la familia retroviridae, incluyendo alfa-retrovirus (por ej., virus de la leucosis aviar), beta-retrovirus (por ej., virus del tumor mamario del ratón), gamma-retrovirus (por ej., virus de la leucemia murina), delta-retrovirus (por ej., virus de la leucemia bovina), épsilon-retrovirus (por ej., virus del sarcoma dérmico de Walley), lentivirus (por ej., VIH-1, VIH-2) y espumavirus (por ej., espumavirus humano). El término "lentivirus", tal como se usa aquí, se refiere al virus de la inmunodeficiencia humana 1 (VIH-1); al virus de la inmunodeficiencia humana 2 (VIH-2); al virus de la inmunodeficiencia de simios (SIV) y al virus de la inmunodeficiencia felina (FIV).

En particular, la integrasa retroviral es una integrasa del virus de la inmunodeficiencia felina (FIV), una integrasa del virus Espumoso (FV), una integrasa del virus de la leucemia murina (MLV), una integrasa de lentivirus u otra integrasa víricamente codificada, o derivado de la misma.

En particular, la integrasa retroviral tiene una secuencia que tiene al menos un 80% de identidad con la SEQ ID NO: 13 o la SEQ ID NO:14.

SEQ ID NO:13 >IN del VIH-1 (Nº Acc. M19921 o K03455:

FLDGIDKAQD EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM
HGQVDCSPGI WQLDCTHLEG KVILVAVHVA SGYIEAEVIP AETGQETAYF
LLKLAGRWPV KTIHTDNGSN FTGATVRAAC WWAGIKQEFQ IPYNPQSQGV
VESMNKELKK IIGQVRDQAE HLKTAVQMAV FIHNFKRKGG IGGYSAGERI
VDIIATDIQT KELQKQITKI QNFRVYYRDS RNPLWKGPAP LLWKGEAVV
IQDNSDIKVV PRRKAKIIRD YGKQMGDDC VASRQDED

SEQ ID NO:14 > IN del MLV (Acc. NP_955592)

ENSSPYTSEH FHYTVTDIKD LTKLGAIDYK TKKYWVYQGK PVMPDQFTFE
LLDFLHLQTH LSFSSMKALL ERSHPYMYL NRDRTLKNIT ETCKACAQVN
ASKSAVKQGT RVRGHRPGTH WEIDFTEIKP GLYGYKYLLV FIDTFSGWIE
AFPTKKETAK VVTKKLLLEI FPRFGMPQVL GTDNGPAFVS KVSQTVADLL
GIDWKLHCAY RPQSSGQVER MNRTIKETLT KLTATGSRD WVLLPLALY
RARNTPGPHG LTPYEILYGA PPPLVNFDP DMTRVTNSPS LQAHLQALYL
VQHEVWRPLA AAYQEQLDRP VVPHYRVGD TVWVRRHQTK
NLEPRWGPY TVLLTPTAL KVDGIAAWIH AAHVKAADPG GGPSSRLTWR
VQRSQNPLKI RLTREAP

En particular, la proteína integrasa está mutada de manera que su interacción con su factor de unión afín queda abolida. En particular, la integrasa de la SEQ ID NO: 13 comprende al menos una mutación seleccionada del grupo consistente en V165A, R166A, L172A/K173A, Q168A y Q168L. En particular, la integrasa de la SEQ ID NO: 14 comprende una mutación en la posición W390P398, L399, K400, R402, y en particular comprende la mutación W390A (Mol Ther Nucleic Acids. 2014 Jul; 3(7): e179.).

En particular, el polipéptido de la presente descripción está fusionado en el extremo C-terminal de la proteína integrasa.
En particular, el polipéptido de la presente descripción está fusionado en el extremo N-terminal de la proteína integrasa.
En particular, el polipéptido de la presente descripción está insertado en la secuencia codificante de la proteína integrasa.

- 5 En particular, el polipéptido de la presente descripción está insertado entre el residuo de aminoácido en la posición 5 y el residuo de aminoácido en la posición 6 en la SEQ ID NO: 13.

- 10 En particular, el polipéptido de la presente descripción el residuo de aminoácido en la posición 390 en la SEQ ID NO:3 está sustituido por el polipéptido de la presente descripción. En particular, el polipéptido de la presente descripción está insertado entre el residuo de aminoácido en la posición 389 y el residuo de aminoácido en la posición 390 en la SEQ ID NO:14, en donde dicho residuo de aminoácido en la posición 390 está sustituido (por ej., W390A).

Otro objeto de la presente descripción se relaciona con una molécula de ácido nucleico que codifica un polipéptido de la presente descripción o una integrasa quimérica de la presente descripción.

- 15 Tal como se usa aquí, el término "molécula de ácido nucleico" tiene su significado general en la técnica y se refiere a una molécula de ADN o ARN. Sin embargo, el término captura secuencias que incluyen cualquiera de los análogos de base conocidos de ADN y ARN, tales como, aunque sin limitación 4-acetilcitosina, 8-hidroxi-N6-metiladenosina, aziridinilcitosina, pseudoisocitosina, 5-(carboxihidroxi)metiluracilo, 5-fluorouracilo, 5-bromouracilo, 5-carboximetilaminometil-2-tiouracilo, 5-carboximetilaminometiluracilo, dihidrouracilo, inosina, N6-isopenteniladenina, 1-metiladenina, 1-metilpseudouracilo, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-metiladenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxycarbonilmetiluracilo, 5-metoxiuracilo, 2-metil-N6-isopenteniladenina, éster metílico del ácido uracil-5-oxiacético, ácido uracil-5-oxiacético, oxibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido -uracil-5-oxiacético, ácido uracil-5-oxiacético, pseudouracilo, queosina, 2-tiocitosina y 2,6-diaminopurina.

- 25 En particular, la integrasa quimérica de la presente descripción está empaquetada en un vector, más preferiblemente un vector retroviral. Por consiguiente, otro objeto de la presente descripción se relaciona con un vector retroviral (por ej., un vector lentiviral) que comprende una integrasa quimérica de la presente descripción.

El término "vector retroviral" se refiere a un vector que contiene elementos genéticos estructurales y funcionales que derivan primariamente de un retrovirus.

- 30 En particular, el vector retroviral de la presente descripción deriva de un retrovirus seleccionado del grupo consistente en alfa-retrovirus (por ej., virus de la leucosis aviar), beta-retrovirus (por ej., virus del tumor mamario del ratón), gamma-retrovirus (por ej., virus de la leucemia murina), delta-retrovirus (por ej., virus de la leucemia bovina), épsilon-retrovirus (por ej., virus del sarcoma dérmico de Walley), lentivirus (por ej., VIH-1, VIH-2) y espumavirus (por ej., espumavirus humano).

- 35 En particular, el vector retroviral de la presente descripción es una partícula vírica retroviral deficiente en replicación, que puede transferir un ARN importado extraño de un gen o un fragmento del mismo o un gen indicador, por ej., un gen terapéutico, en lugar del ARNm retroviral.

- 40 En particular, el vector retroviral de la presente descripción es un vector lentiviral. El término "vector lentiviral" se refiere a un vector que contiene elementos genéticos estructurales y funcionales fuera de los LTR que derivan primariamente de un lentivirus. En particular, se selecciona el vector lentiviral de la presente descripción del grupo consistente en los vectores VIH-1, VIH-2, SIV, FIV, EIAV, BIV, VISNA y CAEV. En particular, el vector lentiviral es un vector VIH-1.

- 45 En particular, la proteína de la envuelta del vector retroviral de la presente descripción está pseudotipificada con la proteína de la envuelta del retrovirus usado para preparar el vector retroviral, o de manera alternativa con una proteína de la envuelta heterogénea que se elige con respecto a las células que se han de dirigir al huésped. En particular, el vector retroviral de la presente descripción está pseudotipificado con una proteína VSV-G. La glicoproteína VSV-G puede originarse de diferentes serotipos del género de los vesiculovirus: serotipo VSV-Indiana, serotipo VSV-New Jersey u otras glicoproteínas de los vesiculovirus, tales como Piry, Chandipura, Isfahan y Cocal. Se escoge la glicoproteína VSV-G entre especies clasificadas en el género vesiculovirus: virus Carajas (CJSV), virus Chandipura (CHPV), virus Cocal (COCV), virus Isfahan (ISFV), virus Maraba (MARAV), virus Piry (PIRYV), virus Alagoas de la estomatitis vesicular (VSAV), virus Indiana de la estomatitis vesicular (VSIV) y virus New Jersey de la estomatitis vesicular (VSNJV) y/o rhabdovirus de la carpa Herbívora, virus BeAn 157575 (BeAn 157575), virus Boteke (BTKV), virus Calchaqui (CQIV), virus Eel Americano (EVA), virus Gray Lodge (GLOV), virus Jurona (JURV), virus Klamath (KLAV), virus Kwatta (KWAV), virus La Joya (LJV), virus Malpais Spring (MSPV), virus del murciélago Mount Elgon (MEBV), virus Perinet (PERV), rhabdovirus de alevines de Lucio (PFRV), virus Porton (PORV), virus Radi (RADIV), virus de la viremia Primavera de la carpa (SVCV), virus Tupaia (TUPV), rhabdovirus de la enfermedad ulcerativa (UDRV) y virus Yug Bogdanovac (YBV). En particular, la proteína VSV-G que se origina de un VSV está modificada con respecto a su forma nativa, especialmente para mejorar la pseudotipificación. En particular, la proteína de la envuelta comprende dominios o fragmentos que se originan a partir de diferente(s) proteína(s) de la envuelta de

diferentes virus, especialmente de diferentes géneros de diferentes especies de VSV. En el caso del VSV, la proteína G comprende o consiste en el dominio de membrana del VSV indiano y el ectodominio de una cepa de un serotipo de VSV diferente. En particular, la(s) proteína(s) de la envuelta comprende(n) el dominio de membrana del VSV indiano y el ectodominio del VSV New-Jersey. En otro aspecto, el vector retroviral de la presente descripción está pseudotipificado con proteína HA (hemaglutinina de la influenza), proteína RD114, envueltas modificadas con ligandos específicos de células insertados o con proteínas de la envuelta vírica originadas de un virus seleccionado en uno o varios de los siguientes órdenes o familias: Arenaviridae, Flaviridae, Togaviridae, Coronaviridae, Orthomyxoviridae, Retroviridae y Mononegavirales, incluyendo Paramyxoviridae, Rhabdoviridae o Filoviridae. Cualquier proteína de envuelta vírica es adecuada para la pseudotipificación, suponiendo que la pseudotipificación sea compatible con las etapas de producción y purificación de partículas de vectores lentivirales.

La estructura y composición del genoma del vector utilizado para preparar los vectores retrovirales de la presente descripción están de acuerdo con las descritas en la técnica. Especialmente, se pueden preparar vectores de administración de genes retrovirales mínimos a partir del genoma de un vector, que sólo contenga, aparte del(de los) polinucleótido(s) heterólogo(s) de interés (es decir, el(los) transgén(es)), las secuencias del genoma retroviral que son regiones no codificantes de dicho genoma, necesarias para proporcionar señales de reconocimiento para la síntesis y el procesamiento de ADN o ARN. Por ello, el genoma de un vector puede ser un vector de reemplazo en el que todas las secuencias codificantes víricas entre las 2 repeticiones terminales largas (LTR) han sido reemplazadas por el(los) polinucleótido(s) de interés (es decir, el transgén). En particular, el genoma del vector retroviral también comprende, además, un polinucleótido consistente en el colgajo de ADN. El colgajo de ADN (definido en Zennou V. *et al.* 2000, Cell vol 101, 173-185, o en WO 99/55892 y WO 01/27304), es una estructura que es central en el genoma de algunos lentivirus, especialmente en retrovirus VIH, en donde da lugar a una estructura de ADN de 3 hélices normalmente sintetizada durante especialmente la transcripción inversa del VIH y que actúa como un determinante *cis* de la importación nuclear del genoma del VIH. El colgajo de ADN permite un suceso de desplazamiento de hebra central controlado en *cis* por el trazo de polipurina central (cPPT) y la secuencia de finalización central (CTS) durante la transcripción inversa. Cuando se inserta en vectores derivados de retrovirus, el polinucleótido que permite la producción del colgajo de ADN durante la retrotranscripción estimula la eficiencia de la transferencia génica y complementa el nivel de importación nuclear hasta los niveles del tipo salvaje (Arhel N. *et al.*, Retrovirology 2006, 3:38, 26 de junio de 2006, Wild-type and central DNA flap defective HIV-1 retroviral vector genomes: intracellular visualization at ultra structural resolution levels). En particular, se inserta el colgajo de ADN inmediatamente secuencia arriba del(de los) transgén(es), ventajosamente para tener una posición central en el genoma del vector. Se puede obtener un colgajo de ADN adecuado para la descripción a partir de un retrovirus, especialmente a partir de un lentivirus, o a partir de un organismo de tipo retrovirus, tal como un retrotransposón, ya sea preparado sintéticamente (síntesis química) o por amplificación del colgajo de ADN a partir de cualquier retrovirus, especialmente a partir del ácido nucleico de un lentivirus, tal como por Reacción en Cadena de Polimerasa (PCR). Se puede obtener el colgajo de ADN a partir de un retrovirus, especialmente un lentivirus, especialmente a partir de un retrovirus o lentivirus humano, y en particular un retrovirus VIH, o del virus CAEV (Virus de la Artritis y Encefalitis Caprina), el virus EIAV (Virus de la Anemia Infecciosa Equina), el virus VISNA, el virus SIV (Virus de la Inmunodeficiencia de Simios) o el virus FIV (Virus de la Inmunodeficiencia Felina). En particular, se obtiene el colgajo de ADN de un retrovirus VIH, por ejemplo virus VIH-1 o VIH-2, incluyendo cualquier aislado de estos dos tipos. Es de observar que el colgajo de ADN se usa como fragmento de ADN aislado de su contexto nucleotídico natural (genoma vírico), es decir, fuera del contexto del gen pol en donde está contenido de manera natural en el genoma del lentivirus. Por lo tanto, se usa el colgajo de ADN, en la presente descripción, con eliminación de las partes 5' y 3' innecesarias del gen pol y se recombina con secuencias de diferente origen. Se puede preparar el colgajo de ADN sintéticamente (síntesis química) o por amplificación del ADN que proporciona el colgajo de ADN a partir de la fuente apropiada como se ha definido anteriormente, tal como por Reacción en Cadena de Polimerasa (PCR). Un polinucleótido apropiado particular que comprende la estructura que proporciona el colgajo de ADN es un fragmento de reacción en cadena de polimerasa (PCR) de 178 pares de bases que abarca las regiones cPPT y CTS del ADN de VIH-1

El genoma del vector retroviral puede también comprender señales reguladoras para la transcripción y expresión de origen no retroviral, tales como un promotor y/o un potenciador. Son ejemplos de promotores que pueden usarse en la provocación de una respuesta inmune CMV, al que también se hace referencia como CMVie (CMV inmediato precoz), promotor EF1 α , promotor CGA, promotor CD11c y promotores de genes constitutivos, tales como el promotor PGK, el promotor de ubiquitina, el promotor de actina, el promotor de histona, el promotor de alfa-tubulina, el promotor de beta-tubulina, el promotor de superóxido dismutasa 1 (SOD-1), el promotor de dihidrofolato reductasa (DHFR), el promotor de hipoxantina fosforibosiltransferasa (HPRT), el promotor de adenosina desaminasa, el promotor de timidilato sintetasa, el promotor de dihidrofolato reductasa PI, el promotor de glucosa-6-fosfato deshidrogenasa o el promotor de nucleolina.

En particular, el transgén está bajo el control de señales reguladoras para la transcripción y la expresión.

En particular, el genoma del vector retroviral comprende todos los elementos necesarios para la importación nuclear y la correcta expresión del polinucleótido de interés (es decir, el transgén). Son ejemplos de elementos que pueden insertarse en el genoma retroviral del vector retroviral de la presente descripción al menos una (preferiblemente dos) repetición terminal larga (LTR), tal como una LTR5' y una LTR3', una secuencia psi implicada en la encapsidación del genoma retroviral, y opcionalmente al menos un colgajo de ADN que comprende un dominio cPPT y un dominio

CTS.

En particular, en la presente descripción, la LTR, preferiblemente la LTR3', está eliminada para el promotor y el potenciador de U3; se ha visto que esta modificación aumenta sustancialmente la transcripción del transgén insertado en el genoma retroviral (WO01/27304).

- 5 En particular, el genoma del vector retroviral puede también comprender elementos seleccionados entre un sitio donante de ajuste (SD), un sitio aceptor de ajuste (SA) y/o un elemento que responde a Rev (RRE).

En particular, el genoma del vector retroviral está desprovisto de genes retrovirales gag, pol y/o env funcionales. Por "funcional", se quiere decir un gen que se transcribe correctamente y/o se expresa correctamente. Por lo tanto, el genoma del vector retroviral de la presente descripción puede contener al menos uno de los genes gag, pol y env que está o bien sin transcribir o bien incompletamente transcrito; la expresión "incompletamente transcrito" se refiere a la alteración en los transcritos gag, gag-pro o gag-pro-pol, no transcribiéndose uno de éstos o varios de éstos. En particular, el genoma retroviral está desprovisto de genes retrovirales gag, pol y/o env.

En particular, el genoma del vector retroviral está también desprovisto de las secuencias codificantes de los genes accesorios Vif-, Vpr-, Vpu- y Nef (para vectores retrovirales VIH-1), o de sus genes completos o funcionales.

- 15 Típicamente, el vector retroviral de la presente descripción es no replicativo, es decir, que el vector y el genoma del vector retroviral no son capaces de formar nuevas partículas que gemen de la célula huésped infectada. Se puede conseguir esto mediante la ausencia en el genoma retroviral de los genes gag, pol o env, como se indicó en el párrafo anterior; también se puede conseguir esto eliminando otra(s) secuencia(s) codificante(s) viral(es) y/o elementos genéticos de acción cis necesarios para la formación de partículas.

- 20 En particular, el vector retroviral de la presente descripción comprende (i) un genoma recombinante que comprende, entre las secuencias LTR 5' y 3' retrovirales, un empaquetamiento retroviral de secuencia psi, un ARN de elemento de exportación nuclear, al menos un transgén y, opcionalmente, un promotor y/o una secuencia que favorece la importación nuclear del ARN, y (ii) una proteína integrasa quimérica de la presente descripción. En particular, el genoma recombinante comprende la secuencia 5' LTR-psi-RRE-cPPT-CTS-transgén(es)-3' LTR. En particular, el genoma recombinante comprende la secuencia 5' LTR-psi-RRE-cPPT-CTS-promotor-transgén(es)-3' LTR. En particular, el genoma recombinante comprende la secuencia 5' LTR-psi-RRE-cPPT-transgén(es)-CTS-3' LTR.

- 30 Tal como se usa aquí, el término "transgén" o "polinucleótido de interés" se refiere a cualquier ácido nucleico que se expresará en una célula de mamífero. Típicamente, el ácido nucleico es un ácido nucleico codificante o no codificante. Puede ser una secuencia no codificante, tal como, por ejemplo, una secuencia de reconocimiento de una enzima (integración específica de sitio, sitio con una afinidad particular por una proteína, etc.). Ésta es preferiblemente una secuencia codificante de un polipéptido o ARN dado activo como tal. Puede incluir un ADNc, un ADNg, un ADN sintético, un ARN, por ejemplo, un ARNs, una ribozima, etc., o una combinación de los mismos. Típicamente, el transgén es un ADN que comprende una secuencia codificante del producto de expresión deseado. El transgén puede incluir también una o más regiones de finalización de la transcripción, típicamente una señal de poliadenilación.

- 35 En particular, se puede seleccionar el transgén de un catalizador de ácido nucleico (interferente, antisentido, ribozima), un suicida de ácido nucleico (por ej., codificante de una toxina) o un ácido nucleico codificante de un péptido biológicamente activo, tal como un factor de crecimiento, un factor trófico, un factor antiangiogénico, una hormona, una citoquina, un anticuerpo, un receptor, un factor de diferenciación, un factor estimulante de colonias, un agente anticanceroso, una enzima, un neurotransmisor o su precursor, etc. En particular, el transgén codifica, por ej., factores tróficos, que incluyen: RdCVF, CNTF, NGF, NT3, NT4, FGF, PDGF, GDNF, etc. O factores antiangiogénicos o enzimas que restauran una actividad metabólica deficiente o que proporcionan una función metabólica particular, por ejemplo: TH, AADC, GTPC, β -glucuronidasa, etc.

- 40 En particular, el transgén codifica, por ejemplo, ARN de interferencia (ARNi) para inhibir específicamente la expresión de proteínas mutadas implicadas en una enfermedad o una enfermedad genética dominante causada por un aumento de función, tal como una enfermedad neurodegenerativa, tal como ELA (Esclerosis Lateral Amiotrófica) mutada), proteína APP, tau, presenilina o BACE (enfermedad de Alzheimer), la α -sinucleína (enfermedad de Parkinson) o Huntington (enfermedad de Huntington).

- En determinadas circunstancias, el transgén codifica una endonucleasa específica de sitio que proporciona una atenuación génica específica de sitio de la función génica. En particular, el transgén codifica una proteína con dedos de zinc (ZFP) que se une al sitio diana en una región de interés (por ej., un gen asociado a enfermedad, un gen altamente expresado, un gen de albúmina u otro gen o gen de puerto seguro) en un genoma, en donde la ZFP comprende uno o más dominios de unión a dedos de zinc de ingeniería. En particular, la ZFP es una nucleasa con dedos de zinc (ZFN) que escinde una región genómica diana de interés, en donde la ZFN comprende uno o más dominios de unión a dedos de zinc de ingeniería y un dominio de escisión o medio dominio de escisión de nucleasa. Se pueden obtener dominios de escisión y medios dominios de escisión, por ejemplo, a partir de diversas endonucleasas de restricción y/o endonucleasas de asentamiento. En particular, los medios dominios de escisión derivan de una endonucleasa de restricción de Tipo IIS (por ej., Fok I). En particular, el dominio con dedos de zinc reconoce un sitio diana en un gen asociado a enfermedad o de puerto seguro, tal como albúmina. En particular, el

transgén codifica una proteína TALE (de tipo activador de la Transcripción) que se une al sitio diana en una región de interés (por ej., un gen altamente expresado, un gen asociado a enfermedad o un gen de puerto seguro) en un genoma, en donde la TALE comprende uno o más dominios de unión a TALE de ingeniería. En particular, la TALE es una nucleasa (TALEN) que escinde una región genómica diana de interés, en donde la TALEN comprende uno o más dominios de unión al ADN de TALE de ingeniería y un dominio de escisión o medio dominio de escisión de nucleasa. Se pueden obtener dominios de escisión y medios dominios de escisión, por ejemplo, a partir de diversas endonucleasas de restricción y/o endonucleasas de asentamiento. En particular, los medios dominios de escisión derivan de una endonucleasa de restricción de Tipo IIS (por ej., Fok I). En particular, el dominio de unión al ADN de TALE reconoce un sitio diana en un gen altamente expresado, asociado a enfermedad o de puerto seguro. En particular, el transgén codifica un sistema CRISPR/Cas que se une a un sitio diana en una región de interés (por ej., un gen altamente expresado, un gen asociado a enfermedad o un gen de puerto seguro) en un genoma, en donde el sistema CRISPR/Cas comprende una nucleasa CRISPR/Cas y un ARNcr/ARNtracr de ingeniería (o ARN de una sola guía). En particular, el sistema CRISPR/Cas reconoce un sitio diana en un gen altamente expresado, asociado a enfermedad o de puerto seguro. El sistema ZFN, TALEN y/o CRISPR/Cas como aquí se describe puede unirse a y/o escindir la región de interés en una región codificante o no codificante en, o adyacente a, el gen, tal como, por ejemplo, una secuencia líder, una secuencia trailer o intrón, o en una región no transcrita, ya sea secuencia arriba o secuencia abajo de la región codificante. En particular, el sistema ZFN, TALEN y/o CRISPR/Cas se une a y/o escinde un gen altamente expresado, por ejemplo, un gen de globina en las células rojas de la sangre (RBC). Véase, por ej., la Solicitud Estadounidense Nº 61/670.451, titulada "Methods and Compositions for Delivery of Biologics", depositada el 11 de julio de 2012. En particular, el sistema ZFN, TALEN y/o CRISPR/Cas se une a y/o escinde un gen de puerto seguro, por ejemplo, un gen CCR5, un gen PPP1R12C (también conocido como AAV SI), albúmina, HPRT o un gen Rosa. Véanse, por ej., las Publicaciones de Patentes Estadounidenses Nº 20080299580, 20080159996 y 201000218264 y las Solicitudes Estadounidenses Nº 61/537.349, 61/560.506, 13/660.821 y la Solicitud Estadounidense Nº 61/670.451, titulada "Methods and Compositions for Regulation of Transgene Expression", depositada el 11 de julio de 2012. Además, para ayudar en la selección, se puede usar el locus HPRT (véanse las solicitudes de patente estadounidenses Nº de Serie 13/660.821 y 13/660.843). En particular, el sistema ZFN, TALEN y/o CRISPR/Cas puede unirse a y/o escindir un gen asociado a enfermedad (por ej., el gen codificante de la hidrolasa lisosómica α -galactosidasa A (AGA), relacionada con la Enfermedad de Fabry).

En particular, el transgén codifica un polipéptido antigénico. Se dice que un polipéptido es antigénico cuando su secuencia contiene al menos un péptido (epítipo) capaz de provocar una respuesta inmune cuando se pone en contacto con células presentadoras de antígeno (APC). Típicamente, el polipéptido antigénico comprende al menos un epítipo B, capaz de provocar una respuesta inmune humoral, particularmente una respuesta humoral protectora, o un epítipo T capaz de provocar una respuesta inmune celular. En particular, al menos un polipéptido está codificado por una secuencia nucleotídica que se origina del genoma de un patógeno, tal como un virus, especialmente un retrovirus, lentivirus, flavivirus o coronavirus, de una bacteria o de un parásito. En particular, el polipéptido antigénico de la presente descripción comprende o consiste en antígenos de superficie, tales como proteínas de la envuelta vírica u otras proteínas de membrana, y sus fragmentos, por ejemplo, envuelta de los virus del SIDA, incluyendo VIH-1 o VIH-2, o, por ejemplo, envuelta del Virus de la Fiebre Amarilla, el Virus West Nile, el virus del Dengue (DV), el virus de la encefalitis japonesa (JEV) o el coronavirus asociado al SARS. Otros polipéptidos víricos interesantes proceden de la cápside del VIH. De manera alternativa, el polipéptido antigénico deriva de un antígeno tumoral o un epítipo tumoral. Son polipéptidos particulares (o parte de los mismos) aquéllos que se expresan sobre la superficie celular de células tumorales. Estos polipéptidos (o parte de los mismos) pueden originarse a partir de la célula (autopéptido) en una forma de tipo salvaje o mutada; también pueden originarse a partir de un virus que transforma una célula normal en una célula tumoral (virus tumoral). Son ejemplos de tales virus, agentes etiológicos para el cáncer humano, el Virus del Papiloma Humano (HPV), que causa cáncer de cuello, el Virus de Epstein-Barr, que causa linfoma a través del antígeno de membrana inducido por EBV (EBMA), HTLV-1, que causa leucemia aguda de células T (ACT) a través de la proteína tax de HTLV-1, el virus del herpes humano de tipo 8 (HHV8), el virus de la hepatitis B (HBV) y el virus de la hepatitis C (HCV).

Se pone típicamente el transgén bajo el control de un promotor transcripcional, que puede ser homólogo al transgén - o promotor heterólogo, tal como uno celular, vírico, sintético, quimérico, etc. El promotor usado puede ser una ARN polimerasa 2 ó 3 dependiente constitutiva o regulada, débil o fuerte, específica de tejido o ubicua, etc. Típicamente utiliza un promotor vírico, tal como CMV, RSV LTR, TK, etc., o preferiblemente un promotor celular, tal como PGK, Rho, EF1 α , etc. Se pueden usar promotores específicos de tejido. Puede tratarse, por ejemplo, de los promotores ENO, GFAP, NSE, un promotor de ARN polimerasa III, tal como U6 o H1, posiblemente modificado, etc. El promotor usado para dirigir la expresión del transgén puede ser, por ejemplo, un promotor vírico seleccionado del promotor génico CMV, RSV LTR o TK.

Se pueden producir los vectores retrovíricos de la presente descripción por cualquier método bien conocido en la técnica, incluyendo por transfección(es) transitoria(s), en líneas celulares estables y/o por medio de virus auxiliar. Típicamente, se puede obtener el vector retrovírico de la presente descripción por un sistema de transcomplementación (sistema de vector/empaquetamiento) transfectando *in vitro* una célula permisiva (tal como células 293T) con un plásmido que contiene el genoma del vector retrovírico de la presente descripción, y al menos otro plásmido que proporciona, *in trans*, las secuencias gag, pol y env codificantes de los polipéptidos GAG, POL y la(s) proteína(s) de la envuelta, o de una porción de estos polipéptidos suficiente para permitir la formación de

partículas retrovíricas. Como ejemplo, se transfectan células permisivas con a) plásmido de transcomplementación, que carece de la señal de empaquetamiento psi y que comprende una secuencia retrovítica gag y pol codificante de una proteína integrasa quimérica de la presente descripción; se desprovee opcionalmente al plásmido de los genes accesorios vif, nef, vpu y/o vpr; b) un segundo plásmido (plásmido de expresión de la envuelta o plásmido env de pseudotipificación) que comprende un gen codificante de una proteína(s) de la envuelta (tal como VSV-G), y c) un vector plasmídico que comprende un genoma retrovítico recombinante, opcionalmente eliminado de la región promotora de la 3' LTR o secuencia potenciadora U3 de la 3' LTR, incluyendo, entre las secuencias LTR 5' y 3' retrovíricas, una secuencia de encapsidación psi, un elemento de exportación nuclear (preferiblemente un elemento RRE de VIH u otros retrovirus equivalentes), el transgén y opcionalmente un promotor y/o una secuencia de importación nuclear (secuencia cPPT, por ej., CTS) del ARN. Ventajosamente, los tres plásmidos usados no contienen secuencia homóloga suficiente para recombinación. Se pueden preparar ventajosamente ácidos nucleicos codificantes de ADNc de gag, pol y env según técnicas convencionales, a partir de secuencias génicas víricas disponibles en la técnica anterior y en las bases de datos. El plásmido de transcomplementación proporciona un ácido nucleico codificante de las proteínas retrovíricas gag y pol. Estas proteínas derivan de un lentivirus, y lo más preferiblemente de VIH-1. El plásmido está desprovisto de secuencia de encapsidación, de secuencia codificante de una envuelta y de genes accesorios, y ventajosamente también carece de LTR retrovíricas. Por lo tanto, se ponen ventajosamente las secuencias codificantes de las proteínas gag y pol bajo el control de un promotor heterólogo, por ej., celular, vírico, etc., que puede ser constitutivo o regulado, débil o fuerte. Es preferiblemente un plásmido que contiene una secuencia transcomplementaria Δ psi-CMV-gag-pol-PolyA. Este plásmido permite la expresión de todas las proteínas necesarias para la formación de viriones vacíos, excepto las glicoproteínas de la envuelta. El plásmido de transcomplementación puede comprender ventajosamente los genes TAT y REV. El plásmido de transcomplementación está ventajosamente desprovisto de genes accesorios vif, vpr, vpu y/o nef. Se entiende que los genes gag y pol y los genes TAT y REV también pueden ser llevados por diferentes plásmidos, posiblemente separados. En este caso, se usan varios plásmidos de transcomplementación, codificando cada uno una o más de dichas proteínas. Los promotores usados en el plásmido de transcomplementación, el plásmido de la envuelta y el vector plasmídico, respectivamente, para promover la expresión de gag y pol de la proteína de la cubierta, el ARNm del genoma del vector y el transgén son promotores idénticos o diferentes, elegidos ventajosamente de promotores ubicuos o específicos, por ejemplo, de los promotores víricos CMV, TK, RSV LTR y el promotor de la ARN polimerasa III, tal como U6 o HI, o promotores de virus auxiliares codificantes de env, gag y pol (es decir, virus adenovíricos, baculovíricos, herpes). Para la producción del vector retrovítico de la presente descripción, se pueden introducir los plásmidos antes descritos en células competentes y se recogen los virus producidos. Las células usadas pueden ser cualquier célula competente, en particular células eucarióticas, en particular de mamífero, por ej., humano o animal. Pueden ser somáticas o madre embrionarias o diferenciadas. Típicamente, las células incluyen células 293T, células fibroblásticas, hepatocitos, células musculares (esqueléticas, cardíacas, lisas, de vaso sanguíneo, etc.), células nerviosas (neuronas, células gliales, astrocitos) o células epiteliales, renales, oculares, etc. También se pueden incluir células de insectos, de plantas, de levadura o procarióticas. También pueden ser células transformadas por el antígeno SV40 T. Se pueden introducir los genes gag, pol y env codificados en plásmidos o virus auxiliares en células por cualquier método conocido en la técnica, adecuado para el tipo de célula considerado. Normalmente, las células y el sistema vector contactan en un dispositivo adecuado (placa, placa, tubo, bolsa, etc.), durante un período de tiempo suficiente para permitir la transferencia del sistema vector o del plásmido en las células. Típicamente, se introduce el sistema vector o el plásmido en las células por precipitación con fosfato de calcio, electroporación, transducción o usando un compuesto facilitador de la transfección, tal como lípidos, polímeros, liposomas y péptidos, etc. Se prefiere la precipitación con fosfato de calcio. Se cultivan las células en cualquier medio adecuado, tal como RPMI, DMEM, un medio específico para un cultivo en ausencia de suero fetal de ternera, etc. Una vez transfectados, se pueden purificar los vectores retrovíricos de la presente descripción del sobrenadante de las células. Se puede realizar la purificación del vector retrovítico para aumentar la concentración por cualquier método adecuado, tal como por purificación en gradiente de densidad (por ej., cloruro de cesio (CsCl)) o por técnicas cromatográficas (por ej., cromatografía en columna o de lotes). Por ejemplo, se puede someter el vector de la presente descripción a dos o tres etapas de purificación en gradiente de densidad con CsCl. Se purifica deseablemente el vector a partir de células infectadas usando un método que comprende la lisis de las células infectadas con adenovirus, la aplicación del lisado a una resina cromatográfica, la elución del adenovirus de la resina cromatográfica y la recogida de una fracción que contiene el vector retrovítico de la presente descripción.

El vector retrovítico de la presente descripción es, en particular, adecuado para conducir la integración dirigida en el genoma de la célula infectada. Por lo tanto, otro objeto de la presente descripción se relaciona con un método para expresar un transgén en el genoma de una pluralidad de células que comprende infectar la pluralidad de células con el vector retrovítico de la presente descripción. Se puede usar el vector retrovítico de la presente descripción, por lo tanto, para expresar el transgén en una célula de mamífero de interés, más en particular en células que no se dividen, o para la expresión transitoria del transgén en células en división que son refractarias a otros métodos de transfección o transducción. Los vectores retrovíricos de la presente descripción son capaces de transducir diversos tipos de células, tales como, por ejemplo, células hepáticas (por ej., hepatocitos), células musculares, células cerebrales, células renales, células retinianas y células hematopoyéticas. En particular, las células diana de la presente descripción son células "que no se dividen". Estas células incluyen células tales como células neuronales que normalmente no se dividen. Sin embargo, no se pretende que la presente descripción se limite a células que no se dividen (incluyendo, aunque sin limitación, células musculares, células blancas de la sangre, células esplénicas, células hepáticas, células oculares, células epiteliales, etc.). Las posibles aplicaciones de los vectores retrovíricos de la presente descripción son de varios tipos e incluyen terapia génica, es decir, la transferencia de genes en cualquier

célula de mamífero, en particular en células humanas. Pueden ser células en división o células quiescentes, células pertenecientes a los órganos centrales o a los órganos periféricos, tales como el hígado, el páncreas, el músculo, el corazón, etc.. Ésta es preferiblemente una transferencia génica en células quiescentes (que no se dividen). La terapia génica puede permitir la expresión de proteínas, por ej., factores neurotróficos, enzimas, factores de transcripción, receptores, etc. También permite implementar una estrategia de terapia celular con "oligonucleótidos" (ARN de interferencia o antisentido, ribozimas, etc.), es decir, la expresión de factores de diferenciación en células progenitoras para guiar el destino celular a uno seleccionado antes del trasplante o de la transducción *ex vivo* de las células para expresar un factor de interés, seguido de trasplante de dichas células.

Los vectores retrovíricos de la presente descripción pueden ser también, en particular, adecuados para fines de investigación.

Los vectores retrovíricos de la presente descripción son también, en particular, adecuados para preparar organismos genéticamente modificados. Por "genéticamente modificado", se quiere decir un gen que está alterado con respecto a su estado nativo (por ej., por mutación de inserción, mutación de delección, mutación en la secuencia de ácido nucleico, u otra mutación), o que un producto génico está alterado con respecto a su estado natural (por ej., por administración de un transgén que trabaja en trans sobre un ARNm o proteína codificados por un gen, tal como administración de ARN inhibitorio o administración de un transgén negativo dominante). Típicamente, dicho organismo genéticamente modificado puede ser cualquier organismo de origen vegetal o animal que se use como alimento o piensos, para producir cultivos o material de fabricación, o para producir animales transgénicos para alimentos o piensos, o para la cría de ganado.

Los vectores retrovíricos de la presente descripción pueden también ser, en particular, adecuados para la producción de vacunas o para provocar una respuesta vacunal en un sujeto que lo necesite.

También se puede usar el vector retrovílico de la presente descripción como un medicamento. El vector retrovílico de la presente descripción puede ser, en particular, adecuado para tratar una enfermedad en un sujeto.

En particular, se puede usar el vector retrovílico de la presente descripción para tratar un cáncer. Como ejemplos no limitativos de cánceres que pueden tratarse mediante la presente descripción, se incluyen cáncer de mama, cáncer de próstata, linfoma, cáncer de piel, cáncer pancreático, cáncer de colon, melanoma, melanoma maligno, cáncer de ovarios, cáncer de cerebro, carcinoma cerebral primario, cáncer de cabeza y cuello, glioma, glioblastoma, cáncer de hígado, cáncer de vejiga, cáncer pulmonar de células no pequeñas, carcinoma de cabeza o cuello, carcinoma de mama, carcinoma de ovarios, carcinoma de pulmón, carcinoma de pulmón de células pequeñas, tumor de Wilms, carcinoma cervical, carcinoma testicular, carcinoma de vejiga, carcinoma pancreático, carcinoma de estómago, carcinoma de colon, carcinoma prostático, carcinoma genitourinario, carcinoma tiroideo, carcinoma esofágico, mieloma, mieloma múltiple, carcinoma adrenal, carcinoma de células renales, carcinoma endometrial, carcinoma de la corteza adrenal, insulinooma pancreático maligno, carcinoma carcinoide maligno, coriocarcinoma, micosis fungoides, hipercalcemia maligna, hiperplasia cervical, leucemia, leucemia linfocítica aguda, leucemia linfocítica crónica, leucemia mielógena aguda, leucemia mielógena crónica, leucemia granulocítica crónica, leucemia granulocítica aguda, leucemia de células pilosas, neuroblastoma, rhabdomyosarcoma, sarcoma de Kaposi, policitemia vera, trombocitosis esencial, enfermedad de Hodgkin, linfoma no Hodgkin, sarcoma de tejidos blandos, mesotelioma, sarcoma osteogénico, macroglobulinemia primaria y retinoblastoma, y similares.

En particular, se puede usar el vector retrovílico de la presente descripción para tratar un trastorno autoinmune, incluyendo, aunque sin limitación, un trastorno seleccionado del grupo consistente en Hepatitis Crónica Activa Autoinmune por Aclorhidria, Encefalomiелitis Diseminada Aguda, leucoencefalitis hemorrágica Aguda, Enfermedad de Addison, gammaglobulinemia, Agammaglobulinemia, Alopecia areata, Esclerosis Lateral Amiotrófica, Espondilitis Anquilosante, Nefritis Anti-GBM/TBM, síndrome Antifosfolípido, síndrome Antisintetasa, Artritis, alergia Atópica, Dermatitis Atópica, Anemia Aplásica, miocardiopatía Autoinmune, anemia hemolítica Autoinmune, hepatitis Autoinmune, enfermedad Autoinmune del oído interno, síndrome linfoproliferativo Autoinmune, neuropatía periférica Autoinmune, pancreatitis Autoinmune, síndrome poliendocrino Autoinmune Tipos I, II y III, dermatitis Autoinmune por progesterona, púrpura trombocitopénica Autoinmune, uveítis Autoinmune, enfermedad de Balo/esclerosis concéntrica de Balo, Síndrome de Bechets, enfermedad de Berger, encefalitis de Bickerstaff, síndrome de Blau, Penfigoide Buloso, enfermedad de Castelman, Síndrome de Disfunción Inmune con Fatiga Crónica, polineuropatía desmielinizante inflamatoria crónica, osteomielitis multifocal recurrente Crónica, síndrome de Churg-Strauss, Penfigoide Cicatricial, Enfermedad Celíaca, síndrome de Cogan, enfermedad de aglutininas por Frío, deficiencia en el componente 2 del Complemento, arteritis Craneal, síndrome de CREST, Enfermedad de Crohn, Síndrome de Cushing, angieitis leucocitoclástica Cutánea, enfermedad de Dego, Dermatitis herpetiforme, Dermatomiositis, Diabetes mellitus tipo 1, esclerosis sistémica cutánea Difusa, síndrome de Dressler, lupus eritematoso Discoide, eczema, artritis relacionada con Entesitis, fascitis Eosinofílica, Epidermolisis bullosa adquirida, Eritema nodoso, crioglobulinemia mixta Esencial, síndrome de Evan, Fibrodisplasia osificante progresiva., Fibromiositis, alveolitis Fibrosante, Gastritis, penfigoide Gastrointestinal. Arteritis de células Gigantes, síndrome de Goodpasture, enfermedad de Graves, síndrome de Guillain-Barre (GBS), encefalitis de Hashimoto, tiroiditis de Hashimoto, anemia Hemolítica, púrpura de Henoc-Schonlein, Herpes gestationis, síndrome de Hughes (o síndrome Antifosfolípidos). Hipogammaglobulinemia, Enfermedades Desmielinizantes Inflamatorias Idiopáticas, fibrosis pulmonar Idiopática, púrpura trombocitopénica Idiopática, nefropatía por IgA (o enfermedad de Bergefs), miositis por cuerpos de inclusión, polineuropatía

desmielinizante inflamatoria, artritis idiopática Juvenil, artritis reumatoide Juvenil, síndrome miasténico de Lambert-Eaton, vasculitis Leucocitoclástica, Liquen plano, Liquen escleroso, enfermedad de IgA Lineal (LAD), Enfermedad de Lou Gehrig, hepatitis Lupoide, Lupus eritematoso, síndrome de Majeed, enfermedad de Meniere, poliangiitis Microscópica, síndrome de Miller-Fisher, Enfermedad Mixta del Tejido Conectivo, enfermedad de Mucha-Habermann, síndrome de Muckle-Wells, Mieloma Múltiple, Miastenia gravis, Miositis, Narcolepsia, Neuromielitis óptica (también Enfermedad de Devic), penfigoide cicatricial Ocular, tiroiditis de Ord, reumatismo Palindrómico, PANDAS (Trastornos Neuropsiquiátricos Autoinmunes Pediátricos Asociados a Streptococcus), degeneración cerebelar Paraneoplásica, degeneración cerebelar Paraneoplásica, síndrome de Parry Romberg, síndrome de Parsonnage-Turner, planitis de Pars, Pénfigo, Pénfigo vulgar, anemia Perniciosa, encefalomiелitis Perivenosa, síndrome de POEMS, Poliarteritis nodosa, Polimialgia reumática, Polimiositis, cirrosis biliar Primaria, psoriasis, artritis psoriásica, Pioderma gangrenosa, aplasia pura de células rojas, encefalitis de Rasmussen, fenómeno de Raynaud, policondritis Recidivante, síndrome de Reiter, fibrosis Retroperitoneal, artritis Reumatoide, fiebre Reumatoide, síndrome de Schmidt, síndrome de Schnitzler, Escleritis, síndrome de Sjogren, Espondiloartropatía, síndrome de la sangre pegajosa, Enfermedad de Still, endocarditis bacteriana Subaguda (SBE), síndrome de Susac, síndrome de Sweet, Corea de Sydenham, oftalmia Simpática, arteritis de Takayasu, arteritis Temporal, síndrome de Tolosa-Hunt, Mielitis Transversa, Colitis Ulcerativa, enfermedad del tejido conectivo No diferenciado, espondiloartropatía No Diferenciada, vasculitis, granulomatosis de Wegener, síndrome de Wilson, y síndrome de Wiskott-Aldrich.

En particular, se puede usar el vector retroviral de la presente descripción para tratar un trastorno ocular que incluye, aunque sin limitación, un trastorno seleccionado del grupo consistente en glaucoma, incluyendo Glaucoma de Ángulo Abierto (por ej., Glaucoma de Ángulo Abierto Primario, Glaucoma Pigmentario y Glaucoma Exfoliativo, Glaucoma de Baja Tensión), Glaucoma con Cierre de Ángulo (también conocido clínicamente como glaucoma de ángulo cerrado, glaucoma de ángulo estrecho, glaucoma de bloque pupilar y glaucoma de bloque ciliar) (por ej., Glaucoma con Cierre de Ángulo Agudo y Glaucoma con Cierre de Ángulo Crónico), Glaucoma Anirídico, Glaucoma Congénito, Glaucoma Juvenil, Glaucoma Inducido por Lente, Glaucoma Neovascular (por ej., usando vectores compuestos por señuelo de Factor de Crecimiento Endotelial Vascular (VEGF), Factor de Crecimiento Derivado de Pigmento (PDGF), Endostatina, Angiostatina o Angiopoyetina-1), Glaucoma Postraumático, Glaucoma Inducido por Esteroides, Glaucoma del Síndrome de Sturge-Weber y Glaucoma Inducido por Uveítis, retinopatía diabética (por ej., usando vectores compuestos por señuelo de VEGF, PDGF, Endostatina, Angiostatina o Angiopoyetina-1), degeneración macular (por ej., vectores compuestos por señuelo de VEGF, PDGF, Endostatina, Angiostatina, Angiopoyetina-1, Casete de Unión a ATP Subfamilia A Miembro 4), degeneración macular (por ej., usando vectores compuestos por señuelo de VEGF, PDGF, Endostatina, Angiostatina, Angiopoyetina-1, Casete de Unión a ATP Subfamilia A Miembro 4), neovascularización coroidal (por ej., usando vectores compuestos por señuelo de VEGF, PDGF, Endostatina, Angiostatina o Angiopoyetina-1), fuga vascular y/o edema retiniano, conjuntivitis bacteriana, conjuntivitis fúngica, conjuntivitis vírica, uveítis, precipitados queráticos, edema macular (por ej., usando vectores compuestos por señuelo de VEGF, PDGF, Endostatina, Angiostatina o Angiopoyetina-1), respuesta inflamatoria tras implantación de lentes intraoculares, síndromes de uveítis (por ejemplo, iridociclitis crónica o endoftalmitis crónica), vasculitis retiniana (por ejemplo, como se ve en la artritis reumatoide, en la artritis reumatoide juvenil, en el lupus eritematoso sistémico, en la escleritis sistémica progresiva, en la poliarteritis nodosa, en la granulomatosis de Wegener, en la arteritis temporal, en la enfermedad de Adamantiades Bechcet, en la policondritis recidivante de Sjogren y en la espondilitis asociada a HLA-B27), sarcoidosis, enfermedad de Eales, necrosis retiniana aguda, síndrome de Vogt Koyanagi Harada, toxoplasmosis ocular, retinopatía por radiación, vitreoretinopatía proliferativa, endoftalmitis, glaucomas oculares (por ejemplo, glaucomas inflamatorios), neuritis óptica, neuropatía óptica isquémica (por ej., vectores compuestos por NADH deshidrogenasa Alotópica Unidad 4), orbitopatía asociada al tiroides, pseudotumor orbitario, síndrome de la dispersión de pigmento (glaucoma pigmentario), escleritis, epiescleritis, coroidopatías (por ejemplo, síndromes de "punto Blanco", incluyendo, aunque sin limitación, placoide posterior multifocal agudo), retinopatías (por ejemplo, edema macular cistoide, coroidopatía serosa central y síndrome de presunta histoplasmosis ocular (por ej., vectores compuestos por Factor Neurotrópico Derivado de Células Gliales, Periferina-2)), enfermedad vascular retiniana (por ejemplo, retinopatía diabética, enfermedad de Coat y macroaneurisma arterial retiniano), oclusiones de la arteria retiniana, oclusiones de la vena retiniana, retinopatía de premadurez, retinitis pigmentosa (por ej., vectores compuestos por proteína de 65kDa Específica de Pigmento Retiniano), vitreoretinopatía exudativa familiar (FEVR), vasculopatía coroidal polipoidal idiopática, membranas maculares epirretinianas y cataratas.

En particular, se puede usar el vector retroviral de la presente descripción para tratar un trastorno de la sangre, que incluye, aunque sin limitación, un trastorno de la sangre seleccionado del grupo consistente en anemia, sangrado y trastornos de la coagulación (por ej., coagulación intravascular diseminada (CID), hemofilia, Púrpura de Henoch-Schonlien, telangiectasia hemorrágica hereditaria, trombocitopenia (ITP, TTP), trombofilia, enfermedad de Von Willebrand), leucemias (por ej., leucemia linfocítica aguda, leucemia mielocítica aguda, leucemia linfocítica crónica, leucemia mielocítica crónica), linfomas (por ej., linfoma de Hodgkin, linfoma no Hodgkin), trastornos mieloproliferativos (por ej., mielofibrosis, Policitemia Vera, trombocitemia), trastornos de las células plasmáticas (por ej., macroglobulinemia, gammopatías monoclonales de significación indeterminada, mieloma múltiple), trastornos del bazo, trastornos de las células blancas de la sangre (por ej., trastorno basofílico, trastorno eosinofílico, linfocitopenia, trastornos de los monocitos, neutropenia, leucocitosis neutrofilica), trombosis, trombosis venosa profunda (DVT), hemocromatosis, menorragia, enfermedad de las células falciformes y talasemia.

En particular, se puede usar el vector retrovítico de la presente descripción para tratar un trastorno neurológico, que incluye, aunque sin limitación, un trastorno neurológico seleccionado del grupo consistente en enfermedad de Gaucher, enfermedad de Parkinson, enfermedad de Alzheimer, esclerosis lateral amiotrófica (ALS), esclerosis múltiple (MS), enfermedad de Huntington, ataxia de Fredrich, Deterioro Cognitivo Leve, Angiopatia Amiloide Cerebral, Enfermedad del Parkinsonismo, Enfermedad de los Cuerpos de Lewy, Demencia Frontotemporal (FTD), Atrofia de Múltiples Sistemas (MSA), Parálisis Supranuclear Progresiva y trastornos del movimiento (incluyendo ataxia, parálisis cerebral, coreoatetosis, distonía, síndrome de Tourette, kernícterus) y trastornos de temblores, y leucodistrofias (incluyendo adrenoleucodistrofia, leucodistrofia metacromática, enfermedad de Canavan, enfermedad de Alexander, enfermedad de Pelizaeus-Merzbacher), lipofuscinosis ceroides neuronales, ataxia telangiectasia, Síndrome de Rett, alfa-sinucleinopatía (por ej., Enfermedad de los Cuerpos de Lewy, Atrofia de Múltiples Sistemas, enfermedad de Hallervorden-Spatz, o Demencia Frontotemporal), enfermedad de Niemann-Pick Tipo C (NPCD), ataxia espinocerebelar Tipo 1, Tipo 2 y Tipo 3 y atrofia palidoluisiana dentatorrubral (DRLPA).

En particular, se puede usar el vector retrovítico de la presente descripción para tratar un trastorno del pulmón, que incluye, aunque sin limitación, un trastorno del pulmón seleccionado del grupo consistente en asma, atelectasia, bronquitis, EPOC (enfermedad pulmonar obstructiva crónica), enfisema, cáncer de Pulmón, mesotelioma, neumonía, asbestosis, Aspergiloma, Aspergilosis, Aspergilosis - bronquiectasia invasiva aguda, bronquiolitis obliterante con neumonía organizativa (BOOP), neumonía eosinofílica, neumonía necrotizante, efusión pleural, neumoniosis, neumotórax, actinomicosis pulmonar, proteinosis alveolar pulmonar, ántrax pulmonar, malformación arteriovenosa pulmonar, fibrosis pulmonar, émbolo pulmonar, histiocitosis pulmonar X (granuloma eosinofílico), hipertensión pulmonar, edema pulmonar, hemorragia pulmonar, noocardiosis pulmonar, tuberculosis pulmonar, enfermedad venooclusiva pulmonar, enfermedad pulmonar reumatoide, sarcoidosis, fibrosis por radiación, neumonitis por hipersensibilidad, síndrome de distrés respiratorio agudo (ARDS), síndrome de distrés respiratorio del niño, fibrosis pulmonar idiopática, neumonía intersticial idiopática, linfangioleiomiomatosis, histiocitosis pulmonar de células de Langerhans, proteinosis alveolar pulmonar, sinusitis, tonsillitis, otitis media, faringitis, laringitis, hamartoma Pulmonar, secuestro pulmonar, malformación adenomatoide quística congénita (CCAM) y fibrosis quística.

En particular, se puede usar el vector retrovítico de la presente descripción para tratar una enfermedad infecciosa en un humano, que incluye, aunque sin limitación, una enfermedad infecciosa seleccionada del grupo consistente en enfermedades fúngicas, tales como dermatofitosis (por ej., tricofitosis, tiña o infecciones de tiña), pie de atleta, paroniquia, pitiriasis versicolor, eritrasma, intétrigo, sarpullido fúngico del pañal, Candida vulvitis, Candida balanitis, otitis externa, candidiasis (cutánea y mucocutánea), mucocandidiasis crónica (por ej., muguet y candidiasis vaginal), criptococcosis, geotricosis, tricosporosis, aspergilosis, peniciliosis, fusariosis, zigomicosis, esporotricosis, cromomicosis, coccidioidomicosis, histoplasmosis, blastomicosis, paracoccidioidomicosis, pseudallescheriasis, micetoma, queratitis micótica, otomicosis, neumocistosis y fungemia, infecciones por Acinetobacter, Actinomicosis, enfermedad del sueño africana, SIDA (síndrome de inmunodeficiencia Adquirida), Amebiasis, Anaplasmosis, Ántrax, infección por Arcanobacterium haemolyticum, fiebre hemorrágica argentina, Ascariasis, Aspergilosis, infección por atrovirus, Babesiosis, infección por Bacillus cereus, neumonía Bacteriana, vaginosis Bacteriana (BV), infección por Bacteroides, Balantidiasis, infección por Bailisascaris, infección por el virus BK, piedra Negra, infección por Blastocistis hominis, infección por Borrelia, Botulismo (y botulismo Infantil), fiebre hemorrágica brasileña, Brucelosis, infección por Burkholderia, úlcera de Buruli, infección por Calicivirus (Norovirus y Sapovirus), Candidiasis, enfermedad del arañazo del gato, Celulitis, Enfermedad de Chagas (trpanosomiasis americana), Chancroide, Varicela, Clamidia, Cólera, Cromoblastomicosis, Clonorchiasis, Clostridium difficile, Coccidioidomicosis, fiebre por garrapatas de Colorado (CTF), resfriado Común (rinofaringitis vírica aguda; coriza Aguda), enfermedad de Creutzfeldt-Jakob (CJD), Criptococcosis, Criptosporidiosis, larva migratoria cutánea (CLM), fiebre del Dengue, Dientamoebiasis, Difteria, Difilobotriasis, Difilobotriasis, Dracunculiasis, fiebre hemorrágica del Ébola, Equinococosis, Ehrlichiosis, Enterobiasis (infección por Oxiuros), infección por Enterococcus, infección por Enterovirus, tifus Epidémico, Eritema infeccioso, Exantema súbito, Fasciolosis, Fasciolosis, insomnio familiar Fatal (FFI), Filariasis, infección por Fusobacterium, gangrena Gaseosa (mionecrosis por Clostridio), Geotricosis, síndrome de Gerstmann-Straussler-Scheinker (GSS), Giardiasis Glanders, Gnatostomiasis, Gonorrea, Granuloma inguinal (Donovanosis), infección estreptocócica del Grupo A, infección estreptocócica del Grupo B, Haemophilus influenzae, enfermedad de la mano, el pie y la boca (HFMD), Síndrome Pulmonar por Hantavirus (HPS), infección por Helicobacter pilori, síndrome urémico hemolítico (HUS), fiebre Hemorrágica con síndrome renal (HFRS), Hepatitis A, B, C, D, E, Herpes simplex, Histoplasmosis, infección por Anquilostomas, infección por bocavirus, ehrlichiosis ewingii Humana, anaplasmosis granulocítica Humana (HGA), anaplasmosis granulocítica Humana (HGA), ehrlichiosis monocítica Humana, infección por papilomavirus Humano (HPV), infección por el virus de la parainfluenza Humana, Himenolepiasis, Mononucleosis Infecciosa por el Virus Epstein-Barr (Mono), Influenza (flu), Isosporiasis, enfermedad de Kawasaki, Queratitis, infección por Kingella kingae, Kuru, fiebre de Lassa, Legionelosis (enfermedad de los Legionarios), Legionelosis (fiebre de Pontiac), Leishmaniasis, Lepra, Leptospirosis, Listeriosis, enfermedad de Lyme (borreliosis de Lyme), filariasis Linfática (Elefantiasis), coriomeningitis Linfocítica, Malaria, fiebre hemorrágica de Marburg (MHF), Sarampión, Melioidosis (enfermedad de Whitmore), Meningitis, enfermedad Meningocócica, Metagonimiasis, Microsporidiosis, Molluscum contagiosum (MC), Papeas, tifus Murino (tifus Endémico), Mycoplasma pneumoniae, Micetoma, Miasis, conjuntivitis Neonatal (Oftalmia de los neonatos), (Nueva) Variante de la enfermedad de Creutzfeldt-Jakob (vCJD, nCJD), Noocardiosis, Oncocercosis (ceguera de los Ríos), Paracoccidioidomicosis (blastomicosis Sudamericana), Paragonimiasis, Pasteurellosis, Pediculosis de la cabeza (piojos de la Cabeza), Pediculosis del cuerpo (piojos del Cuerpo), Pediculosis del pubis (piojos Púlicos, Ladillas), enfermedad inflamatoria Pélvica (PID), Pertussis (Tos ferina), Peste, infección por

neumococos, *Pneumocystis pneumoniae* (PCP), Neumonía, Poliomieltis, *Poliomielitis*, infección por *Prevotella*, meningoencefalitis amebiana primaria (PAM), leucoencefalopatía multifocal Progresiva, Psittacosis, fiebre Q, Rabia, fiebre por mordedura de Rata, infección por el virus sincitial Respiratorio, Rhinosporidiosis, infección por inovirus, infección por *Rickettsias*, *Rickettsiosis* vesicular, fiebre del Valle del Rift (RVF), fiebre maculosa de las Montañas Rocosas (RMSF), infección por Rotavirus, Rubella, Salmonellosis, SARS (Síndrome Respiratorio Agudo Severo), Sarna, Schistosomiasis, Sepsis, Shigellosis (disentería Bacilar), Culebrilla (Herpes zóster), Viruela (Variola), Sporotrichosis, envenenamiento alimentario Estafilocócico, infección Estafilocócica, Strongiloidiasis, Sífilis, Teniasis, tétanos (Tétanos), Tinea de la barba (picor del Barbero), Tinea de la cabeza (Tiña del Cuero Cabelludo), Tinea corporis (Tiña del Cuerpo), Tinea inguinal (tiña Genital), Tinea de la mano (Tiña de la Mano), Tinea negra, Tinea de las uñas (Onicomycosis), Tinea versicolor (Pitiriasis versicolor). Toxocariasis (Larva Migratoria Visceral (VLM)), Toxoplasmosis, Triquinelosis, Tricomonirosis, Tricuriasis (infección por Gusano Látego), Tuberculosis, Tularemia, infección por *Ureaplasma urealyticum*, encefalitis equina Venezolana, fiebre hemorrágica Venezolana, neumonía vírica. Fiebre del Nilo Occidental, plectra Blanca (Tinea blanca), infección por *Yersinia pseudotuberculosis*, Yersiniosis, fiebre Amarilla y Zigomicosis.

Se pueden administrar los vectores retrovíricos de la presente descripción a un sujeto por cualquier vía. En particular, se administra el vector retrovírico de la presente descripción al sujeto parenteralmente, preferiblemente intravascularmente (incluyendo intravenosamente). Cuando se administran parenteralmente, se prefiere dar los vectores en un vehículo farmacéutico adecuado para inyección, tal como una solución o dispersión acuosa estéril. Tras la administración, se monitoriza al sujeto para detectar la expresión del transgén. Se determinan la dosis y la duración del tratamiento individualmente dependiendo de la afección o enfermedad que haya que tratar. Se puede tratar una amplia variedad de afecciones o enfermedades en base a la expresión génica producida mediante la administración del gen de interés en el vector de la presente descripción. La dosificación del vector administrado usando el método de la presente descripción variará dependiendo de la respuesta deseada por parte del huésped y del vector utilizado. En general, se espera poder administrar hasta 100-200 µg de ADN o ARN en una sola dosificación, aunque un intervalo de 0,5 mg/kg de peso corporal a 50 mg/kg de peso corporal será adecuado para la mayoría de las aplicaciones.

Se ilustrará además la descripción mediante las siguientes figuras y ejemplos. Sin embargo, no se han de interpretar estos ejemplos y figuras en modo alguno como limitantes del alcance de la presente descripción.

Figuras:

Figura 1. Ty1 IN interacciona con AC40, pero no con su ortólogo *S. pombe*, AC40sp.

(A) Interacción de dos híbridos entre diferentes dominios de Ty1-IN y AC40 o AC40sp. Izquierda, estructura de Ty1. Dos marcos abiertos de lectura, *GAG* y *POL* (históricamente *TYA* y *TYB*), están flanqueados por dos LTR (triángulos negros). Proteasa (PR), integrasa (IN), transcriptasa inversa-ARNasa H (RT/RH). La flecha izquierda-derecha indica secuencias de Ty1 recuperadas en el cribado de dos híbridos (2-HS) con AC40 como cebo (coordenadas 2706-4342 en Ty1). Se representan regiones de Ty1 IN fusionadas al dominio activador Gal4 (GAD). Derecha, diluciones seriadas con factor de dilución de 10 de cultivos líquidos de células que expresan AC40 o AC40sp fusionado al dominio de unión Gal4 (GBD) y se plaquearon las diversas proteínas GAD-IN en medio no selectivo (Control) o selectivo (Interacción) para detectar interacciones. (B) Coinmunoprecipitación de AC40 e IN-Strep. Se insertó un epítipo 6xHA en el extremo del gen *RPC40* endógeno (AC40-HA) y se fusionó IN con estreptavidina (IN-Strep) y se expresó a partir de un promotor inducible por galactosa. Se realizó la inmunoprecipitación de extractos de proteínas con perlas magnéticas de anti-HA copuladas a IgG. Se revelaron las proteínas en transferencias Western con anticuerpos anti-HA 12CA5 y anti-Strep. (C) Coinmunoprecipitación de AC40 e IN expresada a partir de un elemento Ty1 funcional inducible por galactosa. Las condiciones de Inmunoprecipitación son como en la Fig. 3B. Se revelaron las proteínas con anticuerpos anti-HA 12CA5 y anti-IN 8B11.

Figura 2. La interacción entre Ty1-IN y AC40 es importante para que Ty1 dirija secuencia arriba dos genes de ARNt, pero no para la movilidad global de Ty1.

(A) Integración de Ty1 en el locus *URA3-tG(GCC)B*. El histograma muestra el índice medio de colonias 5FOA^R obtenidas de cultivos (32 y 55 para cepas que expresan AC40 y AC40sp, respectivamente) e inducidas para la retrotransposición de Ty1 a 20°C (barra gris) y de inserciones de Ty1 en colonias *URA3/5FOA^R* identificadas por PCR (barras negras) con los errores estándar. (B) Análisis de PCR de colonias 5FOA^R de cultivos independientes (30 y 27 para cepas que expresan AC40 y AC40sp, respectivamente) para detectar sucesos de inserción de Ty1 en *URA3* (cebadores indicados por triángulos negros hacia la izquierda y hacia la derecha). Se muestran resultados de PCR representativos de 4 colonias por cada cepa, una banda de 6 kb indica un elemento Ty1, una banda de 2,5 kb que no hay inserción. Se muestra un control negativo (C): locus *URA3-tG(GCC)B* de una cepa cultivada a la temperatura no permisiva de 30°C para la retrotransposición de Ty1. (C) Detección de inserciones *de novo* de Ty1 secuencia arriba de *tG(GCC)C* en cepas que expresan AC40 y AC40sp. Superior: locus *tG(GCC)C* y diseño experimental. Se preparó ADN a partir de 5 cultivos independientes de cada cepa cultivada a 20°C y que expresaba Ty1 a partir del promotor *GAL1* inducible por galactosa. Se llevaron a cabo las reacciones PCR con cebadores específicos para *tG(GCC)C* y Ty1 (triángulos negros hacia la izquierda y hacia la derecha). La inserción de Ty1 se produce con igual frecuencia en ambas orientaciones (3, 4). Sólo se analizaron los sucesos de inserción en orientación opuesta a *tG(GCC)C* por PCR.

N.E. una banda no específica correspondiente a la amplificación de un fragmento a partir de elementos Ty1 y Ty2 adyacentes en el genoma. (D) Frecuencia de retrotransposición de Ty1 *his3A1* en cepas que expresan AC40 y AC40sp, indicada como el número de prototifos His⁺ dividido por el número total de células. Se representan los datos de 6 cultivos independientes como gráficos de cajas.

- 5 Figura 3. Asociación de las inserciones de Ty1 con las características cromosómicas en WT (AC40) y cepas mutantes con pérdida de interacción (AC40sp).

(A) Se agregaron inserciones de Ty1 secuencia arriba de los 275 genes de ARNt de *S. cerevisiae* en una distancia de medición de una sola distribución al inicio de la transcripción (posición 0 en el eje de las x). Los puntos negros por encima y por debajo del eje de las x representan integraciones convergentes y divergentes, en relación al gen de ARNt, respectivamente. Las curvas LOESS (alisamiento de diagrama de dispersión localmente ponderado) en rojo y azul indican las tendencias generales. El fondo gris indica el nivel de nucleosoma en cada posición, en base a la secuenciación del ADN cromosómico tras digestión por nucleasa microcócica (30). Se normaliza la escala del eje de las y al número de integraciones totales en cada muestra. Se llevaron linealmente a escala los nucleosomas para la visibilidad de la periodicidad. El panel izquierdo muestra inserciones Ty1 en WT (AC40) y el panel derecho inserciones Ty1 en el mutante (AC40sp). (B) Gráfico de un subgrupo de características representativas con AUC positiva y negativa AUC - 0,5 (AUC, valores del Área Bajo la Curva de la Característica del Operador Receptor). 0 indica un modelo de ningún poder predictivo, valores de magnitud 0,5 indican una perfecta predicción. Los valores negativos y positivos representan una característica que se asocia al control y a la integración, respectivamente. Se obtuvieron valores del AUC de los nucleosomas de regiones cromosómicas para las que se dispone de datos de nucleosomas; se obtuvieron los otros valores del AUC del conjunto de datos de todo el genoma. (C) Superior: Distribución de las inserciones de Ty1 en el cromosoma III. Las barras verticales indican el número de integraciones no ambiguas en un sitio particular. Las barras rojas y azules por encima y por debajo del eje de las x indican inserciones en WT (AC40) y el mutante con pérdida de interacción (AC40sp), respectivamente. Las líneas verticales indican genes de ARNt en orientaciones hacia delante (azul) o a la inversa (rosa). Inferior: Zoom de la distribución de inserciones de Ty1 en el subtelómero derecho del cromosoma III (coordinada 292388 a 316620). Se representan el locus *HMR* (línea gruesa verde), el gen de ARNt tT(AGU)C (triángulo marrón claro), una gran región subtelomérica que contiene genes no esenciales (línea gruesa gris), un ARS (línea gruesa azul) y el telómero TEL03R (línea gruesa negra). El eje de las x representa la posición a lo largo del cromosoma a una resolución de 1.000 pb para todo el cromosoma y 100 pb para el subtelómero derecho. La escala del eje de las y muestra el porcentaje de integraciones totales en el genoma por recipiente.

- 30 Figura 4: Esquema del sistema de dos híbridos. Los factores de transcripción eucarióticos consisten en dos dominios principales que se pliegan independientemente: un dominio de unión a ADN (BD) que se une a secuencias diana específicas en un promotor y un dominio de activación (AD) que recluta la maquinaria de transcripción. El concepto básico de nuestro cribado de dos híbridos de levadura se basa en la reconstitución de un factor de transcripción activo resultante de la interacción entre dos proteínas, la proteína "cebo", aquí AC40, fusionada al BD de GAL4 (GBD), y la otra proteína, la "presa", aquí las diferentes construcciones IN fusionadas al AD de GAL4. La base del ensayo de dos híbridos es que la transcripción del gen indicador HIS3 guarda correlación con la interacción entre las proteínas cebo y presa.

- Figura 5: Caracterización del TD de Ty1 por ensayo de dos híbridos. En la parte superior, se representa el TD de Ty5 como un ejemplo. Era el primer dominio de abordaje identificado en la integrasa de un retroelemento. Corresponde a la región mínima que interactúa con la proteína Sir4, que dirige la integración de Ty5 en regiones de heterocromatina. Se predijo el TD de Ty1 en base a las alineaciones de las integrasas de la región C-terminal Ty1, Ty2 y Ty4, ya que los tres retrotransposones tienen la misma preferencia para integrarse secuencia arriba de genes transcritos por Pol III. Los rectángulos indican regiones/aminoácidos que estaban bien conservados entre las 3 integrasas. Se realizó una sustitución de alanina de LE601-602, VS609-610, T616 (junto con la eliminación de W614), K617, N618, M619, R620, S621, L622, E623 y SLE621-623 en GAD-IN578-635 y se estudió la interacción de las diferentes proteínas de fusión con GBD-AC40. El crecimiento indica una interacción. (+) indica una interacción entre IN y AC40, (-) no hay interacción.

- Figura 6: Caracterización del TD de Ty1 mínimo por ensayo de dos híbridos. A) Se fusionó el péptido 617KNMRSLEPP625 (SEQ ID NO: 7) de Ty1 IN con GAD y se estudió la interacción entre las dos fusiones proteicas por ensayo de dos híbridos. Control positivo: IN₅₇₈₋₆₃₅, control negativo: IN₅₇₈₋₆₃₅ con la sustitución R620A, Ty#125: control positivo, que corresponde a una secuencia obtenida en el cribado de dos híbridos original descrito en el Ejemplo 1. Los resultados indican que el péptido 617KNMRSLEPP625 interacciona débilmente con AC40. B) Las secuencias IN₅₉₆₋₆₃₀ e IN₆₁₇₋₆₂₃ (SEQ ID NO: 4) fusionadas a la secuencia tag de estreptavidina interaccionan con AC40. Control positivo: IN₅₇₈₋₆₃₅, control negativo IN₁₋₅₇₈.

- Figura 7: El reemplazo de la secuencia 617KNMRSLE622 por la secuencia 617KEMDSL622 (SEQ ID N: 2) en GAD-IN578-635 mantiene la interacción con GBD-AC40. Se estudió la interacción entre GBD-AC40 y las diferentes construcciones de proteína GAD-IN por ensayo de dos híbridos. Control negativo (GAD-IN578-635 con la sustitución S621A), control positivo, GAD-IN578-635 con la sustitución E623A).

Figura 8: Las mutaciones que abolen la interacción IN/AC40 disminuyen la integración de Ty1 secuencia arriba de genes transcritos por Pol III, pero no la movilidad global de Ty1.

A) La movilidad global de Ty1 disminuye no más del doble en mutantes de Ty1 que llevan mutaciones que abolen la interacción IN/AC40. Frecuencia relativa de retrotransposición de Ty1 en cepas de levadura que expresan diversos mutantes de *Tilhis3AI*, en donde se modificaron cualquiera de los residuos K617, M619, R619, S621 o L622 en A (alanina). Para cada cepa, se dividió el número de células His⁺ por el número total de células. Se ajustaron los valores al 100% de retrotransposición relativa para WT Ty1 y se normalizaron para elementos mutantes de Ty1.

B) La integración de Ty1 secuencia arriba de genes Pol III se ve severamente afectada en mutantes de Ty1 que llevan mutaciones que abolen la interacción IN/AC40. Detección de sucesos de inserción de Ty1 secuencia arriba de *tG(GCC)C* en cepas de levadura que expresan diversos mutantes de Ty1 (en donde cualquiera de los residuos S621 o L622 se modificaron en A (alanina)). Un patrón de bandas periódico es característico de un elemento Ty1 funcional (WT). La pérdida de este patrón en mutantes indica que han perdido la capacidad de integrarse secuencia arriba de genes Pol III. También se obtuvo la pérdida de este patrón en mutantes de Ty1 en donde se modificaron cualquiera de los residuos K617 o M619 en A (alanina). Se da el diseño experimental en la leyenda de la Fig 2C.

Ejemplos:

Ejemplo 1: Una subunidad de la ARN polimerasa III determina sitios de integración de retrotransposones

Existe un equilibrio entre los efectos perjudiciales a corto plazo de elementos transposables como mutágenos y el papel positivo a largo plazo que tienen como agentes de plasticidad y evolución del genoma (1). Es importante para mantener este equilibrio la elección del sitio de integración (ISC). La integración es típicamente no aleatoria y con frecuencia se produce en regiones del genoma en donde la inserción es benigna. Para los retrotransposones - parásitos genómicos obligados - la integración preferentemente se produce en regiones que son pobres en genes, probablemente la consecuencia de selección de mecanismos que favorezcan inserciones no deletéreas. Ty1, el LTR-retrotransposón más activo y abundante de *S. cerevisiae*, preferentemente se integra en una ventana de 1 kb secuencia arriba de genes transcritos con Pol III (2). El abordaje de Ty1 requiere un promotor de Pol III funcional en el gen diana (2-4) y está influenciado por el factor de remodelación de cromatina Isw2 y la subunidad Bdp1 de TFIIIB (5, 6). Sin embargo, aún se desconocen el determinante primario de Ty1 ISC y su contribución a la protección del genoma frente a inserciones deletéreas.

En un intento por identificar interactores proteicos de Pol III, realizamos cribados sistemáticos de dos híbridos de levadura usando las 17 subunidades Pol III como cebos y una biblioteca de ADN de levadura como presa. Identificamos una interacción específica entre la subunidad AC40 y una región de la proteína TyB/POL de Ty1 que abarca el extremo C de la integrasa (IN) y el extremo N de la transcriptasa inversa (Fig. 1A). El cribado también recuperó Ty2 y Ty4 solapando los mismos dominios proteicos que Ty1, lo que es consistente con preferencias de integración similares para estos 3 elementos (7). AC40 marcada con HA (AC40-HA) coimmunoprecipitaba con IN cuando se expresaba ectópicamente como fusión con estreptavidina (IN-Strep) (Fig. 1B) o a partir de un elemento Ty1 funcional (Fig. 1C).

Ty1 IN tiene dos regiones filogenéticamente conservadas (un dominio de unión a zinc N-terminal y un núcleo catalítico con la unidad DDX₃₅E) y un extremo C menos conservado que contiene una señal de localización nuclear bipartita necesaria para la localización nuclear de IN y una retrotransposición eficiente (Fig. 1A) (8, 9). Un ensayo de dos híbridos confirmó la interacción entre AC40 y IN de longitud total (IN₁₋₆₃₅) y reveló que el extremo N y el núcleo catalítico de IN eran dispensables (IN₁₉₃₋₆₃₅), mientras que los últimos 57 aminoácidos del extremo C de IN eran necesarios y suficientes (Fig. 1A, IN₁₋₅₇₈ frente a IN₅₇₈₋₆₃₅). Las integrasas de varios retrotransposones y retrovirus interactúan con una proteína unida a ADN celular para mediar en ISC (10-16). De forma similar al retrotransposón de levadura Ty5 y al virus de la leucemia murina (MLV), el dominio de abordaje de Ty1 se localiza en el extremo C de IN, lo que sugiere en algunos casos que esta región evolucionó para interactuar con cofactores de abordaje específicos.

Como AC40 es una proteína esencial, abordamos su papel en Ty1 ISC usando el ortólogo *Schizosaccharomyces pombe rpc40+* de *RPC40*, que codifica AC40. El AC40 de *S. pombe* (AC40sp) podía restaurar la viabilidad de células de *S. cerevisiae* carentes de AC40 y mantener una transcripción normal de Pol III (17). Sin embargo, GBD-AC40sp, cuando se expresa a niveles similares a GBD-AC40, no mostraba interacción de dos híbridos con Ty1 IN (Fig. 1A y SID), lo que implica que en su estado nativo, AC40sp y Ty1 IN no interactúan o lo hacen sólo muy escasamente. Por lo tanto, se usó AC40sp como un mutante de pérdida de interacción. Primeramente, determinamos el papel de AC40 en la integración de Ty1 en genes transcritos con Pol III usando un ensayo cuantitativo que se basa en un gen *URA3* localizado secuencia arriba del gen de ARNt de la glicina *tG(GCC)B*, un punto caliente de la integración de Ty1. En este ensayo, las inserciones de Ty1 secuencia arriba de *tG(GCC)B* inactivan *URA3*, dando como resultado células Ura⁻ que crecen en placas 5FOA. Se indujo retrotransposición de Ty1 endógena en cepas *rpc40Δ* que expresaban AC40 o AC40sp a niveles similares cambiando el crecimiento a 20°C, una temperatura permisiva para la retrotransposición de Ty1. Se observó una disminución <3,5 veces en el número de colonias 5FOA^R por célula en presencia de AC40sp (Fig. 2A). La fracción de células 5FOA^R que surgen de la integración de Ty1 en células que expresan AC40sp era <30 veces inferior a en células que expresan AC40 (Fig. 2A y 2B). Por lo tanto, la interacción AC40/IN es determinante para la integración de Ty1 en este gen indicador.

Demostramos además el papel de AC40 en Ty1 ISC analizando inserciones no seleccionadas en el gen de ARNt *tG(GCC)C*, otro punto caliente de la integración de Ty1, en cepas que expresan un elemento Ty1 inducible por

galactosa (Fig. 2C). En presencia de AC40, observamos un patrón de bandas periódico <70 pb característico de inserción de Ty1 en el ADN nucleosómico secuencia arriba de los genes de tARN (3, 4, 18). En cultivos que expresan AC40sp, el patrón se alteró drásticamente, mostrando tanto una disminución significativa en la integración de Ty1 como una aparente periodicidad relajada entre puntos calientes (Fig. 2C). Se observó la misma alteración en el patrón de integración en el locus *tG(GCC)C* en células que expresan AC40sp para elementos Ty1 endógenos. Por ello, la interacción AC40/IN es crítica para la especificidad de la integración de Ty1 secuencia arriba de genes transcritos con Pol III.

Para establecer si la frecuencia de integración de Ty1 depende de la interacción AC40/IN, usamos un elemento Ty1 cromosómico marcado con un gen indicador *his3AI*, que confiere prototrofia His⁺ a las células por retrotransposición. Las frecuencias de células His⁺ generadas en cepas que expresan AC40 o AC40sp eran similares (Fig. 2D). Es concebible que en el mutante de pérdida de interacción una sobrerrepresentación de inserciones de Ty1 *HIS3* pudiera ser debida a recombinación homóloga (HR) con elementos endógenos (19). Determinamos la frecuencia de sucesos de inserción de Ty1 *HIS3* dependiente de IN en un mutante *rad52Δ*, que es deficiente en HR (Fig. 2D). Los sucesos His⁺ eran ligeramente superiores en células mutantes *rad52Δ*, lo cual es consistente con estudios previos que muestran que, aunque la ruta Rad52 contribuye a la HR de Ty1, globalmente inhibe la retrotransposición de Ty1 (20). Notablemente, la frecuencia de integración dependiente de IN es 2 veces inferior en células AC40sp en comparación con células AC40, lo que sugiere que una mayor proporción de inserciones de Ty1 podrían ser debidas a recombinación en ausencia de interacción. Sin embargo, esta ligera disminución no puede explicar el efecto drástico observado en la integración de Ty1 en el gen indicador *URA3-tG(GCC)B*. En conjunto, estos resultados indican que la interacción AC40/IN es necesaria para el patrón de abordaje de Ty1 canónico secuencia arriba de dos genes de ARNt representativos, pero la interacción no influye en la eficacia de la integración. Por lo tanto, la interacción con AC40 podría funcionar específicamente para restringir la integración de Ty1 secuencia arriba de genes transcritos con Pol III para evitar inserciones potencialmente perjudiciales en dianas secundarias en el genoma.

Comparamos los patrones de integración de todo el genoma de Ty1 en las cepas WT y en las mutantes de pérdida de interacción AC40sp. La cepa WT exhibía un patrón de integración típico en genes transcritos con Pol III (3, 4) (Fig. 3A). Por el contrario, el mutante de pérdida de interacción exhibía un perfil fundamentalmente diferente con sólo un ligero sesgo para integrar genes de ARNt próximos. Una interacción residual entre IN y AC40sp podría explicar este modesto sesgo. Para determinar si otras características genómicas influyen en Ty1 ISC en ausencia de la interacción AC40/IN, usamos modelos de regresión logísticos uni- (Fig. 3B) y multidimensionales para asociar las características genómicas con los puntos calientes de integración. Las inserciones de Ty1 en la cepa WT se asociaban a regiones secuencia arriba de genes de ARNt o características asociadas a estos sitios, tales como retrotransposones preexistentes (Fig. 3B). También se favorecieron regiones con nucleosomas bien posicionados y se evitaron marcos abiertos de lectura (ORF). En el mutante de pérdida de interacción AC40sp, los genes de ARNt ya no eran el determinante de abordaje primario, pero los nucleosomas aún se veían favorecidos. Los telómeros y las regiones subteloméricas eran también dianas preferidas en el WT e incluso más en el mutante (Fig. 3B y 3C). Se observó esta propensión a que se produzcan inserciones en dominios heterocromáticos para todos los cromosomas. Es de tener en cuenta que Ty5, un retrotransposón de levadura relacionado, también se dirige a la heterocromatina interaccionando específicamente con la proteína heterocromatina Sir4 (11). Por lo tanto, una interacción entre Ty1 IN y proteínas heterocromatinas podría estar implicada en la preferencia de Ty1 de sitios diana secundarios.

Identificamos una interacción entre Ty1 IN y la subunidad de Pol III AC40 que determina Ty1 ISC secuencia arriba de los genes Pol III. Estudios previos indicaron que la integración de Ty1 requiere un promotor Pol III funcional, y nuestra caracterización de AC40 como cofactor de la integración de Ty1 explica este requerimiento. Los retroelementos Ty3 de *S. cerevisiae* y TRE5-A de *Dictyostelium discoideum* también se integran en genes transcritos por Pol III, pero su abordaje conlleva una interacción con TFIIIB (21, 22). Nuestro estudio demuestra un papel directo de una ARN polimerasa en el abordaje de la integración. En Pol III, AC40 se localiza en la periferia del complejo en estrecha proximidad a la región de ADN secuencia arriba (23). Esta localización podría explicar por qué Ty1 se integra mayormente secuencia arriba de genes transcritos por Pol III y raramente secuencia abajo. Ty1 no se integra cerca de o en genes transcritos por Pol I, que también contiene AC40 (3, 4). La interacción AC40/IN podría requerir cofactor(es) adicional(es) específico(s) para la transcripción por Pol III. Es de observar que AC40 en Pol III está en estrecho contacto con TFIIIB (24), que contiene Bdp1, una proteína necesaria para la integración periódica dependiente de Isw2 de Ty1 (6). Este cofactor podría participar directa o indirectamente en la interacción. Ty1 se integra preferentemente en nucleosomas, y la cromatina en los promotores de Pol III es de estructura relativamente abierta con nucleosomas bien organizados (25, 26). Esta organización no se encuentra en los genes Pol I, que se transcriben directamente o se empaquetan en una estructura nucleosómica apretada y represiva (27) - estados de cromatina que podrían desfavorecer la integración de Ty1.

El complejo de transcripción Pol III, que incluye AC40, está regulado por rutas de señalización de nutrientes y estrés (28). Como resultado, Ty1 exhibe el potencial de actuar como un agente mutagénico en condiciones que afectan a la interacción AC40/IN o a la transcripción de Pol III. Nuestro trabajo revela que Ty1 también se dirige a subtelómeros, especialmente en ausencia de interacción AC40/IN. Los extremos de los cromosomas de levaduras contienen familias de genes de rápida evolución no esenciales generalmente necesarias para responder a cambios medioambientales (29). Por lo tanto, dirigir la integración de Ty1 a los subtelómeros podría además proteger al genoma de la levadura frente a la movilidad de Ty1, promoviendo potencialmente al mismo tiempo la adaptación evolutiva y la innovación

génica en respuesta al estrés.

Entender el mecanismo molecular de la integración dirigida del retroelemento Ty1 ha sido un Santo Grial de la investigación de elementos genéticos móviles durante décadas. Con nuestro estudio, elucidamos este mecanismo molecular indicando que la subunidad AC40 de Pol III es la proteína celular mayor que dirige la integración de Ty1 a los genes Pol III, a través de su interacción con integrasa Ty1. Es importante que la unión de complejos de integración al genoma celular, a través de una interacción entre la integrasa y proteínas celulares unidas en sitios de inserción favorecidos, se conserva con retrovirus MLV y VIH. La principal diferencia es que las integrasas de MLV y VIH interaccionan con factores celulares que dirigen la integración retroviral en el promotor o unidad de transcripción de genes transcritos con Pol II, para dar lugar a una potencial desregulación o inactivación génica tras su integración.

La integración estable de retroelementos en el genoma de la célula huésped constituye una importante ventaja para un sistema de administración génica basado en retrovirus, especialmente para terapia génica, cuyo objetivo es la corrección a largo plazo de defectos genéticos, pero que también produce preocupaciones sobre una potencial mutagénesis y activación de oncogenes. La terapia génica con vectores retrovirales derivados de MLV para tratar una inmunodeficiencia severa obtuvo un impresionante éxito terapéutico. Sin embargo, el desarrollo de leucemia causada por mutagénesis de inserción en algunos pacientes descubrió un grave efecto colateral no anticipado debido a la transactivación de oncogenes adyacentes por la actividad potenciadora de la LTR viral. Un avance clave fue la creación de vectores "autoinactivantes" (SIN) más sofisticados, diseñados para evitar tales sucesos de transactivación tras la integración. Sin embargo, estos vectores aún se integran cerca de protooncogenes, manteniendo una espada de Damocles sobre la seguridad de dichos vectores. Por lo tanto, una mejora potencial para el futuro es dirigir los vectores de terapia génica a regiones seguras del genoma. El mecanismo de abordaje de Ty1 secuenciado arriba de genes Pol III individualmente no esenciales proporciona sitios de aterrizaje seguros que previenen las consecuencias perjudiciales sobre la idoneidad celular, y es, por lo tanto, una alternativa prometedora a los vectores basados en retrovirus ya existentes diseñados para terapia génica humana. Por consiguiente, el dominio de Ty1 responsable de la interacción con la ARN polimerasa III podría así ser adecuado para la ingeniería de integrasas quiméricas de retrovirus, para llevar la integración dirigida de un transgén a regiones seguras de un genoma eucariótico, con objeto de reducir el potencial efecto perjudicial.

Ejemplo 2: Caracterización del dominio de abordaje de Ty1 IN

Utilizamos el enfoque de dos híbridos como se representa en la Figura 4 para caracterizar más el dominio TY1 responsable de la interacción con AC40: si se produce la interacción, entonces se activa el HIS3 y se puede producir el crecimiento en ausencia de histidina; si no se produce ninguna interacción, entonces no hay activación de HIS3 y, por consiguiente, no se puede producir el crecimiento en ausencia de histidina. Brevemente, el diseño experimental era como sigue: se transforma la cepa PJ69-4A con los 2 vectores pAS2ΔΔ y pACTII, que contienen o bien GBD-AC40 o bien cualquiera de las construcciones GAD-IN estudiadas en los ejemplos. Se realizaron cultivos durante la noche a 30°C. Se plaquearon diluciones seriadas de factor 10 de alícuotas de 1 DO600, lavadas en 1 ml de H₂O, en medio HC carente de leucina y triptófano (control) o leucina, triptófano e histidina (interacción). Las prototofías de leucina y triptófano garantizan la presencia de los 2 vectores. Se incubaron las placas durante 3 días a temperatura ambiente. Como se muestra en la Figura 5, la secuencia mínima responsable de la interacción de IN578_635 con AC40 es la secuencia que va del residuo de aminoácido en la posición 617 a 623. Demostramos también que los residuos de aminoácido en la posición 618 ó 623 pueden sustituirse. En la Figura 6, mostramos que la secuencia KNMRSLEPP puede también interactuar con AC40. En la figura 7, reemplazamos en GAD-IN578-635, la secuencia 617KNMRSLE622 por KEMDSL (los aminoácidos que se modificaron están subrayados) para demostrar que la secuencia KEMDSL también interactúa con AC40.

Referencias:

A través de toda esta solicitud, diversas referencias describen el estado de la técnica a la que esta descripción pertenece.

1. H. L. Levin, J. V Moran, Nat Rev Genet 12, 615-627 (2011).
2. S. E. Devine, J. D. Boeke, Genes Dev. 10, 620-633 (1996).
3. J. A. Baller, J. Gao, R. Stamenova, M. J. Curcio, D. F. Voytas, Genome Res. 22, 704-13 (2012).
4. L. Mularoni *et al.*, Genome Res 22, 693-703 (2012).
5. M. E. Gelbart, N. Bachman, J. Delrow, J. D. Boeke, T. Tsukiyama, Genes Dev 19, 942-954 (2005).
6. N. Bachman, M. E. Gelbart, T. Tsukiyama, J. D. Boeke, Genes Dev 19, 955-964 (2005).
7. J. M. Kim, S. Vanguri, J. D. Boeke, A. Gabriel, D. F. Voytas, Genome Res. 8, 464-478 (1998).
8. M. Kenna, C. Brachmann, S. Devine, J. Boeke, Mol Cell Biol 18, 1115-1124 (1998).
9. S. P. Moore, L. A. Rinckel, D. J. Garfinkel, Mol Cell Biol 18, 1105-1114 (1998).

10. J. Kirchner, C. M. Connolly, S. B. Sandmeyer, *Science* 267, 1488-1491 (1995).
11. W. Xie *et al.*, *Mol Cell Biol* 21, 6606-6614 (2001).
12. Y.-E. E. Leem *et al.*, *Mol Cell* 30, 98-107 (2008).
13. A. Engelman, P. Cherepanov, *PLoS Pathog* 4, e1000046 (2008).
- 5 14. J. De Rijck *et al.*, *Cell Rep.* 5, 886-894 (2013).
15. S. S. Gupta *et al.*, *J. Virol.* 87, 12721-36 (2013).
16. A. Sharma *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 110, 12036-41 (2013).
17. G. V Shpakovski, E. K. Shematorova, *Curr Genet* 36, 208-214 (1999).
18. N. Bachman, Y. Eby, J. D. Boeke, *Genome Res* 14, 1232-1247 (2004).
- 10 19. G. Sharon, T. J. Burkett, D. J. Garfinkel, *Mol Cell Biol* 14, 6540-51. (1994).
20. A. J. Ratray, B. K. Shafer, D. J. Garfinkel, *Genetics* 154, 543-56. (2000).
21. L. Yieh, G. Kassavetis, E. P. Geiduschek, S. B. Sandmeyer, *J Biol Chem* 275, 29800-29807 (2000).
22. T. Chung, O. Siol, T. Dingermann, T. Winckler, *Mol. Cell. Biol.* 27, 8492-501 (2007).
23. C. Fernández-Tornero *et al.*, *Nature* 502, 644-9 (2013).
- 15 24. A. Vannini, P. Cramer, *Mol. Cell* 45, 439-46 (2012).
25. K. Brogaard, L. Xi, J.-P. Wang, J. Widom, *Nature* 486, 496-501 (2012).
26. Y. Kumar, P. Bhargava, *BMC Genomics* 14, 402 (2013).
27. K. Merz *et al.*, *Genes Dev.* 22, 1190-204 (2008).
28. R. D. Moir, I. M. Willis, *Biochim. Biophys. Acta* 1829, 361-75 (2013).
- 20 29. C. A. Brown, A. W. Murray, K. J. Verstrepen, *Curr. Biol.* 20, 895-903 (2010).
30. N. Kaplan *et al.*, *Nature* 458, 362-366 (2009).

LISTA DE SECUENCIAS

<110> Interm

5 <120> POLIPÉPTIDOS PARA LA INGENIERÍA DE PROTEÍNAS INTEGRASAS QUIMÉRICAS Y SU USO EN TERAPIA GÉNICA

<130> BIO 15039 LESAGE / MC

10 <160> 17

<170> PatentIn versión 3.3

15 <210> 1
<211> 635
<212> PRT
<213> *Saccharomyces cerevisiae*

20 <400> 1
Asn Val His Thr Ser Glu Ser Thr Arg Lys Tyr Pro Tyr Pro Phe Ile
1 5 10 15

His Arg Met Leu Ala His Ala Asn Ala Gln Thr Ile Arg Tyr Ser Leu
20 25 30

Lys Asn Asn Thr Ile Thr Tyr Phe Asn Glu Ser Asp Val Asp Trp Ser
35 40 45

Ser Ala Ile Asp Tyr Gln Cys Pro Asp Cys Leu Ile Gly Lys Ser Thr
50 55 60

Lys His Arg His Ile Lys Gly Ser Arg Leu Lys Tyr Gln Asn Ser Tyr
65 70 75 80

Glu Pro Phe Gln Tyr Leu His Thr Asp Ile Phe Gly Pro Val His Asn
85 90 95

Leu Pro Asn Ser Ala Pro Ser Tyr Phe Ile Ser Phe Thr Asp Glu Thr
100 105 110

Thr Lys Phe Arg Trp Val Tyr Pro Leu His Asp Arg Arg Glu Asp Ser
115 120 125

Ile Leu Asp Val Phe Thr Thr Ile Leu Ala Phe Ile Lys Asn Gln Phe
130 135 140

Gln Ala Ser Val Leu Val Ile Gln Met Asp Arg Gly Ser Glu Tyr Thr
145 150 155 160

Asn Arg Thr Leu His Lys Phe Leu Glu Lys Asn Gly Ile Thr Pro Cys
165 170 175

ES 2 777 534 T3

Tyr Thr Thr Thr Ala Asp Ser Arg Ala His Gly Val Ala Glu Arg Leu
 180 185 190

Asn Arg Thr Leu Leu Asp Asp Cys Arg Thr Gln Leu Gln Cys Ser Gly
 195 200 205

Leu Pro Asn His Leu Trp Phe Ser Ala Ile Glu Phe Ser Thr Ile Val
 210 215 220

Arg Asn Ser Leu Ala Ser Pro Lys Ser Lys Lys Ser Ala Arg Gln His
 225 230 235 240

Ala Gly Leu Ala Gly Leu Asp Ile Ser Thr Leu Leu Pro Phe Gly Gln
 245 250 255

Pro Val Ile Val Asn Asp His Asn Pro Asn Ser Lys Ile His Pro Arg
 260 265 270

Gly Ile Pro Gly Tyr Ala Leu His Pro Ser Arg Asn Ser Tyr Gly Tyr
 275 280 285

Ile Ile Tyr Leu Pro Ser Leu Lys Lys Thr Val Asp Thr Thr Asn Tyr
 290 295 300

Val Ile Leu Gln Gly Lys Glu Ser Arg Leu Asp Gln Phe Asn Tyr Asp
 305 310 315 320

Ala Leu Thr Phe Asp Glu Asp Leu Asn Arg Leu Thr Ala Ser Tyr His
 325 330 335

Ser Phe Ile Ala Ser Asn Glu Ile Gln Glu Ser Asn Asp Leu Asn Ile
 340 345 350

Glu Ser Asp His Asp Phe Gln Ser Asp Ile Glu Leu His Pro Glu Gln
 355 360 365

Pro Arg Asn Val Leu Ser Lys Ala Val Ser Pro Thr Asp Ser Thr Pro
 370 375 380

Pro Ser Thr His Thr Glu Asp Ser Lys Arg Val Ser Lys Thr Asn Ile
 385 390 395 400

Arg Ala Pro Arg Glu Val Asp Pro Asn Ile Ser Glu Ser Asn Ile Leu
 405 410 415

Pro Ser Lys Lys Arg Ser Ser Thr Pro Gln Ile Ser Asn Ile Glu Ser
 420 425 430

ES 2 777 534 T3

Thr Gly Ser Gly Gly Met His Lys Leu Asn Val Pro Leu Leu Ala Pro
435 440 445

Met Ser Gln Ser Asn Thr His Glu Ser Ser His Ala Ser Lys Ser Lys
450 455 460

Asp Phe Arg His Ser Asp Ser Tyr Ser Glu Asn Glu Thr Asn His Thr
465 470 475 480

Asn Val Pro Ile Ser Ser Thr Gly Gly Thr Asn Asn Lys Thr Val Pro
485 490 495

Gln Ile Ser Asp Gln Glu Thr Glu Lys Arg Ile Ile His Arg Ser Pro
500 505 510

Ser Ile Asp Ala Ser Pro Pro Glu Asn Asn Ser Ser His Asn Ile Val
515 520 525

Pro Ile Lys Thr Pro Thr Thr Val Ser Glu Gln Asn Thr Glu Glu Ser
530 535 540

Ile Ile Ala Asp Leu Pro Leu Pro Asp Leu Pro Pro Glu Ser Pro Thr
545 550 555 560

Glu Phe Pro Asp Pro Phe Lys Glu Leu Pro Pro Ile Asn Ser Arg Gln
565 570 575

Thr Asn Ser Ser Leu Gly Gly Ile Gly Asp Ser Asn Ala Tyr Thr Thr
580 585 590

Ile Asn Ser Lys Lys Arg Ser Leu Glu Asp Asn Glu Thr Glu Ile Lys
595 600 605

Val Ser Arg Asp Thr Trp Asn Thr Lys Asn Met Arg Ser Leu Glu Pro
610 615 620

Pro Arg Ser Lys Lys Arg Ile His Leu Ile Ala
625 630 635

<210> 2
<211> 6
<212> PRT
<213> Artificial

<220>
<223> dominio de interacción con AC40

<400> 2
Lys Asn Met Arg Ser Leu
1 5

<210> 3
<211> 6
<212> PRT
<213> Artificial

<220>
<223> dominio de interacción con AC40

<400> 3
 Lys Glu Met Asp Ser Leu
 1 5

5 <210> 4
 <211> 7
 <212> PRT
 <213> Artificial

10 <220>
 <223> dominio de interacción con AC40

<400> 4
 Lys Asn Met Arg Ser Leu Glu
 1 5

15 <210> 5
 <211> 7
 <212> PRT
 <213> Artificial

20 <220>
 <223> dominio de interacción con AC40

<400> 5
 Lys Ala Met Arg Ser Leu Glu
 1 5

25 <210> 6
 <211> 7
 <212> PRT
 <213> Artificial

30 <220>
 <223> dominio de interacción con AC40

35 <400> 6
 Lys Asn Met Arg Ser Leu Ala
 1 5

40 <210> 7
 <211> 9
 <212> PRT
 <213> Artificial

45 <220>
 <223> dominio de interacción con AC40

<400> 7
 Lys Asn Met Arg Ser Leu Glu Pro Pro
 1 5

50 <210> 8
 <211> 15
 <212> PRT
 <213> Artificial

55 <220>
 <223> dominio de interacción con AC40

<400> 8
 Val Ser Arg Asp Thr Trp Asn Thr Lys Asn Met Arg Ser Leu Glu
 1 5 10 15

<210> 9
 <211> 27
 <212> PRT
 5 <213> Artificial

 <220>
 <223> dominio de interacción con AC40

 10 <400> 9
 Ser Leu Glu Asp Asn Glu Thr Glu Ile Lys Val Ser Arg Asp Thr Trp
 1 5 10 15

 Asn Thr Lys Asn Met Arg Ser Leu Glu Pro Pro
 20 25

 <210> 10
 <211> 36
 15 <212> PRT
 <213> Artificial

 <220>
 <223> dominio de interacción con AC40
 20
 <400> 10
 Ser Lys Lys Arg Ser Leu Glu Asp Asn Glu Thr Glu Ile Lys Val Ser
 1 5 10 15

 Arg Asp Thr Trp Asn Thr Lys Asn Met Arg Ser Leu Glu Pro Pro Arg
 20 25 30

 Ser Lys Lys Arg
 35
 25
 <210> 11
 <211> 41
 <212> PRT
 <213> Artificial
 30
 <220>
 <223> dominio de interacción con AC40

 <400> 11
 Ser Lys Lys Arg Ser Leu Glu Asp Asn Glu Thr Glu Ile Lys Val Ser
 1 5 10 15

 Arg Asp Thr Trp Asn Thr Lys Asn Met Arg Ser Leu Glu Pro Pro Arg
 20 25 30

 Ser Lys Lys Arg Ile His Leu Ile Ala
 35 35 40

 <210> 12
 <211> 58
 <212> PRT
 40 <213> Artificial

 <220>
 <223> dominio de interacción con AC40

 45 <400> 12

ES 2 777 534 T3

Asn Ser Ser Leu Gly Gly Ile Gly Asp Ser Asn Ala Tyr Thr Thr Ile
1 5 10 15

Asn Ser Lys Lys Arg Ser Leu Glu Asp Asn Glu Thr Glu Ile Lys Val
20 25 30

Ser Arg Asp Thr Trp Asn Thr Lys Asn Met Arg Ser Leu Glu Pro Pro
35 40 45

Arg Ser Lys Lys Arg Ile His Leu Ile Ala
50 55

<210> 13

<211> 288

<212> PRT

<213> Virus de inmunodeficiencia humana

<400> 13

Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His
1 5 10 15

Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val

ES 2 777 534 T3

20	25	30
Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu		
35	40	45
Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp		
50	55	60
Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala		
65	70	75
Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu		
85	90	95
Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr		
100	105	110
Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala		
115	120	125
Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn		
130	135	140
Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys		
145	150	155
Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val		
165	170	175
Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly		
180	185	190
Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile		
195	200	205
Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg		
210	215	220
Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys		
225	230	235
Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp		
245	250	255
Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly		
260	265	270
Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp		
275	280	285

5 <210> 14
 <211> 407
 <212> PRT
 <213> Virus del sarcoma murino de Moloney

10 <400> 14

ES 2 777 534 T3

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

ES 2 777 534 T3

Ala Tyr Arg Pro Gln Ser Ser Gly Gln Val Glu Arg Met Asn Arg Thr
210 215 220

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

Trp Val Leu Leu Leu Pro Leu Ala Leu Tyr Arg Ala Arg Asn Thr Pro
245 250 255

Gly Pro His Gly Leu Thr Pro Tyr Glu Ile Leu Tyr Gly Ala Pro Pro
260 265 270

Pro Leu Val Asn Phe Pro Asp Pro Asp Met Thr Arg Val Thr Asn Ser
275 280 285

Pro Ser Leu Gln Ala His Leu Gln Ala Leu Tyr Leu Val Gln His Glu
290 295 300

Val Trp Arg Pro Leu Ala Ala Ala Tyr Gln Glu Gln Leu Asp Arg Pro
305 310 315 320

Val Val Pro His Pro Tyr Arg Val Gly Asp Thr Val Trp Val Arg Arg
325 330 335

His Gln Thr Lys Asn Leu Glu Pro Arg Trp Lys Gly Pro Tyr Thr Val
340 345 350

Leu Leu Thr Thr Pro Thr Ala Leu Lys Val Asp Gly Ile Ala Ala Trp
355 360 365

Ile His Ala Ala His Val Lys Ala Ala Asp Pro Gly Gly Gly Pro Ser
370 375 380

Ser Arg Leu Thr Trp Arg Val Gln Arg Ser Gln Asn Pro Leu Lys Ile
385 390 395 400

Arg Leu Thr Arg Glu Ala Pro
405

<210> 15
<211> 21
<212> PRT
<213> Artificial

<220>
<223> dominio de interacciónTy5

<400> 15
Glu Ser Pro Pro Ser Leu Asp Ser Ser Pro Pro Asn Thr Ser Phe Asn
1 5 10 15

Ala Ala Leu Thr Ala
20

<210> 16
<211> 27
<212> PRT

<213> Artificial

<220>

<223> dominio de interacción Ty4

5

<400> 16

Ile Glu Ala Ser Gly Ser Pro Val Gln Thr Val Asn Lys Ser Ala Phe
1 5 10 15

Leu Asn Lys Glu Phe Ser Ser Leu Asn Met Lys
20 25

<210> 17

10

<211> 9

<212> PRT

<213> Artificial

<220>

15

<223> dominio de interacción con AC40

<400> 17

Trp Asn Thr Lys Asn Met Arg Ser Leu
1 5

20

REIVINDICACIONES

1. Un polipéptido que consiste en la secuencia de aminoácidos seleccionada del grupo consistente en la SEQ ID NO:2 a la SEQ ID NO:12.
- 5 2. Una integrasa quimérica que comprende el polipéptido de la reivindicación 1 y que se selecciona del grupo consistente en integrasas del virus de la inmunodeficiencia felina (FIV), integrasas del virus Espumoso (FV), integrasas del virus de la leucemia murina (MLV) e integrasas de lentivirus.
3. La integrasa quimérica de la reivindicación 2, en donde la integrasa retrovítica tiene una secuencia que tiene al menos un 80% de identidad con la SEQ ID NO: 13 o la SEQ ID NO:14.
- 10 4. La integrasa quimérica de la reivindicación 2, en donde el polipéptido de la reivindicación 1 está fusionado al extremo C-terminal de la proteína integrasa o al extremo N-terminal de la proteína integrasa.
5. La integrasa quimérica de la reivindicación 2, en donde el polipéptido de la reivindicación 1 está insertado en la secuencia codificante de la proteína integrasa.
6. Una molécula de ácido nucleico que codifica un polipéptido de la reivindicación 1 o una integrasa quimérica de la reivindicación 2.
- 15 7. Un vector retrovítico que comprende la integrasa quimérica de la reivindicación 2.
8. El vector retrovítico de la reivindicación 7, que deriva de un retrovirus seleccionado del grupo consistente en alfa-retrovirus, beta-retrovirus, gamma-retrovirus, delta-retrovirus, épsilon-retrovirus, lentivirus y espumavirus.
9. El vector retrovítico de la reivindicación 7, que comprende un transgén.
- 20 10. Un método *in vitro* para expresar un transgén en el genoma de una pluralidad de células, que comprende infectar a la pluralidad de células con el vector retrovítico de la reivindicación 7.
11. El vector retrovítico de la reivindicación 7 para uso en terapia génica.

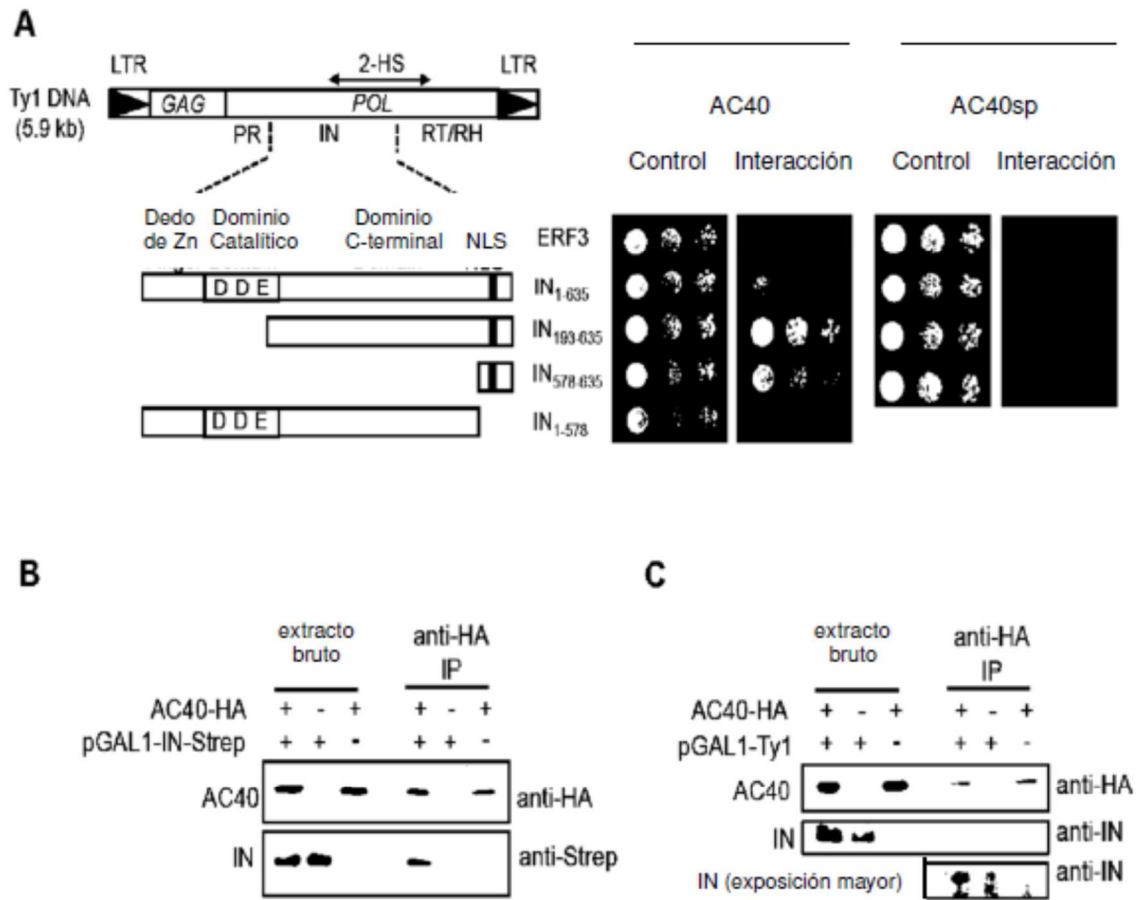


Figura 1

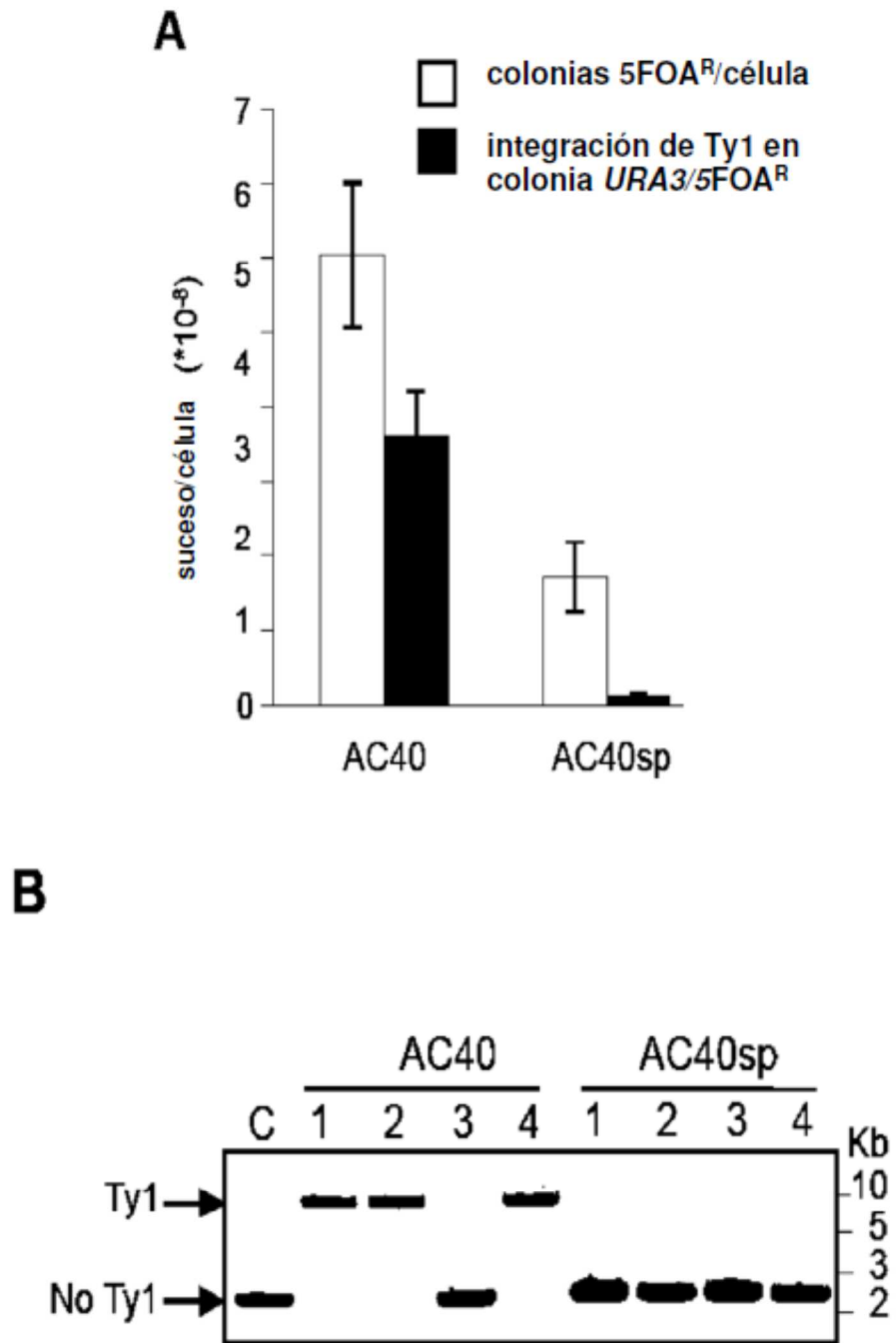


Figura 2 A y B

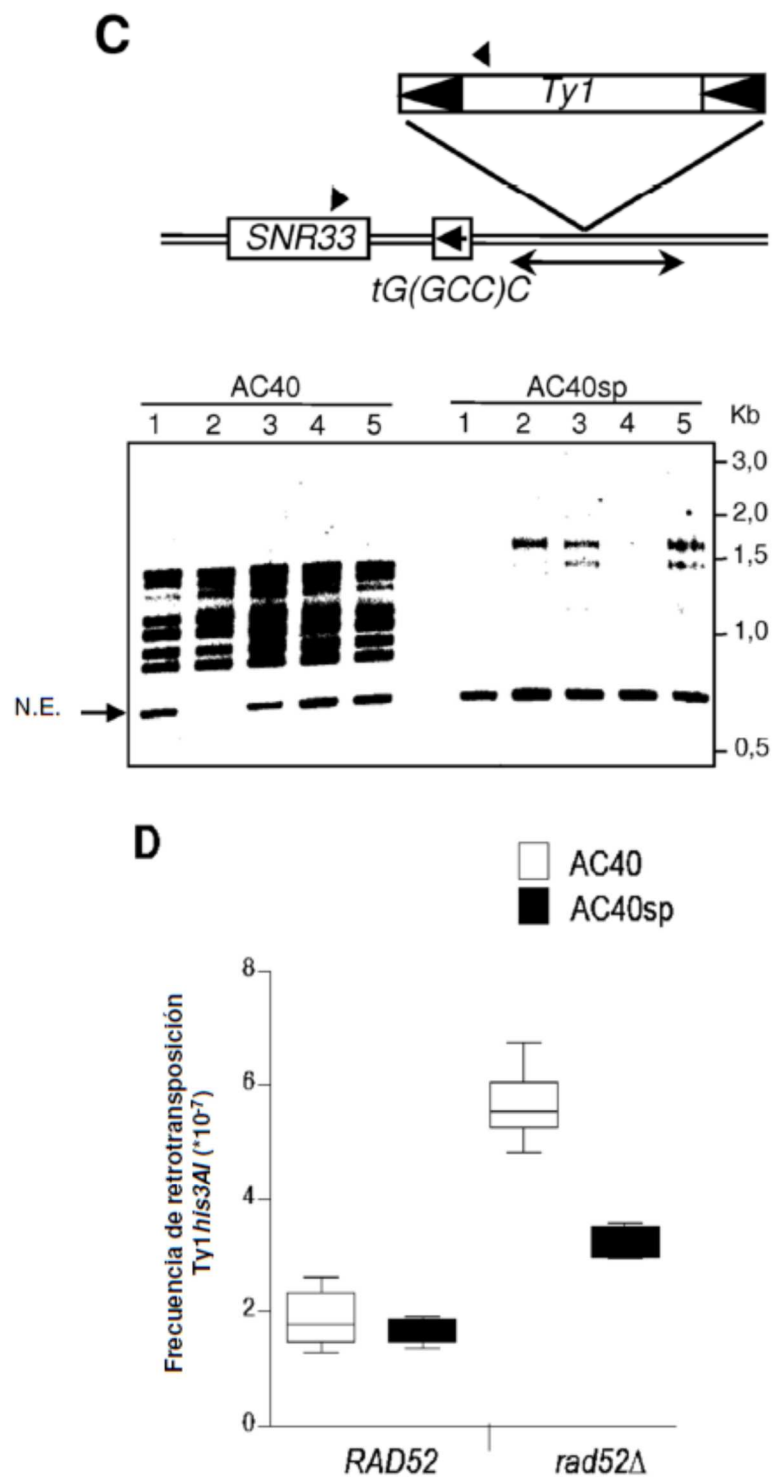


Figura 2 C y D

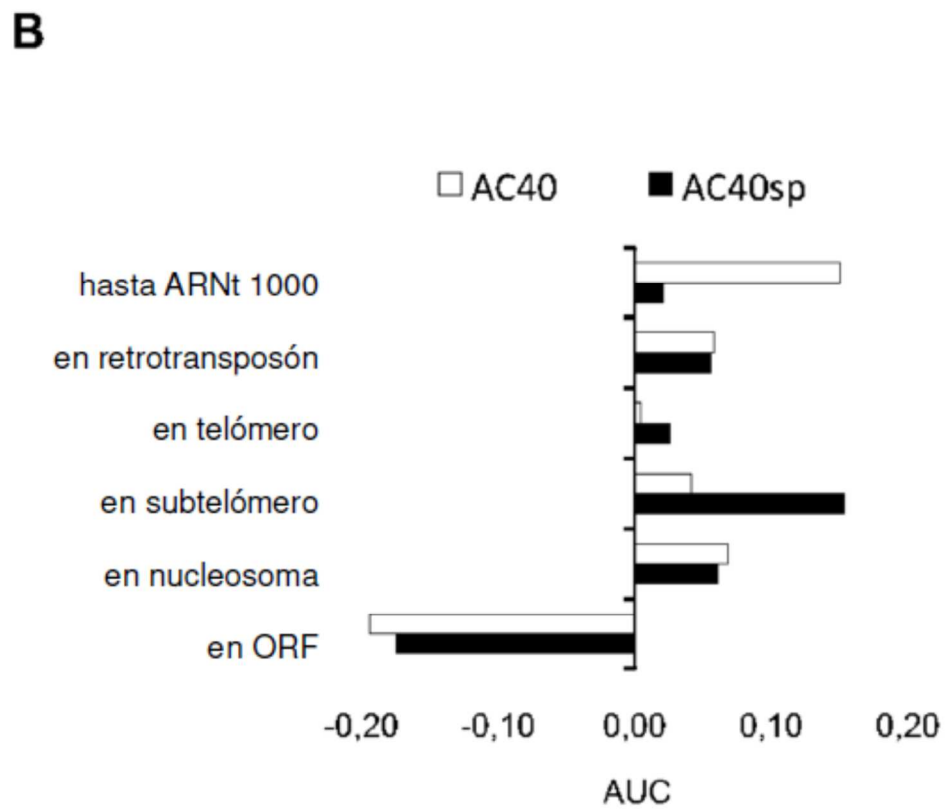
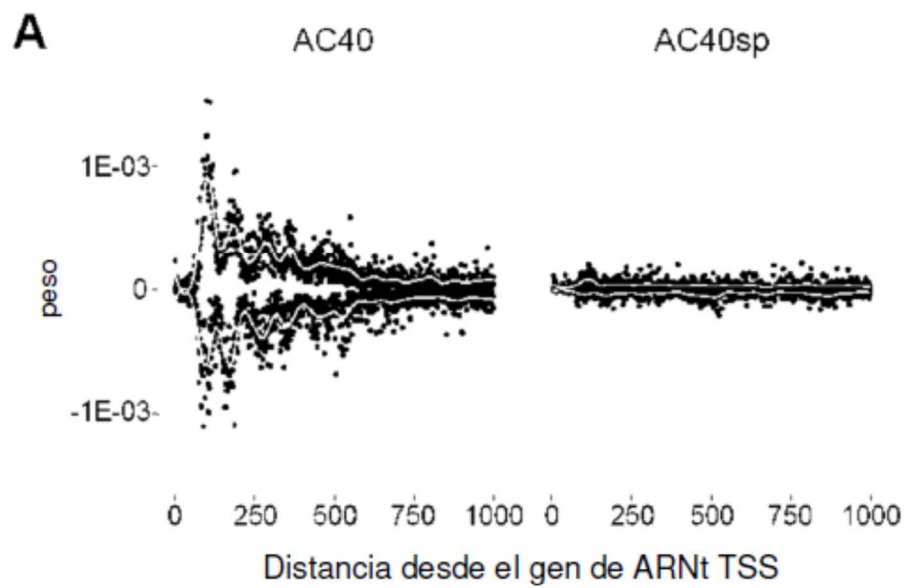


Figura 3 A y B

C

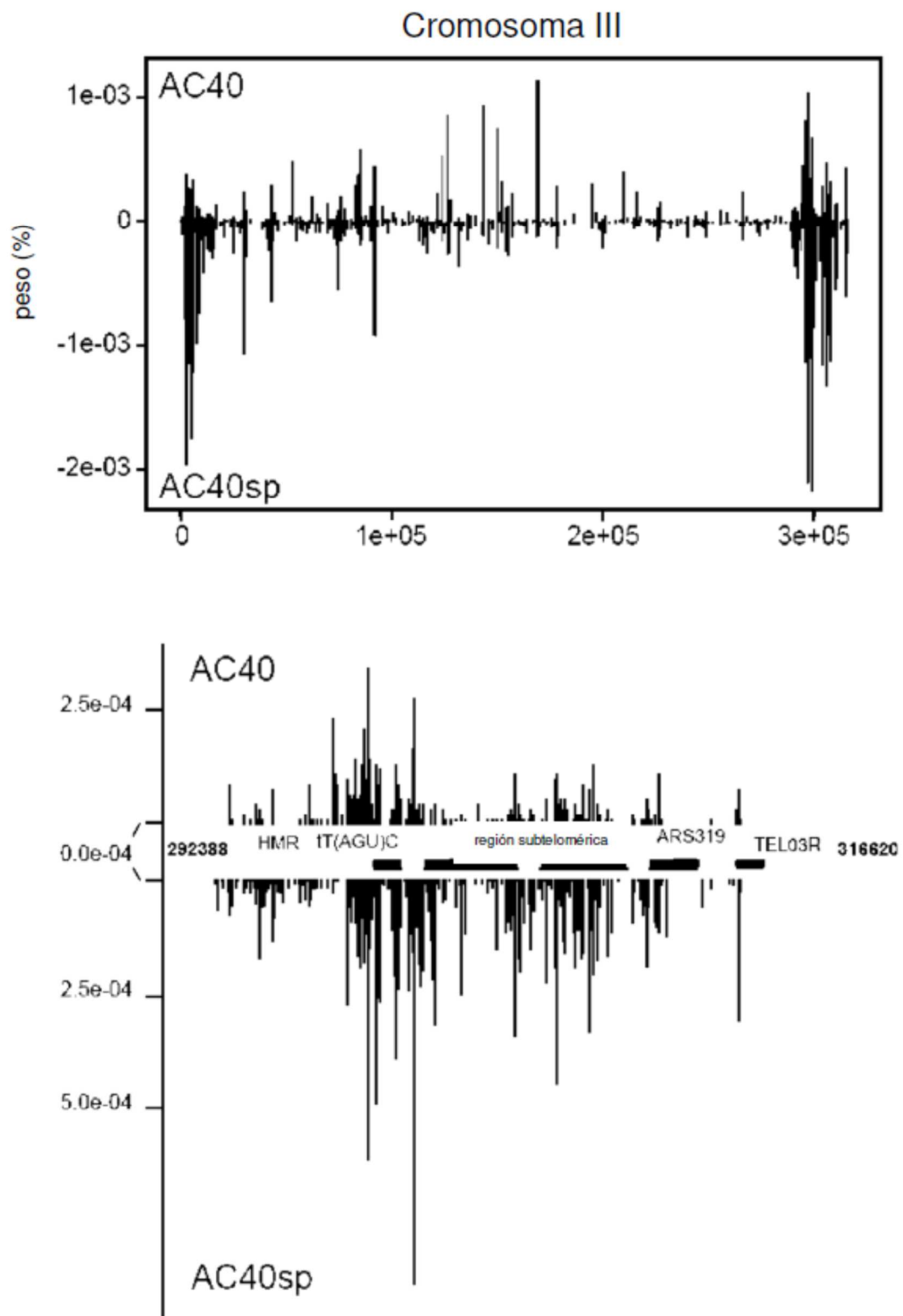
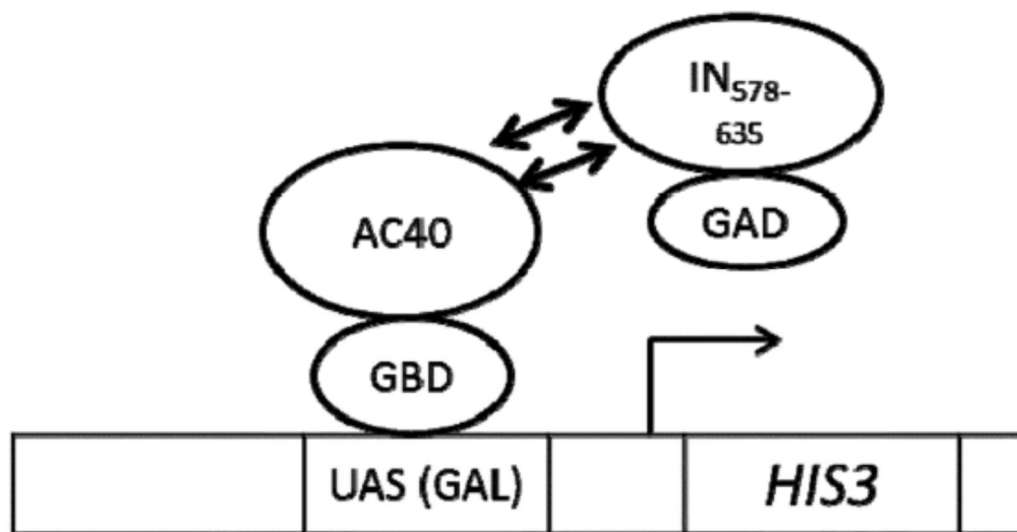


Figura 3C



Si hay interacción: HIS3 activado, crecimiento en ausencia de histidina

Si no hay interacción: no hay activación de HIS3, no hay crecimiento en ausencia de histidina

Vector usado

pAS2ΔΔ (fusión GBD expresada a partir de promotor ADH1, marcador de selección TRP1) pAS2 (fusión GAD expresada a partir de promotor ADH1, marcador de selección LEU2) (Clontech, Fromont-Racine, Nat genet 1997)

Cepa de levadura:

MATa, *trp1-901*, *leu2,3-112*, *ura3-52*, *his3Δ200*, *gal4Δ*, *gal80Δ*, *LYS2::GAL1-HIS3*, *GAL2-ADE2*, *met2::GAL7-lacZ*

James P Genetics (144, pp1425-1436), 1996

Figura 4

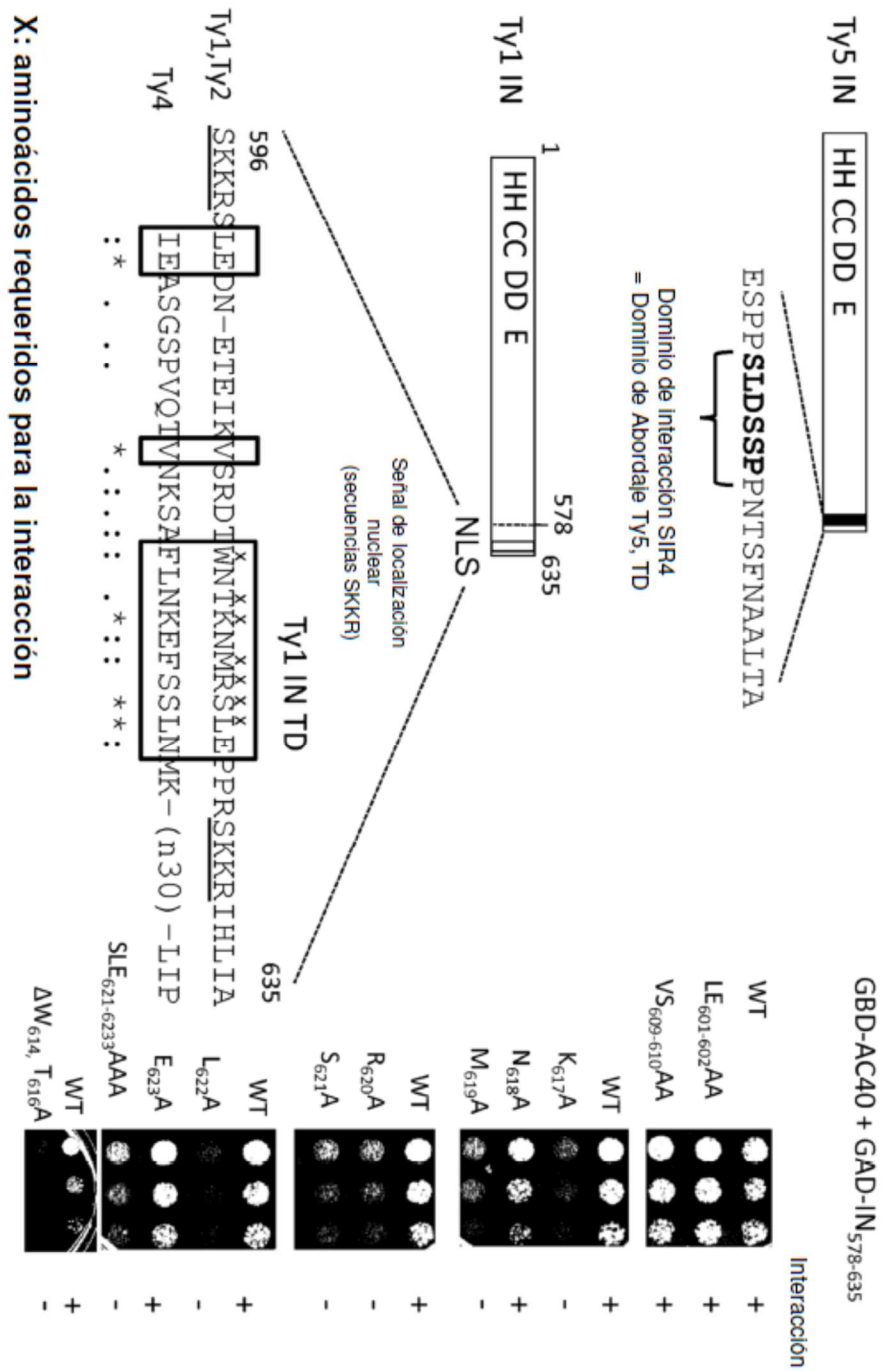
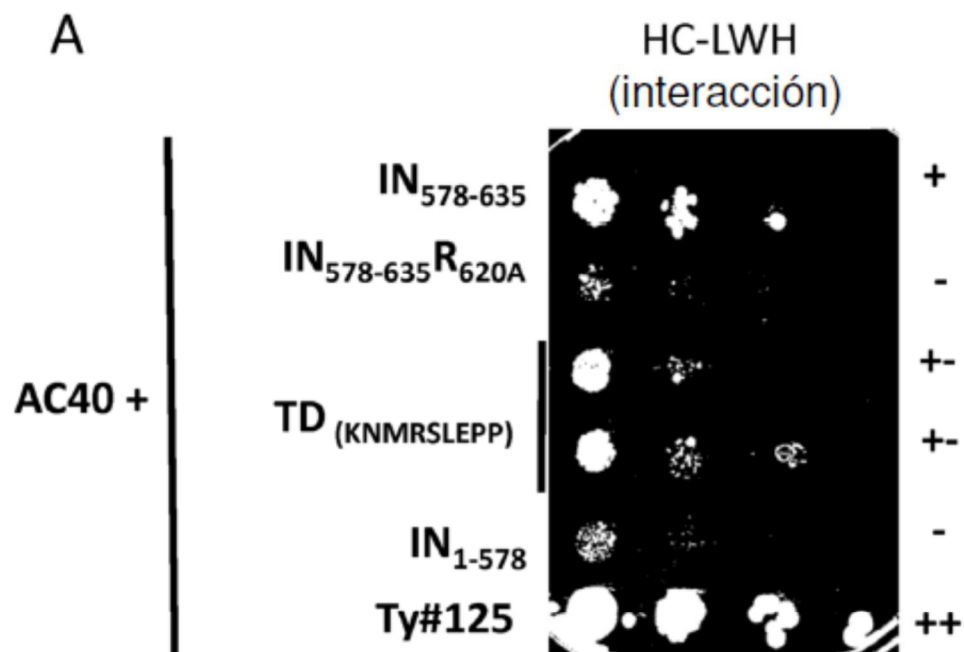


Figura 5



La secuencia KNMRSLEPP interacciona débilmente

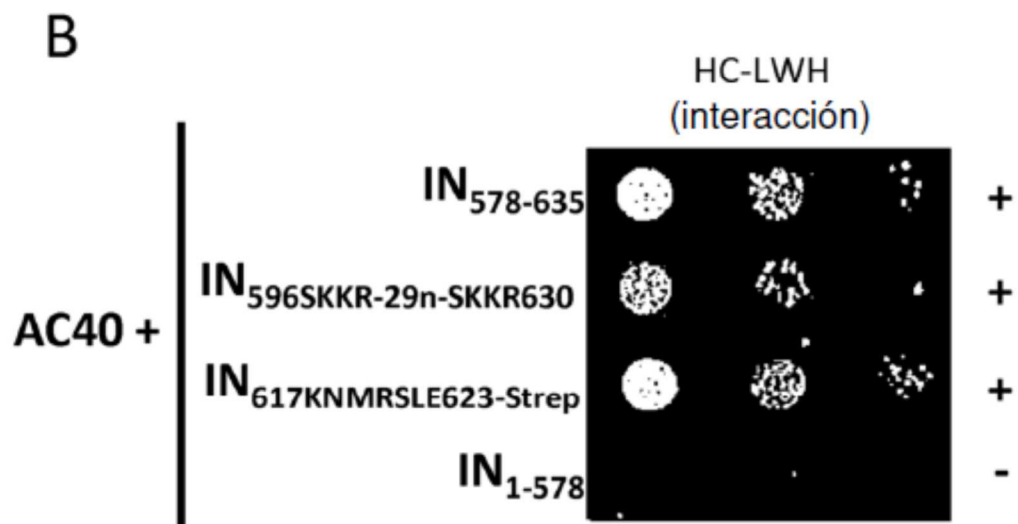


Figura 6 A y B



Figura 7

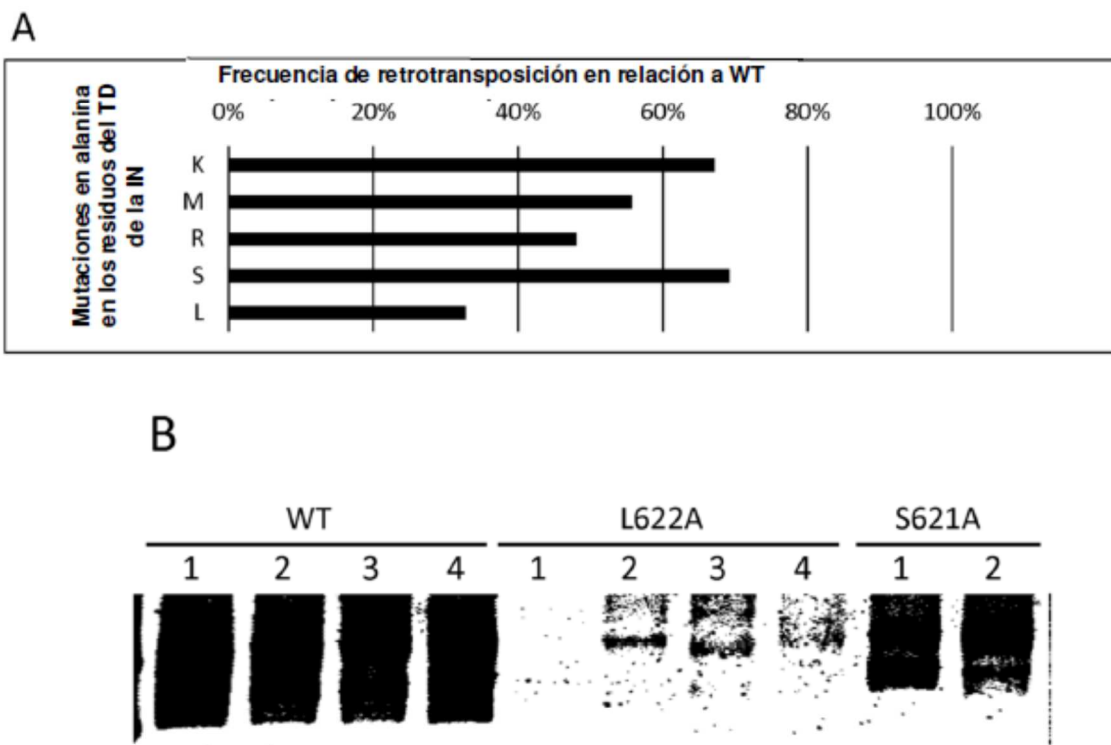


Figura 8 A y B