

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 751 919**

51 Int. Cl.:

**C12Q 1/6869** (2008.01)

**C12Q 1/6851** (2008.01)

**C12Q 1/70** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.11.2015 PCT/EP2015/077882**

87 Fecha y número de publicación internacional: **02.06.2016 WO16083560**

96 Fecha de presentación y número de la solicitud europea: **27.11.2015 E 15801422 (5)**

97 Fecha y número de publicación de la concesión europea: **07.08.2019 EP 3224376**

54 Título: **Impurezas de ADN en una composición que comprende un virión parvoviral**

30 Prioridad:

**28.11.2014 EP 14195464**

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

**02.04.2020**

73 Titular/es:

**UNIQUE IP B.V. (100.0%)  
Paasheuvelweg 25  
1105 BP Amsterdam , NL**

72 Inventor/es:

**LUBELSKI, JACEK y  
HERMENS, WILHELMUS THEODORUS  
JOHANNES M. C.**

74 Agente/Representante:

**TOMAS GIL, Tesifonte Enrique**

ES 2 751 919 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

Impurezas de ADN en una composición que comprende un virión parvoviral

5 Campo de la invención

[0001] La presente invención se refiere al campo de virología y la terapia genética. En particular, la invención se refiere a un método para identificar y/o cuantificar una impureza de ácido nucleico sobrerrepresentada en una composición.

10

Antecedentes de la invención

[0002] Los vectores recombinantes basados en un virus adenoasociado (AAV) se han usado en numerosos ensayos clínicos y conllevan una gran promesa para la terapia genética humana. Las características de principio responsables del amplio éxito del rAAV son su capacidad para establecer una expresión de transgén persistente combinada con buenas características de seguridad. La preparación de rAAV clínico está, entre otras cosas, centrada en la minimización de impurezas de ADN, que pueden codificar potencialmente oncogenes, marcadores antibióticos o péptidos inmunogénicos que comprometen la seguridad del vector (Wright *et al.*, 2008, Gene Ther. 15, 840-848).

20

[0003] Aunque, se ha llevado a cabo un formidable progreso tanto en el procesamiento corriente arriba como corriente abajo del rAAV para asegurar la pureza del producto final, la eliminación completa de ADN no deseado no parece plausible. El empaquetamiento de ADN celular o de vector auxiliar parece ser un subproducto de la biología de AAV y, como tal, está inherentemente enlazado a la encapsidación del ADN de transgén previsto. Una vez que el ADN no relacionado se encapsida en cápsides preformadas, las partículas se vuelven indistinguibles de las cápsides que contienen solo el casete de expresión previsto y es prácticamente imposible separarlas.

25

[0004] Por lo tanto, para apoyar el desarrollo clínico y entender los riesgos relacionados con la presencia de ADN no deseado en vectores rAAV, existe una necesidad de evaluar el potencial para expresión de proteína de estos elementos genéticos en un rango de líneas celulares que refleje el perfil de biodistribución del rAAV usado, para excluir la posibilidad teórica de que una expresión de proteína involuntaria podría tener efectos indeseables tales como, por ejemplo, reestructuraciones celulares, tumorigenicidad o respuestas inmunológicas no deseadas que comprometan potencialmente la seguridad y/o eficacia del vector.

30

35

[0005] Hasta el momento, la presencia y la concentración de impurezas de ADN que contaminan preparaciones de rAAV se ha mencionado en la literatura (Blouin *et al.*, 2004, J. Gene Med. 6 Suppl 1, S223-S228; Nony *et al.*, 2003, J. Virol. 77, 776-781; Chadeuf *et al.*, 2005, Mol. Ther. 12, 744-753; Wright *et al.*, 2008, *supra*). La identidad de estas impurezas se rastreó hasta ADN de plásmido auxiliar o célula huésped (Wright *et al.*, 2008, *supra*). Solo un número limitado de estudios ha abordado la cuestión de la expresión de proteína putativa originada de este ADN residual coempaquetado. Wright y colaboradores han analizado la expresión de cap, amp(r) y dos genes adenovirales E2A y E4 mediante RT-qPCR tras la infección de hepatocitos humanos o de ratones con rAAV y no han descubierto ninguna transcripción detectable (Hauck *et al.*, 2009, Mol Ther. 17(1) 144-152). Por el contrario, Miller *et al.*, han descubierto expresión conducida por impureza de ADN del gen cap usando un ensayo de complementación (Halbert *et al.*, 2011, Gene Ther. 18(4): 411-417).

40

45

[0006] El método actual elegido para analizar las impurezas de ADN en las preparaciones de virión biofarmacéuticas es la qPCR. Usando este método, se determina la presencia y cantidad de una impureza de ADN específica. De manera importante, el experto en la materia selecciona de este modo de antemano la impureza de ADN por detectar, es decir, antes de llevar a cabo una qPCR, dado que existe un consenso general en la técnica con respecto a que las impurezas de ADN se empaquetan de forma aleatoria en el virión. Dicha impureza de ADN (preseleccionada) puede, por ejemplo, comprender ADN de célula huésped, Rep, Cap, o secuencias de nucleótidos de plásmido. Por ejemplo, como indican Thorne *et al.* (2009, Hum. Gene Ther. 20: 707 - 714) las impurezas de ADN de célula huésped se monitorean usando dos objetivos: los genes de transformación de papilomavirus humano (HPV) E6/E7 como la secuencia pertinente para la evaluación de la seguridad para un sistema basado en HeLa y el gen altamente expresado, de copia alta para ARN ribosómico (ARNr) como una etiqueta general sensible. Según Thorne *et al.* (*supra*), las secuencias coempaquetadas más predominantes se derivan del plásmido de empaquetamiento, que incluye rep y cap de AAV y los genes marcadores de selección de células bacterianas y mamíferas. Además, Ye *et al.* (2011, Gene Ther. 18, 135 - 144) revelan que las secuencias de ADN del plásmido de empaquetamiento de HSV se empaquetan de forma aleatoria durante la producción de rAAV en una línea celular mamífera. Según Ye *et al.*, el virión de AAV comprende fragmentos aleatorios a través de todo el genoma de HSV. Además, Chadeuf *et al.* (*supra*) muestran que, en el caso de un plásmido de ADN más pequeño, el plásmido completo y las ITR virales se encapsulan en el virión, que incluye el gen marcador de selección. En la patente WO2011/112089, las impurezas de ADN de baculovirus residual presentes en las muestras de rAAV5 purificadas de afinidad de lote se analizaron usando el

50

55

60

65

método Q-PCR. Por lo tanto, pareció no haber necesidad de detectar secuencias de nucleótidos específicas ya que las impurezas de ADN parecían estar empaquetadas de forma aleatoria en un virión parvoviral.

[0007] A diferencia de las enseñanzas generales de que las impurezas de ADN se empaquetan de forma aleatoria en el virión, la presente invención muestra que algunas impurezas de ADN están, de hecho, sobrerrepresentadas. En consecuencia, los métodos usados actualmente para detectar una impureza de ADN en una composición biofarmacéutica podrían llevar a una subestimación drástica de las impurezas de ADN presentes en una composición. Dicha subestimación de una impureza de ADN en una composición farmacéutica podría resultar en la administración de composiciones que no son lo suficientemente puras clínicamente, lo que conllevaría un riesgo potencial para la seguridad de salud para los pacientes.

[0008] Por lo tanto, existe una necesidad en la técnica de medios y métodos para identificar y cuantificar una impureza de ADN sobrerrepresentada en composiciones biológicas, tales como, por ejemplo, preparaciones biofarmacéuticas. Un objetivo de la presente invención es proporcionar tales medios y métodos.

#### Resumen de la invención

[0009] En un primer aspecto, la presente invención se refiere a un método para identificar y cuantificar una impureza de ácido nucleico sobrerrepresentada en una composición que comprende un vector parvoviral y donde el método comprende las etapas de:

- a) someter la composición a secuenciación de ácido nucleico para obtener lecturas aleatorias de secuencias de nucleótidos, donde la secuenciación de ácido nucleico comprende secuenciación de alto rendimiento;
- b) comparar las lecturas aleatorias de la etapa a) con una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición por la cual una correspondencia entre una lectura aleatoria y una secuencia de nucleótidos de un componente biológico identifica una impureza de ácido nucleico;
- c) determinar el número medio de lecturas por vector parvoviral; y
- d) determinar el número de lecturas por nucleótido de la impureza de ácido nucleico sobrerrepresentada, donde una impureza de ácido nucleico se identifica como una impureza sobrerrepresentada cuando la distribución de lecturas no es aleatoria y la impureza sobrerrepresentada comprende 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o 50 veces el número medio de lecturas del componente biológico, o cuando el número de lecturas por nucleótido de una impureza de ácido nucleico es de al menos el 0,001; 0,01; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0; 2,0; 3,0; 5,0; 7,0 o 10 % del número medio de lecturas por vector parvoviral.

[0010] En una forma de realización preferida, el vector parvoviral es un vector de virus adenoasociado recombinante (rAAV).

[0011] En una forma de realización preferida, la secuencia de nucleótidos de un componente biológico se selecciona a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente del vector parvoviral recombinante y un virus auxiliar, donde preferiblemente el vector es un vector baculoviral.

[0012] En una forma de realización preferida, el virus auxiliar es un adenovirus recombinante y/o un virus del herpes simple recombinante.

[0013] En una forma de realización preferida, la secuencia de nucleótidos del componente biológico comprende una secuencia de nucleótidos codificante para Rep, Cap y/o un transgén, donde, preferiblemente, el componente biológico comprende una secuencia de nucleótidos codificante para un transgén, donde, más preferiblemente, el componente biológico comprende una secuencia de nucleótidos codificante para un transgén flanqueado por al menos una ITR parvoviral y donde, más preferiblemente, el componente biológico comprende una secuencia de nucleótidos codificante para un transgén flanqueado por al menos una ITR parvoviral a cada lado.

[0014] En una forma de realización preferida, la impureza de ácido nucleico sobrerrepresentada se cuantifica en una segunda composición o una composición adicional.

[0015] La presente descripción se refiere a un método para cuantificar una impureza de ácido nucleico en una composición que comprende un vector parvoviral, donde el método comprende la etapa de determinar la abundancia relativa de una impureza de ácido nucleico, donde dicha impureza de ácido nucleico comprende una secuencia de nucleótidos situada entre 1 - 8000 pares de bases, 1 - 5000 pares de bases, 1 - 3000 pares de bases, 1 - 1000 pares de bases, 1 - 500 pares de bases, 1 - 250 pares de bases o 1 - 100 pares de bases inmediatamente adyacentes a una secuencia de ITR parvoviral cuando la secuencia de ITR está presente en un componente biológico usado en un proceso para producir la composición y donde el componente biológico comprende un transgén flanqueado por al menos una copia de la secuencia de ITR parvoviral.

[0016] En una forma de realización preferida de la divulgación, el componente biológico se selecciona a partir de un grupo consistente en una célula huésped, un plásmido, un vector diferente del vector parvoviral recombinante y un virus auxiliar, donde preferiblemente el vector es un vector baculoviral.

5 [0017] En una forma de realización preferida de la divulgación, el vector parvoviral es un vector de virus adenoasociado recombinante (rAAV).

[0018] En una forma de realización preferida de la divulgación, la secuencia de nucleótidos de la impureza de ácido nucleico se localiza inmediatamente adyacente a cada lado de la secuencia de ITR parvoviral cuando la  
10 secuencia de ITR está presente en un componente biológico usado en un proceso de producción de la composición.

[0019] En una forma de realización preferida de la divulgación, la abundancia relativa se determina en comparación con una secuencia de nucleótidos del vector parvoviral, y/o una secuencia de referencia de la  
15 composición.

[0020] En una forma de realización preferida de la divulgación, la abundancia relativa se determina por:

20 a) el número medio de lecturas por ácido nucleico de la impureza de ácido nucleico tal como se ha definido anteriormente; y

- i) el número medio de lecturas por ácido nucleico de la secuencia de referencia; y/o
- ii) el número medio de lecturas por vector parvoviral en la composición;

25 donde el número de lecturas se determina por un método tal y como se ha definido previamente; y/o  
b) la amplificación de la impureza de ácido nucleico tal como se ha definido anteriormente; y

- i) la secuencia de referencia; y/o
- ii) una secuencia de nucleótidos de un vector parvoviral.

30 [0021] En una forma de realización preferida de la divulgación, la abundancia relativa se determina mediante Q-PCR y/o por secuenciación de alto rendimiento.

[0022] En una forma de realización preferida de la divulgación, el método comprende además la etapa de una  
35 hibridación selectiva de un cebador de oligonucleótidos a la impureza de ácido nucleico tal y como se ha definido anteriormente o un complemento de la misma.

[0023] En una forma de realización preferida de la divulgación, el cebador de oligonucleótidos hibrida selectivamente una impureza de ácido nucleico que comprende una parte de una secuencia de baculovirus o un  
40 complemento de la misma.

[0024] En un segundo aspecto, la presente invención se refiere a un método de determinación de si una composición que comprende un vector parvoviral se considera clínicamente pura, donde el método comprende las etapas de:

- i) cuantificar una impureza de ácido nucleico en una composición de vector parvoviral tal y como se ha definido anteriormente; y
- ii) determinar que la composición es clínicamente pura si la impureza de ácido nucleico tal y como se ha definido anteriormente está al menos 10, 100, 250, 1000 veces menos presente que la secuencia de referencia y/o transgén tal y como se determina por la abundancia relativa de la impureza de ácido nucleico.

[0025] En una forma de realización preferida, la composición que comprende el vector parvoviral es una composición farmacéutica. De forma alternativa, o en combinación con otra forma de realización preferida, en una forma de realización preferida de la presente invención la composición que comprende el vector parvoviral comprende una cápside parvoviral donde se empaqueta el vector parvoviral. De forma alternativa, o en combinación con otra forma de realización preferida, en una forma de realización preferida de la presente invención, la composición que comprende el vector parvoviral no comprende una muestra obtenida u obtenible a partir de un mamífero, donde el mamífero es preferiblemente un primate no humano.

#### 60 Descripción de la invención

[0026] La presente divulgación se refiere al descubrimiento de que las impurezas de ADN no se empaquetan de forma aleatoria en el virión parvoviral. En cambio, hay impurezas de ácido nucleico que están sobrerrepresentadas en la composición de virión. Por lo tanto, en un primer aspecto, la divulgación se refiere a un método para identificar una impureza de ácido nucleico en una composición. La composición comprende un vector parvoviral. El método comprende las etapas de: a) someter la composición a secuenciación de ácido

nucleico para obtener lecturas aleatorias de secuencias de nucleótidos; b) comparar las lecturas aleatorias de la etapa a) con una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición, por medio del cual una correspondencia entre una lectura aleatoria y una secuencia de nucleótidos de un componente biológico identifica una impureza de ácido nucleico. Para cuantificar una impureza de ácido nucleico identificada, el método comprende además preferiblemente las etapas de: c) determinar el número medio de lecturas por vector parvoviral; y d) determinar el número de lecturas por nucleótido de una impureza de ácido nucleico identificada, donde una impureza de ácido nucleico se identifica como una impureza sobrerrepresentada cuando la distribución de lecturas no es aleatoria y la impureza sobrerrepresentada tiene 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o 50 veces el número medio de lecturas del componente biológico, o cuando el número de lecturas por nucleótido de una impureza de ácido nucleico es de al menos el 0,001; 0,01; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0; 2,0; 3,0; 5,0; 7,0 o 10 % del número medio de lecturas por vector parvoviral.

[0027] Por lo tanto, el método de la divulgación se refiere a la identificación y cuantificación de una impureza de ácido nucleico en una composición. La impureza de ácido nucleico puede ser una impureza de ADN y/o una impureza de ARN, preferiblemente la impureza de ácido nucleico es una impureza de ADN.

[0028] El término "impureza de ácido nucleico" se entiende en el sentido de que incluye cualquier secuencia de ácidos nucleicos que no se destina a ser empaquetada en el virión parvoviral, tal como, por ejemplo, secuencias de nucleótidos de un componente biológico usado en el proceso de producción de la composición. En particular, una secuencia que no está flanqueada por una ITR parvoviral a cada lado puede constituir una impureza de ácido nucleico.

[0029] Un "vector parvoviral" aquí se refiere a una molécula de ácido nucleico recombinante que comprende una o más secuencias de polinucleótidos de interés (por ejemplo un constructo de expresión para un gen que codifica un producto de interés, es decir, un "transgén" ) que están flanqueados por al menos una (y normalmente dos) secuencia(s) de repetición terminal(es) invertida(es) (ITR) parvoviral(es).

[0030] Una "distribución aleatoria de las lecturas" se define aquí como una distribución de lecturas que se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición. En particular, una distribución aleatoria de lecturas se define como una distribución de lecturas que se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición. Más preferiblemente, la distribución aleatoria de lecturas se define aquí como una distribución de lecturas que se alinean uniformemente a lo largo la longitud de una secuencia de nucleótidos seleccionada a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente de un vector parvoviral recombinante y virus auxiliar, donde preferiblemente el vector es un vector baculoviral. Por lo tanto, más preferiblemente, una distribución aleatoria de lecturas se define en la invención como una distribución de lecturas que se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un vector baculoviral.

[0031] Un alineamiento uniforme se define aquí como una probabilidad igual de que una lectura se alinee con una región específica de una secuencia de nucleótidos en comparación con cualquier otra región de la misma secuencia de nucleótidos. Preferiblemente, en un alineamiento uniforme el número de lecturas que se alinean con un nucleótido no desvía más del 0,1; 0,2; 0,5; 1,0; 2,0; 5,0; 10, 15 o 20 % del número medio de lecturas que se alinean con dicha secuencia de nucleótidos.

[0032] Una "distribución de lecturas no aleatoria" se define aquí como una distribución de lecturas que no se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición. En particular, una distribución no aleatoria de lecturas se define como una distribución de lecturas que no se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición. Más preferiblemente, una distribución no aleatoria de lecturas se define aquí como una distribución de lecturas que no se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos seleccionada a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente de un vector parvoviral recombinante y virus auxiliar, donde preferiblemente el vector es un vector baculoviral. Por lo tanto, más preferiblemente una distribución no aleatoria de lecturas se define en la invención como una distribución de lecturas que no se alinean uniformemente a lo largo de la longitud de una secuencia de nucleótidos de un vector baculoviral, es decir, significa que más lecturas se alinean con regiones específicas del vector baculoviral en comparación con otras regiones del vector baculoviral.

[0033] Preferiblemente, la composición es una composición que comprende un virión parvoviral (recombinante) que comprende o consiste en (al menos) una cápside parvoviral donde se empaqueta el vector parvoviral recombinante. La composición y sus constituyentes son adicional y preferiblemente tal y como se define más adelante. En una forma de realización preferida de la divulgación, el método es para identificar y/o cuantificar impurezas de ácido nucleico que están empaquetadas en un virión parvoviral, es decir, encapsuladas en el virión. En particular, la impureza de ácido nucleico según la divulgación no se degrada después del tratamiento

de nucleasa (por ejemplo, tratamiento de RNasa o DNasa) de la composición que comprende un virión parvoviral.

#### Composiciones

5

[0034] En una forma de realización preferida, un vector parvoviral se encuentra contenido en una composición. Preferiblemente, la composición es una composición farmacéutica. La composición farmacéutica comprende además preferiblemente un portador farmacéuticamente aceptable. Se puede usar cualquier portador o excipiente farmacéuticamente aceptable adecuado en las presentes composiciones (véase, por ejemplo, Remington: The Science and Practice of Pharmacy, Alfonso R. Gennaro (Editor) Mack Publishing Company, abril 1997).

10

15

[0035] Las formas farmacéuticas preferidas serían en combinación con solución salina estéril, solución de dextrosa o solución tamponada, u otros fluidos estériles farmacéuticamente aceptables. De forma alternativa, se puede usar un soporte sólido tal como, por ejemplo, microesferas de microportadores.

20

[0036] De forma alternativa o en combinación con otra forma de realización preferida, en una forma de realización preferida de la presente invención, la composición que comprende el vector parvoviral no comprende una muestra, tal como por ejemplo una muestra de hígado o músculo, obtenido u obtenible a partir de un mamífero. En una forma de realización más preferida, la composición no comprende una muestra obtenida u obtenible a partir de un primate no humano. En una forma de realización más preferida, la composición no comprende ADN genómico procedente de músculo o hígado de un mamífero, tal como, por ejemplo, de un primate no humano.

25

#### Secuenciación de ácido nucleico

[0037] En un método según la divulgación, una impureza de ácido nucleico sobrerrepresentada en una composición se identifica y cuantifica. Los métodos para identificar y cuantificar una impureza de ácido nucleico sobrerrepresentada incluyen, pero no se limitan a, secuenciación de Sanger o secuenciación de alto rendimiento.

30

[0038] En una primera forma de realización, el ADN de vector parvoviral se puede clonar en un plásmido, seguido de una secuenciación de Sanger convencional. La secuenciación de Sanger se define aquí como un método de secuenciación del ADN que se basa en la incorporación selectiva de didesoxinucleótidos terminadores de cadena por polimerasa de ADN durante la replicación de ADN *in vitro*. La secuenciación de Sanger conforme a la presente invención comprende las denominadas secuenciación de Sanger terminadora de cadena y/o secuenciación de Sanger terminadora de coloración. Preferiblemente, la secuenciación de Sanger es una secuenciación de Sanger terminadora de coloración. La secuenciación terminadora de coloración emplea el marcado de los ddNTP terminadores de cadena. En particular, en la secuenciación terminadora de coloración, cada uno de los cuatro didesoxinucleótidos terminadores de cadena se marca con colorantes fluorescentes, cada uno de los cuales emite luz en longitudes de onda diferentes, lo que permite la secuenciación en una reacción única. Otros métodos de secuenciación del ADN pueden ser igualmente adecuados, tal como secuenciación del ADN de nanoporos, secuenciación del ADN de corrientes túnel, secuenciación por hibridación, secuenciación con espectrometría de masas, secuenciación de Sanger microfluídica, técnicas basadas en microscopía y secuenciación de ARNP.

45

[0039] En la divulgación actual, la secuenciación de ácido nucleico comprende la secuenciación de alto rendimiento. La secuenciación de alto rendimiento (también denominada secuenciación de nueva generación o secuenciación profunda) se refiere a tecnologías de secuenciación de ADN de alto rendimiento no basadas en Sanger. Se pueden secuenciar miles, millones o incluso miles de millones de cadenas de ADN en paralelo, produciendo sustancialmente más rendimiento y minimizando la necesidad de los métodos de clonación de fragmentos que se usan a menudo en la secuenciación de Sanger de genomas.

50

[0040] En un método preferido según la divulgación, la secuenciación de alto rendimiento comprende secuenciación de moléculas únicas Heliscope, secuenciación en tiempo real de moléculas únicas (SMRT), secuenciación Ion Torrent (semiconductor iónico), secuenciación 454 (pirosecuenciación, Roche 454 Life Sciences™, Branford, CT), Solexa (secuenciación por síntesis, Illumina, Inc, San Diego, CA) y/o secuenciación SOLiD (secuenciación por ligación, ABI, Applied Biosystems, Indianapolis, IN). En otra forma de realización preferida de la divulgación, la secuenciación de ácido nucleico comprende secuenciación SOLiD, Solexa y/o 454. En una forma de realización más preferida, la secuenciación de ácidos nucleicos comprende secuenciación Solexa/Illumina o 454.

60

[0041] El método según la divulgación no se limita a los métodos de secuenciación de alto rendimiento actualmente conocidos. En particular, se entiende que se desarrollarán nuevos métodos de secuenciación de alto rendimiento a lo largo del tiempo que son igualmente adecuados para usar en un método de la divulgación. En particular, cualquier método de secuenciación que se clasifique como un método de secuenciación de alto

65

rendimiento, es decir, cualquier método que genere miles, millones o miles de millones de lecturas en una única ejecución, se puede usar en un método según la divulgación.

[0042] En un método preferido según la divulgación, se obtienen lecturas aleatorias sometiendo una composición que comprende un vector parvoviral a secuenciación de ácidos nucleicos. Una "lectura" o "conteo" (de secuenciación) se define como una cadena individual de bases producidas durante un método de secuenciación de ácidos nucleicos. Diferentes métodos de secuenciación de alto rendimiento pueden generar diferentes números de lecturas por ejecución (reacción) y lecturas de diferentes longitudes. Por ejemplo, Illumina genera hasta 3 mil millones de lecturas por ejecución y una lectura tiene una longitud media de 50 - 300 pares de bases. Por otro lado, la secuenciación 454 genera aproximadamente 1 millón de lecturas por ejecución con lecturas con una longitud media de aproximadamente 700 pares de bases. El número de lecturas (número de lecturas por ejecución) y la longitud de lectura (número de bases por lectura) puede variar por ejecución de secuenciación y se prevé que la longitud de lectura, así como el número de lecturas por ejecución aumente después del posterior desarrollo de los diferentes métodos de secuenciación de alto rendimiento.

#### Identificación de una impureza de ácido nucleico sobrerrepresentada

[0043] En una forma de realización de la divulgación, las lecturas aleatorias obtenidas tal y como se ha descrito anteriormente se comparan con una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición, donde la comparación de las lecturas aleatorias con la secuencia de nucleótidos del componente biológico tiene como resultado la identificación de la impureza de ácido nucleico sobrerrepresentada. Esta secuencia de nucleótidos de un componente biológico puede ser una fuente esperada o inesperada de secuencias de nucleótidos.

[0044] Una fuente esperada de secuencias de nucleótidos se puede seleccionar a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector y un virus auxiliar. Preferiblemente donde el virus auxiliar es un adenovirus y/o un virus del herpes simple y/o donde el vector es un vector baculoviral. En una forma de realización preferida, la fuente esperada de secuencias de nucleótidos es un vector baculoviral. En un método según la divulgación, las lecturas aleatorias se alinean de este modo o en comparación con una fuente esperada de secuencias de nucleótidos, lo que lleva a la identificación de la impureza de ácido nucleico sobrerrepresentada.

[0045] De forma alternativa o en combinación con las formas de realización precedentes, la fuente de secuencias de nucleótidos es una fuente inesperada de secuencias de nucleótidos, es decir, la fuente de secuencias de nucleótidos no es predeterminada. La fuente inesperada de secuencias de nucleótidos se puede recuperar por ensamblaje *de novo* de las lecturas aleatorias obtenidas según el método de la divulgación y comparando estas secuencias de nucleótidos ensambladas a una base de datos de secuencias de nucleótidos. Dicha base de datos de secuencias de nucleótidos puede ser una base de datos de secuencias de nucleótidos disponible de forma privada o pública. Algunos ejemplos de una base de datos de secuencias de nucleótidos disponibles públicamente incluyen, pero no se limitan a, UCSC Genome Bioinformatics, GenBank, DDBJ, ENA, etc. La comparación de la secuencia ensamblada con una secuencia de la base de datos disponible de forma privada o pública podría llevar a la identificación de la impureza de ácido nucleico.

[0046] En un método preferido de la divulgación, la secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición es una fuente esperada de secuencias de nucleótidos. En particular, en una forma de realización preferida, la secuencia de nucleótidos de un componente biológico se selecciona a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente del vector parvoviral recombinante y un virus auxiliar. En particular, la secuencia de nucleótidos del componente biológico no comprende una secuencia de nucleótidos de un vector parvoviral.

[0047] Una célula huésped según la presente descripción puede ser cualquier célula usada en el proceso de producción de la composición. Una célula huésped se podría seleccionar a partir de un grupo consistente en: una célula vegetal, una célula bacteriana, una célula de levadura y una célula animal. Preferiblemente, la célula huésped es una célula animal y, más preferiblemente, la célula huésped es una célula huésped de mamífero o una célula huésped de insecto. En una forma de realización más preferida de la divulgación, la célula huésped es una célula huésped de insecto. Una secuencia de nucleótidos de una célula huésped usada en el proceso de producción de la composición comprende ADN genómico y/o ADN mitocondrial. Preferiblemente, la secuencia de nucleótidos de la célula huésped comprende ADN genómico. El ADN genómico puede comprender un gen seleccionado a partir de un grupo de genes consistentes en: un transgén, un gen codificante de Rep, un gen codificante de Cap y un gen codificante para una proteína o un ARN con una función auxiliar para producir la composición. Dicho gen codificante para una proteína o ARN con una función auxiliar para producir la composición se puede derivar de un virus animal, tal como un adenovirus y/o un virus de herpes simple o un virus de insecto, tal como baculovirus. De forma alternativa o en combinación, el ADN genómico puede comprender una secuencia de ITR. En una forma de realización preferida, el ADN genómico de la célula huésped comprende un transgén flanqueado por al menos una ITR, y preferiblemente el transgén está flanqueado por una ITR a cada lado.

[0048] Un plásmido según la presente descripción puede ser cualquier plásmido usado en el proceso de producción de la composición. El plásmido comprende preferiblemente un gen seleccionado de un grupo de genes consistentes en: un gen de resistencia, un transgén, un gen que codifica Rep, una gen que codifica Cap y un gen que codifica una proteína o un ARN con una función auxiliar para producir la composición. De forma alternativa o en combinación, el plásmido puede comprender una secuencia de ITR. En una forma de realización preferida, el plásmido comprende un transgén flanqueado por al menos una ITR, y preferiblemente el transgén está flanqueado por una ITR a cada lado.

[0049] Un vector según la presente descripción puede ser cualquier vector usado en el proceso de producción de la composición. Un vector se puede seleccionar de un grupo consistente en: un plásmido, un vector viral, un cósmido y un cromosoma artificial. En una forma de realización preferida de la invención, el vector es un vector viral. En la forma de realización más preferida, el vector es un vector baculoviral. El vector baculoviral usado en un proceso de producción de la composición puede comprender un gen seleccionado a partir de un grupo de genes consistentes en: un transgén, un gen que codifica Rep, un gen que codifica Cap y un gen que codifica una proteína con una función auxiliar para producir la composición. De forma alternativa o en combinación, el vector baculoviral puede comprender una secuencia de ITR. En una forma de realización preferida, el vector baculoviral comprende un transgén flanqueado por al menos una ITR, y preferiblemente el transgén está flanqueado por una ITR a cada lado.

[0050] Un virus auxiliar según la presente descripción puede ser cualquier virus usado en el proceso de producción de la composición. En una forma de realización preferida, el virus auxiliar se usa en un proceso para producir un virión parvoviral. En otra forma de realización preferida, el virus auxiliar se usa en un proceso de producción de un virión adenoasociado recombinante (rAAV). En la forma de realización más preferida, el virus auxiliar es un adenovirus y/o un virus del herpes simple. En una forma de realización más preferida, el virus auxiliar es un adenovirus recombinante y/o un virus del herpes simple recombinante. En otra forma de realización, el virus auxiliar comprende un gen seleccionado de un grupo de genes consistentes en: un transgén, un gen que codifica Rep, un gen que codifica Cap y un gen que codifica una proteína con una función auxiliar para producir la composición. De forma alternativa o en combinación, el virus auxiliar puede comprender una secuencia de ITR. En una forma de realización preferida, el virus auxiliar comprende un transgén flanqueado por al menos una ITR, y preferiblemente el transgén está flanqueado por una ITR a cada lado.

[0051] En una forma de realización preferida, la secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición es a partir de un vector baculoviral. En una forma de realización más preferida, la secuencia de nucleótidos a partir de un componente biológico es a partir de un vector baculoviral que comprende un transgén, flanqueado por al menos una ITR, y preferiblemente el transgén está flanqueado por dos ITR.

[0052] En combinación con o de forma alternativa a lo anterior, la secuencia de nucleótidos del componente biológico comprende una secuencia de nucleótidos que codifica para Rep, Cap y/o un transgén, donde preferiblemente el componente biológico comprende una secuencia de nucleótidos que codifica para un transgén, donde más preferiblemente el componente biológico comprende una secuencia de nucleótidos que codifica para un transgén que está flanqueado por al menos una ITR parvoviral, y donde, de la forma más preferible el componente biológico comprende una secuencia de nucleótidos que codifica un transgén que está flanqueado por al menos una ITR parvoviral a cada lado.

Cuantificación de una impureza de ácido nucleico sobrerrepresentada usando secuenciación de ácido nucleico

[0053] La presente divulgación divulga un método para identificar y cuantificar una impureza de ácido nucleico sobrerrepresentada en una composición, donde el método comprende una etapa de someter la composición a secuenciación de ácidos nucleicos para obtener lecturas aleatorias de secuencias de nucleótidos. La impureza de ácido nucleico se puede identificar tal y como se ha indicado previamente. La impureza de ácido nucleico identificada se puede cuantificar posteriormente determinando el número de lecturas por ácido nucleico de la impureza de ácido nucleico sobrerrepresentada en la composición.

[0054] El número de lecturas por ácido nucleico se define aquí como el número de lecturas que se alinean con un ácido nucleico específico de una secuencia de nucleótidos. De este modo, el número de lecturas por ácido nucleico de una impureza de ácido nucleico en la composición se entiende como el número de lecturas que se alinean específicamente a un ácido nucleico de una impureza de ácido nucleico.

[0055] El número de lecturas que se alinean a un ácido nucleico específico de una secuencia de nucleótidos se traduce en la frecuencia donde este ácido nucleico particular está presente en una composición. Por lo tanto, un alto número de lecturas que se alinean con a un ácido nucleico específico se entiende como un ácido nucleico que es abundante en la composición. De forma alternativa, si solo se alinean unas pocas lecturas con un ácido nucleico específico, se entiende que la presencia de este ácido nucleico es escasa en la composición.



[0056] Según un método de la presente descripción, una impureza de ácido nucleico está sobrerrepresentada cuando la distribución de lecturas no es aleatoria y la impureza sobrerrepresentada comprende 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o 50 veces el número medio de lecturas del componente biológico, o cuando el número de lecturas por ácido nucleico de una impureza de ácido nucleico es de al menos el 0,0005; 0,001; 0,002; 0,003; 0,004; 0,005; 0,006; 0,007; 0,008; 0,009; 0,01; 0,02; 0,03; 0,04; 0,05; 0,05; 0,06; 0,07; 0,08; 0,09; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0; 1,5; 2,0; 3,0; 4,0; 5,0; 6,0; 7,0 o 10 % del número medio de lecturas por ácido nucleico del vector parvoviral.

[0057] El número medio de lecturas por vector parvoviral se define aquí como el número total de lecturas que se alinean con el vector parvoviral dividido por el número total de nucleótidos del vector parvoviral. Preferiblemente, las lecturas y nucleótidos de los 25, 50, 75, 100, 200, 300, 400 o 500 nucleótidos más corriente arriba del vector parvoviral y/o las lecturas y nucleótidos de los 25, 50, 75, 100, 200, 300, 400 o 500 nucleótidos más corriente abajo del vector parvoviral no se consideran cuando se determina el número medio de lecturas por vector parvoviral. Estas lecturas pueden no ser representativas para el número medio de lecturas por vector parvoviral ya que hay una disminución artificial en el número de lecturas que se alinean con los nucleótidos más corriente arriba y/o más corriente abajo del vector parvoviral, es decir, en los terminales del vector.

[0058] El número medio de lecturas de un componente biológico se define aquí como el número total de lecturas que se alinean con un componente biológico dividido por el número total de nucleótidos del componente biológico. El componente biológico puede comprender un componente biológico usado en el proceso de producción de la composición. Más preferiblemente, el componente biológico se selecciona a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente de un vector parvoviral recombinante y virus auxiliar, donde preferiblemente el vector es un vector baculoviral. Más preferiblemente, el componente biológico es un vector baculoviral.

[0059] En una forma de realización de la presente invención, el número de nucleótidos del vector parvoviral puede comprender la secuencia de nucleótidos completa del vector parvoviral, por ejemplo con las secuencias de ITR, las secuencias promotoras, una secuencia de transgén y cualquiera de las otras secuencias entre la ITR izquierda y la ITR derecha. En una forma de realización más preferida, una parte de la secuencia de nucleótidos del vector parvoviral se selecciona para determinar el número medio de lecturas del vector parvoviral. Dicha parte del vector parvoviral puede comprender la secuencia de nucleótidos ITR, la secuencia promotora y/o la secuencia de transgén. En una forma de realización más preferida, la secuencia de un transgén se selecciona para determinar el número medio de lecturas por ácido nucleico del vector parvoviral, es decir el número de lecturas que se alinean con el transgén se divide por el número de nucleótidos de la secuencia de nucleótidos del transgén.

[0060] En otra forma de realización de la presente divulgación, la impureza de ácido nucleico sobrerrepresentada se cuantifica en una segunda composición o una composición adicional. Después de identificar y cuantificar una impureza de ácido nucleico sobrerrepresentada en una primera composición, la impureza de ácido nucleico sobrerrepresentada se puede cuantificar posteriormente en otra composición. La cuantificación de una impureza de ácido nucleico sobrerrepresentada en una segunda composición o composición adicional se puede determinar como se ha descrito previamente. De forma alternativa, la cuantificación de una impureza de ácido nucleico sobrerrepresentada en una segunda composición o composición adicional podría determinarse mediante cualquier otro método adecuado para determinar la cantidad de la una impureza de ácido nucleico. En la técnica se conocen métodos para cuantificar fragmentos específicos de ADN y se aplican igualmente para cuantificar una impureza de ácido nucleico sobrerrepresentada en una segunda composición o una composición adicional. Estos métodos incluyen, pero no se limitan a, secuenciación de alto rendimiento, Q-PCR, PCR de ciclo limitado, un ensayo de hibridación, una micromatriz y una electroforesis de agarosa.

#### 50 Viriones parvovirales

[0061] En una forma de realización preferida de la divulgación, la composición comprende un vector parvoviral. En particular, en un método preferido según la invención, el vector parvoviral es un vector de virus adenoasociado recombinante (rAAV). Los virus de la familia Parvoviridae son pequeños virus animales de ADN. La familia Parvoviridae se puede dividir en dos subfamilias: los Parvovirinae, que infectan vertebrados y los Densovirinae, que infectan insectos. Los miembros de la subfamilia Parvovirinae se designan aquí como los parvovirus e incluyen el género Dependovirus. Como se puede deducir del nombre de su género, los miembros del Dependovirus son únicos en el hecho de que requieren normalmente la coinfección con un virus auxiliar tal como adenovirus o virus del herpes para infectar productivamente en un cultivo celular. El género Dependovirus incluye AAV, que normalmente infecta seres humanos (por ejemplo, serotipos 1, 2,3A, 3B, 4, 5 y 6) o primates (por ejemplo, serotipos 1 y 4), y los virus relacionados que infectan otros animales de sangre caliente (por ejemplo, virus adenoasociados bovinos, caninos, equinos y ovinos). Se describe información adicional sobre los parvovirus y otros miembros de los Parvoviridae en Kenneth I. Berns, "Parvoviridae: The Viruses and Their Replication," Capítulo 69 en Fields Virology (3a Ed. 1996).

[0062] La organización genómica de todos los serotipos de AAV conocidos es muy similar. El genoma de AAV es una molécula de ADN lineal, monocatenaria que es de menos de aproximadamente 5,000 nucleótidos (nt) de longitud. Las repeticiones terminales invertidas (ITR) flanquean las secuencias de nucleótidos de codificación únicas para las proteínas de replicación (Rep) no estructurales y las proteínas estructurales (VP). Las proteínas VP (VP1, -2 y -3) forman la cápside. Los terminales 145 nt son autocomplementarios y están organizados de modo que se puede formar un dúplex intramolecular energéticamente estable que forma una horquilla en forma de T. Estas estructuras en horquilla funcionan como un origen para la replicación de ADN viral, sirviendo de cebadores para el complejo de polimerasa de DNA celular. A continuación de la infección de AAV de tipo silvestre (wt) en células mamíferas, los genes Rep (es decir, Rep78 y Rep52) se expresan a partir del promotor P5 y el promotor P19, respectivamente y ambas proteínas Rep tienen una función en la replicación del genoma viral. Un evento de empalme en el Rep ORF tiene como resultado la expresión de, en realidad, cuatro proteínas Rep (es decir, Rep78, Rep68, Rep52 y Rep40). Sin embargo, se ha demostrado que los ARNm no empalmados, que codifican las proteínas Rep78 y Rep52, en células mamíferas son suficientes para la producción de vector de AAV. También en las células de insecto, las proteínas Rep78 y Rep52 bastan para la producción de vector de AAV.

[0063] Un "vector parvoviral o AAV" (o "vector rAAV" ) se refiere aquí a una molécula de ácido nucleico que comprende una o más secuencias polinucleótidas de interés (por ejemplo un constructo de expresión para gen que codifica un producto de interés, es decir un "transgén") que están flanqueadas por al menos una secuencia de repetición terminal invertida (ITR) parvoviral o de AAV. Dichos vectores de rAAV se pueden replicar y empaquetar en viriones infecciosos en caso de existir en una célula huésped que expresa productos génicos rep y cap de AAV (es decir, proteínas Rep y Cap de AAV). Un vector parvoviral o de AAV es preferiblemente una molécula de ácido nucleico recombinante, es decir una molécula de ácido nucleico que no se produce en la naturaleza y se compone combinando elementos de secuencia que no se producen naturalmente en esta combinación y/o orden.

[0064] Cuando un vector de rAAV se incorpora en un constructo de ácido nucleico mayor (por ejemplo en un cromosoma o en otro vector tal como un plásmido o baculovirus usado para la clonación o transfección), entonces el vector de rAAV se designa típicamente como un "provector" que se puede "rescatar" por replicación y encapsidación en presencia de funciones de empaquetamiento de AAV y funciones auxiliares necesarias.

[0065] Preferiblemente, un producto génico de interés está flanqueado por ITR de AAV en cada lado. Cualquier ITR de AAV se puede usar en el método de la invención, incluyendo ITR de AAV1, AAV2, AAV4, AAV5, AAV6, AAV8, AAV9 y/o AAVrh10. Las ITR de AAV2 son las más preferidas. Ejemplos de secuencias de ITR preferidas para usar en constructos de ácido nucleico preferidos de la invención se proporcionan en la SEQ ID N.º: 1 (ITR izquierda o corriente arriba) y la SEQ ID N.º: 2 (ITR derecha o corriente abajo).

[0066] El AAV es capaz de infectar un número de células mamíferas. Véase, por ejemplo, Tratschin *et al.* (1985, Mol. Cell Biol. 5:3251-3260) y Grimm *et al.* (1999, Hum. Gene Ther. 10:2445-2450). Sin embargo, la transducción de AAV de fibroblastos sinoviales humanos es significativamente más eficiente que en células murinas similares, Jennings *et al.*, Arthritis Res, 3:1 (2001), y la tropicidad celular de AAV difiere entre serotipos. Véase, por ejemplo Goncalves, 2005, Virol J. 2(1):43 que discute métodos para la modificación de tropismo de AAV.

[0067] Un vector rAAV para usar en un método de la presente descripción se puede producir en células mamíferas o en células de insecto. Ambos métodos se describen en la técnica. Por ejemplo, Grimm *et al.* (2003 Molecular Therapy 7(6):839-850) revelan una estrategia para producir vectores de AAV de una manera libre de virus auxiliares y ópticamente controlable, que se basa en la transfección de solo dos plásmidos en células 293T. Estos revelan un método para la producción de un vector de AAV híbrido que comprende proteínas Rep de AAV2, ITR de AAV2 y proteínas de cápside de AAV5. Esta referencia se incluye aquí en su totalidad. También se puede encontrar información adicional en Blits *et al.* (2010) (Journal of Neuroscience methods 185(2):257-263).

[0068] Los términos "híbrido" y "pseudotipado" se usan aquí de manera intercambiable y se utilizan para indicar vectores de los cuales las proteínas Rep, las ITR y/o las proteínas de cápside son de diferentes serotipos. Por ejemplo, las ITR y las proteínas Rep son de AAV2 y las proteínas de cápside son de AAV5. El término "quimérico" se utiliza aquí para describir que un único gen, tal como por ejemplo la cápside, está compuesto por al menos dos secuencias derivadas de distintos serotipos.

#### Métodos de producción de AAV

[0069] Un vector de rAAV en una composición según la presente descripción puede producirse usando un método de producción tradicional para rAAV. Tales métodos de producción tradicionales se basan en protocolos de transfección transitoria de células de objetivo/productor (Merten *et al.*, Gene Ther, 2006, 12: S51 - S61). Este es un enfoque basado en la transcomplementación y la transfección transitoria, que requiere los siguientes elementos genéticos: (i) la secuencia del genoma de rAAV. La secuencia del genoma de rAAV se puede clonar en un plásmido (el llamado plásmido de vector viral). Este plásmido de vector viral comprende normalmente al menos una ITR y un casete de expresión para la expresión de un transgén; (ii) la secuencia que codifica rep y

cap y (iii) las funciones de auxiliar requeridas codificadas por un virus auxiliar natural, tal como un adenovirus y/o un virus del herpes simple.

[0070] Por ejemplo, los rAAV se pueden producir en células mamíferas según el método siguiente, pero limitado a este: el genoma de vector contiene el casete de expresión del transgén flanqueado por dos repeticiones terminales invertidas (ITR) derivadas del AAV del serotipo 2. La longitud total del genoma de vector viral no puede exceder el tamaño de genoma de tipo silvestre de 4,7 kB para mantener una eficiencia de empaquetamiento eficiente. Una única cápside se compone de 60 proteínas virales de VP1 (62kDa), VP2 (73 kDa) o VP3 (87 kDa), en una proporción de 1:1:10. El proceso de fabricación de los vectores de AAV se basa en la transfección de Ca(PO<sub>4</sub>)<sub>2</sub> de dos plásmidos en células de producción embrionarias humanas del riñón (HEK293) en frascos rotativos (área de superficie de 850 cm<sup>2</sup>) seguida por la purificación de los genomas de vector encapsidados mediante técnicas de filtración y cromatografía. El primer plásmido es el plásmido de vector viral y contiene un constructo de expresión flanqueado por ITR de AAV2. El segundo plásmido es el plásmido de empaquetamiento y codifica los genes rep tipo 2 y cap tipo 5 de AAV del serotipo deseado y genes auxiliares tempranos de adenovirus E2A, VA, E4 (pDP5; secuencia de nucleótidos descrita en la SEC ID N.º:3). El genoma de la línea celular de producción comprende el adenovirus E1 para proporcionar funciones auxiliares. A continuación de la cotransfección con los dos plásmidos en medio Iscove's Modified Dulbecco's Medium (IMDM) que contiene un 10 % de suero fetal bovino (FCS), las células se incuban durante tres días en medio Dulbecco's modified Eagle's medium (DMEM) sin suero para permitir que tenga lugar la producción de vector. La producción de vector en frascos rotativos tiene como resultado de media rendimientos de 3x10<sup>3</sup> genomas de vector por célula o 4x10<sup>11</sup> genomas de vector por frasco rotativo (cuantificados mediante qPCR). Posteriormente, el cultivo celular se lisa mediante un tampón que contiene Triton-X-100 y se eliminan restos celulares por centrifugación a velocidad baja. El volumen clarificado se purifica por cromatografía de afinidad de Sefarosa AVB y formula en PBS/5 % de sacarosa por concentración y diafiltración usando un módulo de fibra hueca de 400 kDa (por ejemplo de Spectrum Laboratories).

[0071] De forma alternativa, un rAAV en una composición según la presente descripción se puede producir con un método predominantemente independiente de transfección. Tales métodos podrían estar basados o en el uso de empaquetamiento/líneas celulares productoras, que producen rAAV después de la inducción, o en el uso del sistema baculovirus/célula de insecto.

[0072] Las células de empaquetamiento pueden albergar una parte de todos los elementos genéticos de AAV necesarios, tales como las secuencias de AAV auxiliar rep y cap. La posterior inducción de la producción de rAAV a partir de una línea celular de empaquetamiento podría llevarse a cabo por transfección de un plásmido que contiene de la secuencia de rAAV (el plásmido de vector viral) seguida por la introducción de una secuencia que codifica las funciones de auxiliar requeridas, tales como una infección con adenovirus o virus de herpes simple (de replicación deficiente). Las líneas celulares productoras pueden ser sistemas de transcomplementación completos, que albergan todos los componentes derivados de AAV necesarios integrados en su genoma, es decir, secuencias de AAV auxiliar (rep-cap) junto con la secuencia de vectores virales. La inducción de producción de rAAV puede tener lugar después de la introducción de una secuencia que codifica las funciones del auxiliar requerido.

[0073] Por otro lado, la secuencia del genoma de rAAV, que comprende al menos una ITR y un casete de expresión para la expresión de un transgén, se puede introducir en el genoma de un virus auxiliar, tal como el adenovirus o virus del herpes simple, generando respectivamente un sistema rAAV/Ad-hybrid (Thorne *et al.*, 2009; Hum. Gene Ther. 20; 707-714) o sistema rAAV/HSV-hybrid (Clement *et al.*, 2009; Hum. Gene ther. 20; 796-806. ; Ye *et al.*, 2014; Hum. Gene Ther. 15; 1-6).

[0074] De forma alternativa, un vector de AAV para usar en un método de la presente descripción se puede producir en células de insecto, tal y como ha sido descrito previamente por Urabe *et al.* (Journal of Virology 2006 80(4):1874-1885). En este sistema, la secuencia del genoma de rAAV se puede clonar en un baculovirus recombinante.

[0075] Las impurezas de ADN en una composición que comprende un vector parvoviral pueden derivarse de cualquier componente biológico usado en el proceso de producción de la composición. La composición se puede producir según cualquiera de los métodos tal y como se han descrito previamente.

[0076] Preferiblemente, en un método según la presente descripción, la secuencia de nucleótidos de un componente biológico se selecciona a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector y un virus auxiliar. En una forma de realización preferida, el componente biológico comprende al menos uno de los siguientes elementos genéticos: (i) la secuencia del genoma de rAAV, que comprende preferiblemente al menos una ITR y un casete de expresión para la expresión de un transgén; (ii) una secuencia que codifica rep y/o cap, y/o (iii) una secuencia que codifica las funciones de auxiliar requeridas, que se codifican naturalmente por un virus auxiliar, tal como un adenovirus y/o un virus del herpes simple. Más preferiblemente, el componente biológico comprende al menos una ITR y un casete de expresión para la expresión de un transgén.

[0077] En un método preferido según la divulgación, el vector es un vector baculoviral. Los *Baculoviridae* son una familia de virus de ADN grandes y envueltos. Los Baculovirus infectan preferentemente artrópodos donde la amplia mayoría de especies permisivas se engloban dentro del orden de los lepidópteros. Varias líneas celulares continuas tales como Sf9, Sf21 o High Five que permiten la propagación de baculovirus *in vitro* están disponibles comercialmente y se pueden usar para la producción de una composición según la presente invención.

[0078] Los Baculovirus recombinantes derivados del virus de la poliedrosis multinuclear de *Autographa californica* (AcMNPV) son los que más frecuentemente se usan en biotecnología, en particular para la producción de proteínas recombinantes o de partículas tipo virus (VLP, es decir, recubrimientos desprovistos de ácidos nucleicos virales).

[0079] Las principales ventajas de la producción basada en el sistema de vector de expresión de baculovirus (BEVS) se pueden resumir como sigue: (i) la presencia de promotores muy fuertes (poliedrina o p10) habilita la producción de una alta cantidad de proteínas heterólogas sin limitación de tamaño de gen; (ii) las células de insecto poseen la capacidad de llevar a cabo las principales modificaciones postraduccionales, permitiendo de esta manera la producción de proteínas biológicamente activas; y (iii) la tecnología de baculovirus se puede implementar fácilmente, el crecimiento en escala se consigue fácilmente, las células se cultivan en suspensión y están disponibles comercialmente varios medios sin suero.

[0080] El ensamblaje de partículas virales es un proceso más complejo que la expresión de una única proteína. Sin embargo, ha mostrado que las VLP basadas en HBV, B19 parvovirus, rotavirus, papilomavirus humano podrían producirse exitosamente con el BEVS. Además, se puede usar un sistema de vector de expresión de baculovirus para la producción de rAAV (Merten *et al.*, *supra*, Urabe *et al.*, 2002).

[0081] Por lo tanto, en el método de la divulgación, un virión parvoviral se puede producir usando un sistema de vector de expresión de baculovirus en células mamíferas o en células de insecto. Preferiblemente, el virión parvoviral se produce usando un sistema de vector de expresión de baculovirus en células de insecto.

[0082] En una forma de realización preferida de la divulgación, el vector baculoviral comprende una secuencia de nucleótidos codificantes de Rep, Cap y/o un transgén, donde preferiblemente el vector baculoviral comprende una secuencia de nucleótidos codificantes de un transgén, donde más preferiblemente el vector baculoviral comprende una secuencia de nucleótidos que codifican un transgén que está flanqueado por al menos una ITR parvoviral, y donde más preferiblemente el vector baculoviral comprende una secuencia de nucleótidos que codifica un transgén que está flanqueado por al menos una ITR parvoviral a cada lado.

[0083] También se pueden emplear en la presente descripción modificaciones en las secuencias Rep y VP1, VP2 y VP3 descritas previamente, tales como, por ejemplo, las descritas en publicaciones internacionales WO 2007/046703, WO 2007/148971, WO 2009/014445, WO 2009/104964 y/o WO 2011/112089.

[0084] Las secuencias de ITR y Rep de AAV que se pueden usar en un método de la presente descripción para la producción de vectores de rAAV pueden derivarse del genoma de cualquier serotipo de AAV. Generalmente, los serotipos de AAV tienen secuencias genómicas de homología significativa en los niveles de aminoácido y ácido nucleico. Esto proporciona un conjunto idéntico de funciones genéticas para producir viriones que son equivalente esencial, física y funcionalmente. Para la secuencia genómica de los diversos serotipos de AAV y una visión de conjunto de las similitudes genómicas véase, por ejemplo, número de registro de GenBank U89790; número de registro de GenBank J01901; número de registro de GenBank AF043303; número de registro de GenBank AF085716; Chiorini *et al.* (1997, J. Vir. 71: 6823-33); Srivastava *et al.* (1983, J. Vir. 45:555-64); Chiorini *et al.* (1999, J. Vir. 73:1309-1319); Rutledge *et al.* (1998, J. Vir. 72:309-319) y Wu *et al.* (2000, J. Vir. 74: 8635-47). Se pueden usar los serotipos de rAAV 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 y 12 como una fuente de secuencias de nucleótidos de AAV para usar en el contexto de la presente divulgación, los serotipos de rAAV 1, 2, 3, 4 y 5 son una fuente preferida de secuencias de nucleótidos de AAV. Preferiblemente, las secuencias de ITR de AAV en el contexto de la presente descripción se derivan de AAV1, AAV2 y/o AAV5. Más preferiblemente, las secuencias de ITR en un método de la presente descripción son ITR de AAV2. Asimismo, las secuencias codificantes de Rep (Rep78/68 y Rep52/40) se derivan preferiblemente de AAV1, AAV2 y/o AAV5, más preferiblemente AAV2.

[0085] Las secuencias Rep e ITR de AAV se conservan particularmente entre la mayoría de serotipos. Las proteínas Rep78 de varios serotipos de AAV son, por ejemplo, más de un 89 % idénticas y la identidad de secuencia de nucleótidos total en el nivel de genoma entre AAV2, AAV3A, AAV3B y AAV6 es de alrededor del 82 % (Bantel-Schaal *et al.*, 1999, J. Virol., 73(2):939-947). Además, las secuencias Rep e ITR de muchos serotipos de AAV se conocen por el hecho de que complementan de forma cruzada eficazmente (es decir, sustituyen funcionalmente) las secuencias correspondientes de otros serotipos en la producción de partículas de AAV en células mamíferas. La patente S2003148506 informa de que las secuencias Rep y ITR de AAV también complementan de forma cruzada eficazmente otras secuencias Rep y ITR de AAV en células de insecto.

[0086] Las proteínas VP de AAV se conocen por determinar la tropicidad celular del virión de AAV. Las secuencias codificantes de proteína VP están significativamente menos conservadas que las proteínas Rep y los genes entre serotipos de AAV diferentes. En una forma de realización preferida, el vector de rAAV comprende proteínas VP1. La capacidad que tienen las secuencias Rep e ITR para complementar de forma cruzada secuencias correspondientes de otros serotipos permite la producción de partículas de rAAV pseudotipadas que comprenden las proteínas de cápside de un serotipo (por ejemplo, AAV5) y las secuencias Rep y/o ITR de otro serotipo de AAV (por ejemplo, AAV2). Tales partículas de rAAV pseudotipadas son una parte del método de la presente invención. Aquí, una partícula de rAAV pseudotipada se pueden designar como siendo tipo "x/y", donde "x" indica la fuente de ITR e "y" indica el serotipo de cápside, por ejemplo un partícula 2/5 de rAAV tiene ITR de AAV2 y una cápside de AAV5.

[0087] Las secuencias de "AAV" modificadas también se pueden usar en el contexto de la presente descripción, por ejemplo para la producción de vectores de rAAV en células de insecto. Tales secuencias modificadas incluyen, por ejemplo, secuencias que tienen al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 % o más identidad de secuencia de nucleótidos y/o aminoácidos (por ejemplo, una secuencia con de aproximadamente el 75 % a aproximadamente el 99 % de identidad de secuencia de nucleótidos) con una ITR, Rep, o VP de AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8 o AAV9 se pueden usar en lugar de secuencias ITR, Rep o VP de AAV de tipo silvestre.

[0088] Se pueden modificar secuencias ITR, Rep y Cap como se desee para obtener una producción eficiente de vectores de rAAV o rAAV pseudotipado en células, tales como células de insecto. Por ejemplo, se pueden modificar el codón de inicio de las secuencias Rep, los sitios de empalme de VP se pueden modificar o eliminar, y/o el codón de inicio de VP1 se puede modificar para mejorar la producción de vectores de rAAV en la célula (de insecto), tal y como se describe, por ejemplo, en las patentes WO 2007/046703, WO 2007/148971 y/o WO 2009/014445. En la presente descripción también se incluyen cápsides de AAV quiméricas, donde, por ejemplo la VP1 de AAV5 está parcial o totalmente sustituida por VP1 derivada de AAV2 y VP2 y 3 se derivan de AAV5 (Urabe *et al.*, 2006; WO2000/028004). Los vectores adenovirales preferidos se modifican para reducir la respuesta de huésped como ha sido analizado por Russell (2000, J. Gen. Virol. 81: 2573-2604), o descrito en la patente US20080008690 y por Zaldumbide y Hoeben (Gene Therapy 2008:239-246).

[0089] Preferiblemente, una proteína Rep de AAV comprendida dentro de un vector de terapia génica según la presente descripción es una proteína Rep de serotipo 2 de AAV 2. Aún más preferiblemente, una proteína Rep78 que tiene la secuencia de ácidos nucleicos de la SEQ ID N.º: 4 y/o una secuencia de aminoácidos según la SEQ ID N.º: 5 se emplea en la presente invención y una proteína Rep52 con la secuencia de ácidos nucleicos de la SEQ ID N.º: 6 se emplea en la presente descripción.

#### Quantificación de una impureza de ácido nucleico sobrerrepresentada

[0090] De forma alternativa o en combinación con cualquiera de las formas de realización tal y como se ha descrito previamente, la presente divulgación se refiere además a la cuantificación de una impureza de ácido nucleico sobrerrepresentada. En particular, la presente divulgación se refiere al descubrimiento de que están sobrerrepresentadas impurezas de ADN específicas en una composición que comprende un vector parvoviral. Estas impurezas de ADN comprenden secuencias de nucleótidos que flanquean inmediatamente las ITR de una secuencia de nucleótidos parvoviral durante el proceso de producción de un virión parvoviral. De este modo, por ejemplo, las secuencias corriente arriba de una ITR parvoviral izquierda y/o las secuencias corriente abajo de una ITR parvoviral derecha están sobrerrepresentadas en el virión parvoviral.

[0091] La divulgación se refiere a un método para cuantificar una impureza de ácido nucleico en una composición que comprende un vector parvoviral, donde el método comprende la etapa de determinar la abundancia relativa de una impureza de ácido nucleico, donde dicha impureza de ácido nucleico comprende una secuencia de nucleótido situada entre 1 - 8000 pares de bases, 1 - 5000 pares de bases, 1 - 3000 pares de bases, 1 - 1000 pares de bases, 1 - 500 pares de bases, 1 - 250 pares de bases o 1 - 100 pares de bases inmediatamente adyacentes de una secuencia de ITR parvoviral cuando la secuencia de ITR está presente en un componente biológico usado en un proceso de producción de la composición y donde el componente biológico comprende un transgén flanqueado por al menos una copia de la secuencia de ITR parvoviral. En otra forma de realización, la divulgación se refiere a un método para cuantificar una impureza de ácido nucleico en una composición que comprende un vector parvoviral, donde el método consiste en la etapa de determinar la abundancia relativa de una impureza de ácido nucleico, donde dicha impureza de ácido nucleico comprende una secuencia de nucleótidos situada entre 1 - 8000 pares de bases, 1 - 5000 pares de bases, 1 - 3000 pares de bases, 1 - 1000 pares de bases, 1 - 500 pares de bases, 1 - 250 pares de bases o 1 - 100 pares de bases inmediatamente adyacentes de una secuencia de ITR parvoviral cuando la secuencia de ITR está presente en un componente biológico usado en un proceso de producción de la composición y donde el componente biológico comprende un transgén flanqueado por al menos una copia de la secuencia de ITR parvoviral.

[0092] Tal y como se ha indicado previamente, durante el proceso de producción de viriones parvovirales, una secuencia parvoviral puede estar presente en una célula huésped, un plásmido, un vector y/o un virus auxiliar, donde el vector preferible es un vector baculoviral. La secuencia parvoviral puede comprender al menos una copia de la ITR y un casete de expresión para la expresión de un transgén. Las impurezas de ADN sobrerrepresentadas pueden comprender cualquier secuencia que flanquee inmediatamente la ITR o las ITR parvovirales, tales como secuencias genómicas, secuencias de plásmido, secuencias de vector o secuencias de un virus auxiliar. Por lo tanto, el tipo de impureza de ADN depende de las secuencias que flanquean la ITR durante la producción de viriones parvovirales. Por ejemplo, si la secuencia parvoviral, que comprende al menos una copia de la ITR y preferiblemente un transgén, está presente en un vector baculoviral, las secuencias baculovirales que flanquean inmediatamente la ITR o las ITR estarán sobrerrepresentadas en la composición que comprende un vector parvoviral.

[0093] En una forma de realización de la divulgación, la impureza de ácido nucleico se cuantifica en una composición que comprende un vector parvoviral. El método para cuantificar la impureza de ácido nucleico puede comprender cualquier método conocido en la técnica para cuantificar un ácido nucleico. Tales métodos incluyen, pero no se limitan a, secuenciación de alto rendimiento, Q-PCR, PCR de ciclo limitado, un ensayo de hibridación, una micromatriz y una electrofóresis de agarosa.

[0094] En otra forma de realización, el componente biológico se define tal y como se ha indicado previamente. En particular, el componente biológico se selecciona a partir de un grupo que consiste en una célula huésped, un plásmido, un vector diferente del vector parvoviral recombinante y un virus auxiliar. Preferiblemente, el componente biológico comprende una secuencia parvoviral, donde preferiblemente la secuencia parvoviral comprende al menos una ITR y una secuencia de nucleótidos codificante para un transgén. Más preferiblemente, el componente biológico comprende una secuencia de nucleótidos codificante para un transgén que está flanqueado por al menos una ITR parvoviral a cada lado.

En la forma de realización más preferida, el componente biológico es un vector y donde el vector es un vector baculoviral. El vector baculoviral comprende preferiblemente una secuencia parvoviral. Esta secuencia parvoviral comprende preferiblemente al menos una ITR y una secuencia de nucleótidos codificante para un transgén. Más preferiblemente, el vector baculoviral comprende una secuencia de nucleótidos codificante para un transgén que está flanqueado por al menos una ITR parvoviral a cada lado.

[0095] En un método según la divulgación, la impureza de ácido nucleico comprende una secuencia de nucleótidos que está situada entre 1- 10000 pares de bases, 1 - 9000 pares de bases 1 - 8000 pares de bases, 1 - 7000 pares de bases, 1 - 6000 pares de bases 1 - 5000 pares de bases, 1 - 4000 pares de bases, 1 - 3000 pares de bases, 1 - 2000 pares de bases 1 - 1000 pares de bases, 1 - 800 pares de bases, 1 - 600 pares de bases, 1 - 500 pares de bases, 1 - 400 pares de bases, 1 - 250 pares de bases o 1 - 100 pares de bases inmediatamente adyacentes de una secuencia de ITR parvoviral cuando esté presente en un componente biológico usado en un proceso de producción de la composición. La longitud de la impureza de ADN puede depender de la presencia de otras impurezas de ADN empaquetadas de forma aleatoria y/o del tamaño del transgén ya que hay una capacidad de empaquetamiento máxima de un virión parvoviral. Sin embargo, es conocido que un virión parvoviral puede incorporar secuencias de ADN más largas que la longitud de su propio genoma (Grieger *et al.*, J Virol. 2005 79(15):9933-44).

[0096] En una forma de realización preferida de la divulgación, el vector parvoviral es un vector de virus adenoasociado recombinante (rAAV) tal y como se ha descrito anteriormente. Además, el virión de rAAV se puede producir usando cualquiera de los métodos de producción tal y como se han descrito previamente.

[0097] En otra forma de realización, la secuencia de nucleótidos de la impureza de ácido nucleico está situada inmediatamente adyacente a cada lado de la secuencia de ITR parvoviral cuando la secuencia de ITR está presente en un componente biológico usado en un proceso de producción de la composición. La secuencia parvoviral presente en un componente biológico puede comprender un transgén que está flanqueado por al menos una ITR a cada lado del transgén. En tal caso, la impureza de ácido nucleico puede comprender secuencias de nucleótidos que están presentes en un lado de las ITR, es decir solo inmediatamente adyacentes a la o las ITR izquierdas o solo inmediatamente adyacentes a la o las ITR derechas. De forma alternativa, la secuencia de nucleótidos de la impureza de ácido nucleico puede comprender secuencias de nucleótidos que están presentes a ambos lados de las ITR.

[0098] En una forma de realización de la divulgación, la impureza de ácido nucleico comprende una secuencia de nucleótidos que está situada inmediatamente adyacente a una secuencia de ITR parvoviral, cuando la secuencia de ITR está presente en un componente biológico usado en un proceso de producción de la composición. "Inmediatamente adyacente" se define aquí como sigue: en caso de una ITR corriente arriba del transgén, "inmediatamente adyacente" designa cualquier secuencia de nucleótidos que termina al menos 0, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40 o 50 pares de bases corriente arriba de la ITR. En el caso de que la ITR esté situada corriente abajo del transgén, "inmediatamente adyacente" designa cualquier secuencia de nucleótidos que comience al menos 0, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40 o 50 pares de bases corriente abajo de la ITR.

Abundancia relativa

5 [0099] En una forma de realización de la divulgación, el método comprende la etapa de determinar la abundancia relativa de una impureza de ácido nucleico. La abundancia relativa se define aquí como la presencia de (parte de) la primera molécula de ácido nucleico en comparación con la presencia de (parte de) una segunda molécula de ácido nucleico en la misma composición o en otra. En el caso de que la abundancia relativa se determine entre dos o más composiciones, la segunda molécula de ácido nucleico puede comprender la misma u otra secuencia de nucleótidos que la primera molécula de ácido nucleico. En el caso de que la abundancia relativa se determine en una composición, la primera y la segunda moléculas de ácido nucleico comprenden al menos  
10 parcialmente una secuencia de nucleótidos diferente.

15 [0100] En el método de la divulgación, la abundancia relativa entre una impureza de ácido nucleico y una segunda molécula de ácido nucleico se determina preferiblemente en la misma composición, aunque también constituye una forma de realización de la divulgación que la abundancia relativa de una impureza de ácido nucleico se determine entre composiciones diferentes. En un método preferido de la divulgación, la abundancia relativa se determina en comparación con una secuencia de nucleótidos del vector parvoviral y/o una secuencia de referencia en la composición. Preferiblemente, el transgén y/o secuencia de referencia está presente en la misma composición que la impureza de ácido nucleico.

20 [0101] Una secuencia de nucleótidos del vector parvoviral puede comprender cualquier secuencia del vector parvoviral, tal como el vector parvoviral completo, o (una parte de la) secuencia de nucleótidos del transgén, el promotor o la o las ITR. Sin embargo, cualquier otra secuencia de nucleótidos del vector parvoviral puede ser igualmente adecuada para usar en el método de la divulgación.

25 [0102] La secuencia de referencia puede ser cualquier (parte de una) secuencia de nucleótidos adecuada. Preferiblemente, la secuencia de referencia es (parte de una) secuencia de un gen constitutivo, una secuencia de nucleótidos de un componente biológico y/o la secuencia de referencia es una secuencia de un ácido nucleico que se usa para impulsar la composición.

30 [0103] En el caso de que la secuencia de referencia sea una secuencia de un componente biológico, la secuencia no flanquea inmediatamente una ITR parvoviral y la secuencia de referencia se deriva preferiblemente de una célula huésped, un plásmido o un vector usado en el proceso de producción de la composición. Preferiblemente, la secuencia de referencia se deriva del mismo componente biológico que comprende un transgén y al menos una ITR. Más preferiblemente, la secuencia de referencia comprende una secuencia a partir  
35 de un vector baculoviral, donde dicho vector baculoviral se usa en un proceso de producción de la composición, que comprende un transgén y al menos una ITR y donde dicha secuencia baculoviral de referencia no flanquea inmediatamente una ITR parvoviral.

40 [0104] En el caso de la secuencia de que la referencia sea un ácido nucleico usado para impulsar la composición, se prevé que este ácido nucleico no esté presente después de la producción de la composición sino que se añade en punto temporal posterior, pero antes de determinar la abundancia relativa de la impureza de ácido nucleico. La molécula de ácido nucleico usada para impulsar la composición puede ser cualquier molécula de ácido nucleico adecuada, tal como una molécula de ARN o ADN lineal o circular pequeña de al menos 10, 30, 50, 100, 150, 200, 500, 1000 o más pares de bases. Dicha molécula de ácido nucleico puede  
45 comprender una región codificante y/o no codificante. En otra forma de realización de la divulgación, la abundancia relativa se determina mediante:

50 a) el número medio de lecturas por ácido nucleico de la impureza de ácido nucleico tal y como se ha definido anteriormente; y

- i) el número medio de lecturas por ácido nucleico de la secuencia de referencia; y/o
- ii) el número medio de lecturas por vector parvoviral en la composición;

55 donde el número de lecturas se determina por cualquier método tal y como se ha resumido previamente; y/o b) la amplificación de la impureza de ácido nucleico tal y como se ha definido anteriormente;

- i) la secuencia de referencia; y/o
- ii) la secuencia de nucleótidos del vector parvoviral.

60 [0105] En una forma de realización preferida, el número medio de lecturas se define tal y como se ha indicado previamente. Además, el vector parvoviral puede estar relacionado a cualquier secuencia del vector parvoviral, tal como el vector parvoviral completo, o solo (una parte de la) secuencia de nucleótidos del transgén, el promotor o la o las ITR. Además, cualquier otra secuencia de nucleótidos del vector parvoviral puede ser igualmente adecuada para usar en el método de la divulgación.  
65

[0106] En otra forma de realización, la abundancia relativa de la impureza de ácido nucleico se puede determinar mediante cualquier método adecuado para determinar la cantidad de una molécula de ácido nucleico tal y como se ha indicado previamente. Más preferiblemente, la abundancia relativa se determina mediante Q-PCR y/o por secuenciación de alto rendimiento. Cualquier método de Q-PCR o secuenciación de alto rendimiento que tenga como resultado la cuantificación de un ácido nucleico es adecuado para usar en el método de la invención. Los métodos de Q-PCR (reacciones en cadena de polimerasa en tiempo real) son bien conocidos en la técnica y la técnica puede usar fluorocromos no específicos o sondas de hibridación. En una forma de realización preferida de la divulgación, la Q-PCR se lleva a cabo con sondas de hibridación específicas.

[0107] Preferiblemente, el método comprende además la etapa de una hibridación selectiva de un cebador de oligonucleótidos a la impureza de ácido nucleico tal y como se ha definido previamente.

[0108] La hibridación selectiva de un cebador de oligonucleótidos a una impureza de ácido nucleico se entiende con el significado de que el oligonucleótido forma un dúplex productivo o positivo con la impureza de ácido nucleico. La formación de dicho dúplex productivo o positivo se entiende como la formación de un dúplex entre el oligonucleótido y la impureza de ácido nucleico que se puede detectar mediante la formación de un amplicón en el ensayo de Q-PCR. En la práctica, esto significará que el extremo del cebador de oligonucleótidos formará un dúplex con la impureza de ácido nucleico, de manera que el oligonucleótido se puede alargar por una polimerasa o ligar a una molécula de poli o oligonucleótido apareada de base de forma adyacente. Tal y como se utiliza en la presente, un "amplicón" se refiere a un segmento de ácido nucleico bicatenario que tiene un tamaño y secuencia definidos resultantes de un procedimiento de amplificación, tal como un procedimiento de PCR. El tamaño del amplicón se determina por los sitios en las dos cadenas de un dúplex de ácido nucleico a las que se enlazan los cebadores de oligonucleótidos. Tal y como se explica en la patente de EE. UU. N.º 4,683,195, dicho segmento del ácido nucleico de producto se convierte en el producto predominante del procedimiento de amplificación después un pequeño número de ciclos de amplificación. Además, una secuencia es "específica" o "selectiva" para la impureza de ácido nucleico siempre y cuando hibride eficazmente a la secuencia objetivo pero no hibride a cualquier secuencia que no sea una impureza de ácido nucleico tal y como se ha definido anteriormente, bajo las condiciones usadas en circunstancias experimentales determinadas.

[0109] En una forma de realización preferida, el cebador de oligonucleótidos hibrida selectivamente a una impureza de ácido nucleico que comprende una parte de una secuencia de baculovirus o un complemento de la misma. El término "complemento" o "secuencia complementaria" de una primera secuencia se entiende aquí para designar la segunda secuencia que puede formar una estructura bicatenaria o dúplex con la primera secuencia mediante emparejamiento de pares de bases, por ejemplo la secuencia complementaria de G-T-A-C es C-A-T-G.

[0110] Por lo tanto, se prefiere que la impureza de ácido nucleico se derive de un vector baculoviral usado en el proceso de producción de la composición. En particular, se prefiere que dicho vector baculoviral comprenda un transgén flanqueado por al menos una ITR parvoviral. En una forma de realización preferida, el cebador de oligonucleótidos hibrida selectivamente a una impureza de ácido nucleico, donde la impureza de ácido nucleico comprende una secuencia de nucleótidos derivada del baculovirus que se sitúa entre 1- 10000 pares de bases, 1 - 9000 pares de bases 1 - 8000 pares de bases, 1 - 7000 pares de bases, 1 - 6000 pares de bases 1 - 5000 pares de bases, 1 - 4000 pares de bases, 1 - 3000 pares de bases, 1 - 2000 pares de bases 1 - 1000 pares de bases, 1 - 800 pares de bases, 1 - 600 pares de bases, 1 - 500 pares de bases, 1 - 400 pares de bases, 1 - 250 pares de bases o 1 - 100 pares de bases inmediatamente adyacentes de la secuencia de ITR parvoviral, cuando la secuencia de ITR parvoviral está presente en el vector baculoviral.

#### Aplicación clínica

[0111] Una composición que comprende un vector parvoviral no debería contener un alto grado de impurezas de ácido nucleico, especialmente si la composición se va a usar en un tratamiento médico. En particular, tales impurezas de ácido nucleico podrían causar reacciones adversas normalmente en pacientes ya vulnerables, lo que podría provocar complicaciones graves. La divulgación actual se refiere al descubrimiento de que las impurezas de ADN no están encapsuladas de forma aleatoria dentro de un virión parvoviral. En cambio, están sobrerrepresentadas secuencias que flanquean la ITR durante la producción de viriones parvovirales. En otro aspecto, la divulgación actual se refiere, por lo tanto, a un método de determinación de si una composición que comprende un vector parvoviral se considera clínicamente puro, donde el método comprende las etapas de:

- i) cuantificar una impureza de ácido nucleico en una composición que comprende un vector parvoviral según cualquier método tal y como se ha resumido previamente; y
- ii) determinar la composición como clínicamente pura si la impureza de ácido nucleico tal y como se ha definido anteriormente está al menos 10, 100, 250, 1000 veces menos presente que la secuencia de referencia y/o transgén tal y como se ha determinado por la abundancia relativa de la impureza de ácido nucleico.



[0112] Clínicamente puro se define aquí como un producto de alta calidad farmacéutica, que es una composición que se considera segura para su administración a animales, preferiblemente un producto de alta calidad farmacéutica es un producto que se considera seguro para su administración a mamíferos, más preferiblemente la composición se considera segura para su administración a seres humanos.

[0113] En una forma de realización preferida de la invención, la impureza de ácido nucleico tal y como se ha definido anteriormente está al menos 10, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 10 000 o 100 000 veces menos presente que el transgén o la secuencia de referencia. La presencia de la impureza de ácido nucleico y el transgén o secuencia de referencia se puede cuantificar mediante cualquier método convencional para cuantificar una secuencia de ADN específica.

[0114] Por clínicamente puro se entiende aquí adicionalmente que la composición que comprende el vector parvoviral se considera lo suficientemente pura para el tratamiento clínico. En particular, una composición clínicamente pura según la invención comprende un grado bajo de impurezas de ADN tal y como se solicita en las pautas de calidad y seguridad la Conferencia Internacional sobre armonización de requisitos técnicos para el registro de productos farmacéuticos para uso humano (ICH). Preferiblemente, el nivel de impurezas de ADN se encuentra por debajo de un nivel que cause cualquier efecto adverso en los pacientes.

[0115] En este documento y en sus reivindicaciones, el verbo "comprender" y sus conjugaciones se usa en su sentido no limitativo para designar que los elementos que siguen a la palabra están incluidos, pero que los elementos que no se mencionen específicamente no están excluidos. Además, la referencia a un elemento mediante el artículo indefinido "un" o "uno" no excluye la posibilidad de que esté presente más de uno del elemento, a menos que el contexto requiera claramente que haya uno y solo uno de los elementos. El artículo indefinido "un" o "uno" designa, por lo tanto, normalmente "al menos uno".

[0116] Los ejemplos siguientes se proporcionan únicamente con fines ilustrativos, y no se destinan a limitar el alcance de la presente invención en modo alguno.

#### Referencias

[0117]

1. Fujita, R., Matsuyama, T., Yamagishi, J., Sahara, K., Asano, S., y Bando, H. (2006) Expression of Autographa californica multiple nucleopolyhedrovirus genes in mammalian cells and upregulation of the host beta-actin gene, *J. Virol.* 80, 2390-2395.
2. Liu, C. Y., Wang, C. H., Wang, J. C., y Chao, Y. C. (2007) Stimulation of baculovirus transcriptome expression in mammalian cells by baculoviral transcriptional activators, *J. Gen. Virol.* 88, 2176-2184.
3. Laakkonen, J. P., Kaikkonen, M. U., Ronkainen, P. H., Ihalainen, T. O., Niskanen, E. A., Hakkinen, M., Salminen, M., Kulomaa, M. S., Yla-Herttuala, S., Airene, K. J., y Vihinen-Ranta, M. (2008) Baculovirus-mediated immediate-early gene expression and nuclear reorganization in human cells, *Cell Microbiol.* 10, 667-681.
4. Blouin, V., Brument, N., Toublanc, E., Raimbaud, I., Moullier, P., y Salvetti, A. (2004) Improving rAAV production and purification: towards the definition of a scaleable process, *J. Gene Med.* 6 Suppl 1, S223-S228.
5. Nony, P., Chadeuf, G., Tessier, J., Moullier, P., y Salvetti, A. (2003) Evidence for packaging of rep-cap sequences into adeno-associated virus (AAV) type 2 capsids in the absence of inverted terminal repeats: a model for generation of rep-positive AAV particles, *J. Virol.* 77, 776-781.
6. Chadeuf, G., Ciron, C., Moullier, P., y Salvetti, A. (2005) Evidence for encapsidation of prokaryotic sequences during recombinant adeno-associated virus production and their in vivo persistence after vector delivery, *Mol. Ther.* 12, 744-753.
7. Wright, J. F. (2008) Manufacturing and characterizing AAV-based vectors for use in clinical studies, *Gene Ther.* 15, 840-848.
8. Arsalan Haseeb Zaidi, Patrick J. Bakkes, Jacek Lubelski, Herfita Agustindari, Oscar P. Kuipers, y Arnold J. M. Driessen (2008) The ABC-Type Multidrug Resistance Transporter LmrCD Is Responsible for an Extrusion-Based Mechanism of Bile Acid Resistance in *Lactococcus lactis* *Journal of Bacteriology*, 7357-7366.
9. Jean - Marie Rouillard, Michael Zuker y Erdogan Gulari (2003) OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach, *Nucleic Acids Research*, Vol. 31, N.º. 12 3057-3062.
10. Van Hijum SA., de la Nava GJ, Trelles O., Kok J., Kuipers OP., (2003) MicroPreP: a cDNA microarray data pre-processing framework. *Appl Bioinformatics*, 2(4):241-4.
11. P. Baldi y A.D. Long, (2001) A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized t-Test and Statistical Inferences of Gene Changes, *Bioinformatics*, 17, 6, 509-519.
12. Krappa, R., Roncarati, R., Knebel-Morsdorf, D., (1995) Expression of PE38 and IE2, Viral members of the C3HC4 finger family, during baculovirus infection: PE38 and IE2 localize to distinct nuclear regions, *J Virol*, 5287-5293.

13. Gerhard Schwarz, Stefan BaÈumler, Annette Block, Friedrich G. Felsenstein y Gerhard Wenzel (2004) Determination of detection and quantification limits for SNP allele frequency estimation in DNA pools using real time PCR *Nucleic Acids Research*, Vol. 32, N.º. 3 e24.

14. Yaffe David, Saxel Ora, (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle, *Nature* 270, 725-727.

15. Manno, CS, Pierce, GF, Arruda, VR, Glader, B, Ragni, M, Rasko, J *et al.* (2006). Successful transduction of liver in hemophilia by AAV-factor IX and limitations imposed by the host immune response. *Nat Med* 12: 342-347.

16. Mingozzi, F, Maus, MV, Hui, DJ, Sabatino, DE, Murphy, SL, Rasko, JE *et al.* (2007). CD8+ T-cell responses to adeno-associated virus capsid in humans. *Nat Med* 13: 419-422.

17. Christine L. Halbert\*, Michael J. Metzger\*, Siu-Ling Lam, y A. Dusty Miller (2011) Capsid-expressing DNA in AAV vectors and its elimination by use of an oversize capsid gene for vector production. *Gene Ther.* 18(4): 411-417

18. Bernd Hauck, Samuel L Murphy, Peter H Smith, Guang Qu, Xingge Liu, Olga Zeleniaia, Federico Mingozzi, Jürg M Sommer, Katherine A High y J Fraser Wright (2009) Undetectable Transcription of cap in a Clinical AAV Vector: Implications for Preformed Capsid in Immune Responses. *Mol Ther.* 17(1) 144-152

### Descripción de las figuras

[0118]

Figura 1. Resistencia de transgén de AAV1 (paneles superiores) y ADN de baculovirus (paneles centrales e inferiores) a la DNasa I. La cantidad de ADN detectada por Q-PCR usando tres conjuntos de cebadores diferentes sin o con tratamiento de DNasa. Para cada conjunto de cebadores, se evaluaron dos lotes. Conjunto de cebadores A) y B) 59/60, conjunto de cebadores C) y D) 180/181, conjunto de cebadores E) y F) 340/341.

Figura 2. Mapa de secuencia del plásmido baculoviral Bac.VD. Se indican los conjuntos de cebadores e ITR usados.

Figura 3. Cantidad relativa de copias de genoma detectadas por los diferentes conjuntos de cebadores. En el eje se indica la ubicación de los amplicones. El amplicón 5214-5284 representa el promotor de CMV del casete transgén de AAV. El amplicón 73555-73604 es el objetivo del conjunto de cebadores 340/341 y se sitúa lo más lejos posible del casete de transgén de AAV. Cada punto representa una medición.

Figura 4. Se analizó el rAAV que comprende un transgén mediante secuenciación profunda. Las lecturas obtenidas se alinearon con el transgén (A), el casete de cap (B) o el casete de rep (c).

Figura 5. Se analizó el rAAV mediante secuenciación profunda. Las lecturas obtenidas se alinearon con el genoma baculoviral. Se representa la distribución de lecturas por nucleótido de la estructura baculoviral. El nucleótido 1 es la ITR derecha tal y como se indica en la figura 2.

Figura 6. Se evaluaron cinco lotes diferentes de vectores de rAAV para impurezas de ADN usando Q-PCR o secuenciación profunda con Illumina o Roche 454.

### Ejemplos

#### Ejemplo 1: impurezas de ADN en vectores de rAAV fabricados

##### 1.1 Material y métodos

[0119] Para investigar si el ADN residual estaba empaquetado en partículas de AAV1, se evaluó si el ADN residual era resistente a la DNasa. Las muestras se trataron con benzonasa (9U/mL) y se analizó la cantidad de ADN mediante Q-PCR.

[0120] Se aisló ADN de las muestras seguido de una Q-PCR usando tres conjuntos de cebadores diferentes (59/60, 180/181, 340/341). Para estudiar la resistencia a la DNasa del ADN de baculovirus, para algunas muestras se omitió la etapa de DNasa (indicado como sin DNasa). Los datos se analizaron mediante análisis de PLA y para cada muestra se determinaron las proporciones de las cantidades de ADN amplificadas por los diferentes conjuntos de cebadores. La cantidad de ADN de AAV1 se determinó mediante Q-PCR con el conjunto de cebadores 59/90 con el promotor de CMV del vector de transgén de AAV1 como objetivo. La cuantificación de ADN de baculovirus residual se llevó a cabo mediante Q-PCR con cebadores específicos de baculovirus. Los experimentos se llevaron a cabo usando dos conjuntos de cebadores diferentes; el conjunto de cebadores 180/181 tiene como objetivo el ORF 1629 del ADN de baculovirus cerca del casete de transgén de AAV y el conjunto de cebadores 340/341 tiene como objetivo la secuencia de baculovirus hr3, detectando el ADN de

baculovirus situado lejos del casete de transgén de AAV1. Para estos experimentos se incluyeron dos estándares: línea estándar de plásmido (PVD) y ADN de baculovirus purificado de clon de VD.

[0121] Para determinar la cantidad de ADN de baculovirus usando el conjunto de cebadores 180/181, se utilizó como estándar pVD con el conjunto de cebadores 180/181. La concentración de pVD se determinó con mediciones de densidad óptica (OD). Para determinar la cantidad de ADN de baculovirus usando los conjuntos de cebadores 340/341, se utilizó como estándar BacVD con el conjunto de cebadores 340/341. La cantidad de BacVD para la línea estándar se determinó mediante Q-PCR usando el conjunto de cebadores 180/181 con pVD como estándar. La cantidad de ADN (GC/ml) se calculó usando la fórmula:

$$[ADN] = S \cdot D \cdot C$$

donde:

- S = cantidad media medida (gc)
- D = factor de dilución de ADN viral (500 veces o 1000 veces)
- C = factor de corrección para calcular de muestra de 10 µl a muestra de 1 mL (100)

Para calcular la cantidad de ADN en µg/mL, la fórmula se extendió a:

$$[ADN] = \frac{S \cdot D \cdot C \cdot X}{A} \cdot PM \Rightarrow \dots \mu g / mL$$

donde:

- X = factor de conversión para g a µg ( $10^6$ )
- A = número de Avogadro ( $6,022 \times 10^{23}$ )
- PM = peso molar de ADN. El genoma de baculovirus consiste en 135 kbp de ADN bicatenario. El peso molar medio por pares de bases es de 649 Da. Como PM para el ADN de baculovirus (después de la determinación usando los conjuntos de cebadores 180/181 o 340/341), se usó un PM de  $13500 \cdot 649 = 8,76 \times 10^7$  Da.
- El genoma de AAV1 consiste en ADN monocatenario de 3630 pares de bases. Para calcular la cantidad de ADN de AAV1, se usó un PM de  $3630 \text{ pares de bases} \cdot 340 \text{ Da} = 1,23 \times 10^6$  Da.

## 1.2 Resultados

[0122] La cantidad de ADN de baculovirus se determinó mediante Q-PCR usando dos conjuntos de cebadores diferentes. El conjunto de cebadores 180/181 detecta una secuencia en el ORF 1629, cerca del casete de transgén de AAV, mientras que la secuencia para el conjunto de cebadores 340/341 está situada lejos del casete. Los resultados de la tabla 1 muestran que los dos conjuntos de cebadores producen valores muy diferentes para la cantidad de GC/ml de ADN de baculovirus (el conjunto de cebadores 180/181 produce de media valores 20 veces más altos que los obtenidos por el conjunto de cebadores 340/341).

Tabla 1. Concentración de genoma de rAAV previsto y DAN contaminante

Lote	ADN de AAV (cebadores 59...90)	ADN de baculovirus (cebador 180...181)	ADN baculovirus (cebador 340..341)
1	$5,9 \times 10^{12}$	$3,2 \times 10^{11}$	$1,4 \times 10^{10}$
2	$5,6 \times 10^{12}$	$2,9 \times 10^{11}$	$1,5 \times 10^{10}$
3	$6,5 \times 10^{12}$	$3,0 \times 10^{11}$	$1,3 \times 10^{10}$
4	$7,1 \times 10^{11}$	$3,7 \times 10^{10}$	$1,5 \times 10^9$
5	$7,0 \times 10^{12}$	$3,0 \times 10^{11}$	$1,2 \times 10^{10}$
6	$8,5 \times 10^{11}$	$3,5 \times 10^{10}$	$2,7 \times 10^9$
7	$9,1 \times 10^{12}$	$3,0 \times 10^{11}$	$2,0 \times 10^{10}$
8	$8,9 \times 10^{11}$	$3,9 \times 10^{10}$	$2,0 \times 10^9$
9	$8,4 \times 10^{11}$	$3,9 \times 10^{10}$	$1,9 \times 10^9$
10	$1,1 \times 10^{12}$	$3,6 \times 10^{10}$	$1,7 \times 10^9$
11	$3,0 \times 10^{12}$	$1,3 \times 10^{11}$	$5,8 \times 10^9$
12	$3,3 \times 10^{12}$	$1,3 \times 10^{11}$	$5,9 \times 10^9$
13	$2,9 \times 10^{12}$	$1,2 \times 10^{11}$	$6,1 \times 10^9$

[0123] La concentración del estándar de Bac.VD se corrigió mediante Q-PCR, por lo tanto, se puede excluir que esta diferencia esté relacionada con el estándar. Por lo tanto, estos datos indican que el ADN de baculovirus

cercano a las ITR (detectado con el conjunto de cebadores 180/181) está mucho más presente que el ADN distante de las ITR (detectado por el conjunto de cebadores 340/341).

## Ejemplo 2: determinación de impurezas de ácido nucleico mediante Q-PCR

5

### 2.1 Material y métodos

[0124] Para investigar adicionalmente qué partes del genoma de baculovirus estaban presentes en las muestras y posibles diferencias en la cantidad de secuencias diferentes, se llevó a cabo una Q-PCR con conjuntos de cebadores diferentes (véase la figura 2). Para cada conjunto de cebadores, se incluyó una línea estándar en el experimento. La cantidad de copias de transgén se determinó utilizando el conjunto de cebadores 59/60. Posteriormente, se determinó la cantidad relativa de copias de genoma en comparación con las copias de transgén. Dado que es conocido el hecho de que una partícula de AAV puede incorporar secuencias de ADN más largas que la longitud de su propio genoma (Grieger *et al.* 2005, Allocca *et al.* 2008), los cebadores se eligieron al principio y en el extremo de los ORF que flanqueaban el casete de transgén y 10 kb corriente arriba y abajo de las ITR.

### 2.2 Resultados

[0125] La cantidad de ADN de baculovirus se determinó utilizando el conjunto de cebadores 340/34, amplificando el amplicón 73555-73604 que está situado cerca de la secuencia hr3 del baculovirus. Se asumió que la cantidad de copias de genoma determinadas con estos cebadores es representativa de todo el genoma de baculovirus. Sin embargo, experimentos similares usando conjuntos de cebadores diferentes que tienen como objetivo secuencias más cercanas al casete de transgén de AAV mostraron que se halló un número superior de copias de genoma cuando el amplicón estaba situado más cerca del casete de transgén de AAV que contiene las ITR (figura 3). Dado que es conocido el hecho de que el AAV puede ser capaz de empaquetar secuencias de ADN mayores (hasta 8,9 kb, probablemente incluso mayores (Allocca *et al.*, 2008)), se espera que dichas secuencias estén empaquetadas dentro de la partícula. Esto implica que hay dos tipos de secuencias de ADN residual de baculovirus; 1) secuencias aleatorias, como se determinó con el conjunto de cebadores 340/341, que solo se hallan en un 0,1 % de la cantidad de copias de genoma del transgén y 2) secuencias de baculovirus dentro de 10 kb del casete de transgén (en rango de límite de empaquetamiento de AAV), que se hallan entre el 1 y el 2,5 % de la cantidad de copias de genoma de transgén. Ambas secuencias están presumiblemente empaquetadas dentro de la partícula de AAV1 o asociadas a la cápside.

## Ejemplo 3: determinación de impurezas de ácido nucleico mediante secuenciación de nueva generación

### 3.1 Material y métodos

[0126] Para investigar el alcance y el origen de las impurezas de ADN en vectores de rAAV fabricados, se analizaron cuatro lotes diferentes de vectores de rAAV mediante secuenciación profunda. Se aisló ADN de estos vectores de rAAV usando Nucleospin extract II kit (Macherey Nagel, Duren, Alemania).

[0127] Este ADN se utilizó para preparar las bibliotecas de secuenciación profunda.

[0128] Para crear características de secuenciación separadas se lleva a cabo una hibridación *in situ*. Las agrupaciones se llevan a cabo mediante diluciones límite de un material inicial. Los fragmentos de ADN se derriten y las cadenas simples se retienen dentro de la célula de flujo que está cubierta por una densa superficie de cebadores. La amplificación local posterior (PCR de puente) provoca la formación de agrupación de aproximadamente 1000 moléculas idénticas por micrómetro cuadrado. La incorporación de base comienza por la adición de cebadores, polimerasa y cuatro desoxirribonucleótidos trifosfato marcados con fluoróforo. Los dNTP actúan como terminadores reversibles, es decir, solo se añade una única base por molécula en cada ciclo. La fluorescencia de la agrupación se mide para identificar qué base se ha incorporado. Un láser verde identifica la incorporación de las bases G y T y un láser de lectura identifica las bases A y C. Se utilizan dos filtros diferentes para distinguir entre G/T y A/C, respectivamente. Después de la detección de señal, el fluoróforo y la modificación terminante del nucleótido se eliminan (Dohm, J. C., Lottaz, C., Borodina, T., y Himmelbauer, H. (2008), *Nucleic Acids Res.* 36, e105; Shendure, J. y Ji, H. (2008), *Nat. Biotechnol.* 26, 1135-1145; Rothberg, J. M. y Leamon, J. H. (2008), *Nat. Biotechnol.* 26, 1117-1124; Kahvejian, A., Quackenbush, J., y Thompson, J. F. (2008), *Nat.* Este método puede ser particularmente útil para determinar qué tipo de secuencia está presente como una impureza y cuáles son las proporciones entre las poblaciones de secuencia específica. El análisis fue realizado por ServiceXS (Leiden).

[0129] El experimento de secuenciación de nueva generación estándar tiene como resultado > 20 millones de lecturas cortas que necesitan ser alineadas con una secuencia de referencia o *de novo* ensamblada para producir secuencias contiguas. Aquí, tras la secuenciación del contenido total, se alinearon las lecturas con un número de secuencias de referencia. Estas secuencias de referencia representan moléculas de ADN que se conocen por estar presentes en las preparaciones de vector de rAAV. Esto incluye el genoma previsto y las

impurezas de ADN relacionadas con la producción. El alineamiento fue realizado mediante CLC\_bio aligner. La frecuencia en la cual se lee cada base en el experimento proporciona información acerca de su incidencia relativa en comparación con otras secuencias medidas. Las lecturas por nucleótido se recuperaron para cada secuencia de referencia (figura 4). Generalmente, se acepta que cuando los nucleótidos de la secuencia de referencia se leen más de 8-12 veces, la información de secuencia tiene un alto nivel de confianza (Schuster, S. C. (2008), Nat. Methods 5, 16-18).

3.2 Resultados

[0130] La secuenciación del ADN total se usó para analizar la composición de ADN de diferentes lotes de AAV. El análisis fue llevado a cabo por Baseclear (Leiden, El Países Bajos) basado en un procedimiento de secuenciación de lectura única de Illumina GAI-II. Los datos de secuencia brutos recortados de calidad resultantes se analizaron con una ayuda de CLC bio bioinformatic software. Las lecturas se ensamblaron por referencia sobre las secuencias de referencia que representaban las moléculas de ADN potencialmente presentadas en las preparaciones de vector de rAAV, es decir, estructura de baculovirus, ADN específico de cap, específico de rep y específico de transgén. Como se esperaba, la gran mayoría > 99,7 % del ~20 millón de lecturas generadas se ensambló con el casete de transgén de ADN destinado y con las impurezas de ADN relacionadas con la producción conocidas. Todas las demás secuencias (por debajo del 0,3 %) que no se ensamblaron con ninguna de las secuencias de referencia mencionadas pueden representar errores de secuenciación, multimerización de enlace, lecturas de baja calidad y otras secuencias de ADN.

[0131] Se recuperaron cuentas por nucleótido para el casete de transgén, el casete de cap, el casete de rep y el genoma de baculovirus y se fijaron contra un número de nucleótido (véanse las figuras 4 y 5). La distribución de las frecuencias de lectura por nucleótido fue altamente consistente entre las diferentes preparaciones de lote. Además, resultó evidente que la distribución de genoma de baculovirus no es aleatoria. Los segmentos de genoma que flanquean las ITR estaban claramente sobrerrepresentados (figura 5). Se ha usado una distribución media de las lecturas recuperada de los experimentos de secuenciación como una entrada para calcular la incidencia relativa de varias secuencias de ADN encontradas en las preparaciones de rAAV (tabla 2).

Tabla 2. Distribución media de lecturas (S) en 5 lotes de rAAV diferentes (lote #)

	Lote#1	Lote#2	Lote#3	Lote#4	Lote#5
casete de transgén	569837	588620	600677	589597	544717
Baculovirus	1184	1449	1236	1316	1087
casete de Cap	60	47	43	102	46
casete de Rep	460	676	596	691	562

La tabla 2 muestra frecuencias medias (S) recuperadas por secuencia determinada. Estas frecuencias se presentan en relación con el ADN principal en la muestra, es decir, el casete de transgén en la tabla 3. El porcentaje de una impureza determinada se calcula en relación con el casete de transgén según la fórmula presentada más adelante y tiene en cuenta el factor de corrección de tamaño.

$$X_{bac} = S_{bac} / S_{transgén} * C_{bac} * 100\%$$

Donde:

X<sub>bac</sub> - porcentaje de impurezas de ADN de baculovirus en relación con el casete de transgén

S<sub>bac</sub> - cuentas medias recuperadas para la estructura de baculovirus

S<sub>transgén</sub> - cuentas medias recuperadas para el casete de transgén

C<sub>bac</sub> - factor de corrección de longitud de molécula donde C<sub>bac</sub> = longitud de la estructura de baculovirus (nt) / longitud del casete de transgén (nt)

Tabla 3. Abundancia relativa de las diversas impurezas de ADN en comparación con el casete de transgén. Se presenta la distribución de cuenta media (S) de diferentes moléculas en relación con la distribución de cuenta de transgén (S<sub>transgén</sub>)

	S <sub>transgén</sub> /S <sub>transgén</sub>	S <sub>bac</sub> /S <sub>transgén</sub>	S <sub>cap</sub> /S <sub>transgén</sub>	S <sub>rep</sub> /S <sub>transgén</sub>
Lote#1	1	2,077E-03	1,048E-04	8,073E-04
Lote#2	1	2,462E-03	8,068E-05	1,148E-03
Lote#3	1	2,057E-03	7,207E-05	9,918E-04
Lote#4	1	2,232E-03	1,727E-04	1,173E-03
Lote#5	1	1,996E-03	8,467E-05	1,031E-03

Tabla 4. Porcentaje de varias impurezas presentes en varios lotes de AAV en relación con el casete de Ipl (basado en la fórmula descrita en el texto).

	transgén	rep	cap	Estructura de Baculovirus
Longitud molecular (nt)	3645	2785	3088	133894
Factor de corrección de longitud molecular (C)	1	0,76406	0,847188	36,73361
% de transgén presente en el lote 1	N/A	0,061682 %	0,00888 %	7,630121 %
% de transgén presente en el lote 2	N/A	0,087726 %	0,006835 %	9,04252 %
% de transgén presente en el lote 3	N/A	0,075782 %	0,006106 %	7,555691 %
% de transgén presente en el lote 4	N/A	0,089591 %	0,01463 %	8,200541 %
% de transgén presente en el lote 5	N/A	0,078764 %	0,007173 %	7,333003 %

#### Ejemplo 4: Distribución no aleatoria de las impurezas de ADN

##### 4.1 Material y métodos

5 [0132] La etapa siguiente fue la determinación del origen exacto de las impurezas de ADN obtenidas de Baculovirus. Con este objetivo, se secuenciaron diferentes lotes de vectores de rAAV mediante secuenciación profunda en la plataforma Illumina tal y como se ha descrito previamente. El alineamiento de las lecturas con el genoma de baculovirus proporcionó un medio para examinar la frecuencia de cada nucleótido (derivado del baculovirus) en la biblioteca de secuenciación profunda. Además, la frecuencia media se calculó dividiendo el número total de lecturas que se mapeaban al genoma de baculovirus con el número de nucleótidos.

##### 4.2 Resultados

15 [0133] La figura 5 representa el alineamiento de las lecturas con el genoma de Baculovirus después de la secuenciación profunda. Si las impurezas de ADN se derivaran de forma aleatoria del genoma de Baculovirus, debería observarse una distribución relativamente homogénea con aproximadamente 1200 lecturas por nucleótido. Una distribución homogénea se observa de hecho en medio del genoma de baculovirus. Sin embargo, al principio y al final del genoma de Baculovirus, se observa un fuerte aumento en el número de lecturas. Esto indica que estas regiones están sobrerrepresentadas como impurezas de ADN en el rAAV.

#### Ejemplo 5: evaluación del control de calidad usando Q-PCR o secuenciación profunda

##### 5.1 Material y métodos

25 [0134] Para investigar las capacidades cuantitativas de diferentes metodologías de NGS, es decir Solexa y 454 Roche, la distribución de lectura NGS obtenida se comparó con las mediciones de varios objetivos situados a través del genoma de baculovirus con qPCR (figura 6). Los objetivos de QPCR representan regiones: altamente sobrerrepresentada (180/181), que se corresponde con la distribución media (426/427, 428/429, 1018/1019, 30 1020/1021, 1024/1025) e infrarrepresentada (340/341). La última región se utilizó como un calibrador para todas las demás mediciones.

##### 5.2 Resultados

35 [0135] Se investigaron tres técnicas diferentes para evaluar el nivel de impurezas de ADN en los vectores de rAAV. Tal y como se muestra en la figura 6, las tres técnicas se correlacionaron bien entre sí y pueden, por lo tanto, utilizarse conjuntamente para la detección de impurezas de ADN en los vectores de rAAV.

40 [0136] Tal y como se indica en la figura 6, el análisis NGS demuestra claramente que la elección aleatoria de amplicón de ADN para una PCR cuantitativa puede provocar una medición imprecisa de una impureza de ADN particular en la preparación de vector. La presencia de impureza de ADN determinada se calcula basándose en la medición de amplicón bajo la suposición de que todas las partes de la molécula de ADN investigada (que a veces tienen 13 6000 pares de bases de longitud, por ejemplo la estructura de baculovirus) están distribuidas con la misma frecuencia. El análisis aquí presentado indica claramente que varios segmentos de moléculas de ADN largas, por ejemplo, genoma de baculovirus, pueden contaminar la preparación de vector con diferentes 45 frecuencias debido al empaquetamiento desigual de diferentes secuencias de ADN.

ES 2 751 919 T3

Tabla 5. Cebadores de Q-PCR usados en los experimentos

SEQ ID N.º	Nombre	Secuencia	Objetivo	dirección	amplificación
7	pr59	AATGGGCGGTAGGC GTGTA	Promotor de CMV	hacia adelante	5214-5284
8	pr60	AGGCGATCTGACGG TTCATAA	Promotor de CMV	inversa	5214-5284
9	pr180	CGAACCGATGGCTG GACTATC	Orf 1629 (sistema baculovirus de protein sciences)	hacia adelante	8760-8830
10	pr181	TGCTGCTACAAGAT TTGGCAAGT	Orf 1629 (sistema baculovirus de protein sciences)	inversa	8760-8830
11	pr340	ATACAACCGTTGGT TGCACG	región hr3 baculovirus corriente abajo	hacia adelante	73555-73604
12	pr341	CGGGACACGCCATG TATT	región hr3 baculovirus corriente abajo	inversa	73555-73604
13	pr402	GGGAGTGGCGGCGT TGATTT	ADN de Baculovirus 10 kb izquierda	sentido	135323-135403
14	pr403	GCACAGTTCAAGCC TCACAGCCTA	ADN de Baculovirus 10kb izquierda	antisentido	135323-135403
15	pr404	CAAACGTGGTTTCG TGTGCCAA	ADN de Baculovirus izquierda ORF603	sentido	3501-3603
16	pr405	GATGCATGACTTCA CCCACACACTT	ADN de Baculovirus izquierda ORF603	antisentido	3501-3603
17	pr406	ACAGCCATTGTAAT GAGACGCACAA	ADN de Baculovirus derecha ORF603	sentido	4357-4438
18	pr407	CCTAGCGCCCGATC AGCAACTATAT	ADN de Baculovirus derecha ORF603	antisentido	4357-4438
19	pr408	TACCGACTCTGCTG AAGAGGAGGAA	ADN de Baculovirus izquierda ORF1629	sentido	8421-8499
20	pr409	TGCGTCTGGTGCAA ACTCCTTA	ADN de Baculovirus izquierda ORF1629	antisentido	8421-8499
21	pr410	GATTCGTCATGGCC ACCACAAA	ADN de Baculovirus derecha Orf 1629	sentido	10178-10261
22	pr411	CCAAAGCGCCCGTT GATTATTTT	ADN de Baculovirus derecha Orf 1629	antisentido	10178-10261
23	pr412	GCGTACTTGCGGCT GTCGTTGTA	ADN de Baculovirus 10kb derecha	sentido	14503-14605
24	pr413	CGAGGTCAAGTTCA AAGGGCAACAT	ADN de Baculovirus 10kb derecha	antisentido	14503-14605
25	Pr426	GCATTTCGCAGCTC TCCTTCAATT	genoma de Bac	sentido	32837-32935

ES 2 751 919 T3

SEQ ID N.º	Nombre	Secuencia	Objetivo	dirección	amplificón
26	Pr427	CTTCAAGCGAGAAC GCAGCAATT	genoma de Bac	antisentido	32837-32935
27	Pr428	GTGGCGTTTGCCGT GGAAAA	genoma de Bac	sentido	116615-116699
28	Pr429	TGCAGCTGTGCGTT TTGAATGAA	genoma de Bac	antisentido	116615-116699
29	Pr1018	TTGTTATGTCAATTT GTAGCGC	genoma de Bac	sentido	18230-18296
30	Pr1019	TGCATAAAGACACA GTACAACG	genoma de Bac	antisentido	18230-18296
31	Pr1020	GACATAGTTCGTTT GAAAATTATCCC	genoma de Bac	sentido	25963-26024
32	Pr1021	AACGATCAAGCTGT TAATAAACG	genoma de Bac	antisentido	25963-26024
33	Pr1024	CGCTTCGGCGTAGT TTACC	genoma de Bac	sentido	109100-109151
34	Pr1025	CGCTATAAGCGCGG GTTAC	genoma de Bac	antisentido	109100-109151

LISTADO DE SECUENCIAS

[0137]

5

<110> uniQure IP B.V.

<120> impurezas de ADN en una composición que comprende un virión parvoviral

10

<130> p6042305pct

<160> 34

<170> versión de PatentIn 3.3

15

<210> 1

<211> 145

<212> ADN

<213> virus adenoasociado 2

20

<220>

<221> misc\_feature

<223> ITR corriente arriba

25

<400> 1

ttggccactc cctctctgcg cgctcgctcg ctactgagg cgggcgacc aaaggtcgcc 60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120

gccaaactcca tctactagggg ttcct 145

<210> 2

30

<211> 183

<212> ADN



ES 2 751 919 T3

<213> virus adenoasociado 2

<220>

<221> misc\_feature

5 <223> ITR corriente abajo

<400> 2

```

gcgcggtacc ccatggagga acccctagtg atggagttgg ccaactccctc tctgcgcgct      60
cgctcgctca ctgaggccgc ccgggcaaaag cccgggcgctc gggcgacctt tggtcgcccg      120
gcctcagtga gcgagcgcgc ggcagagag ggagtgccca agatatccca tggggtaccg      180
cgc                                                                              183

```

10

<210> 3

<211> 2203

<212> ADN

<213> virus adenoasociado 5

15

<220>

<221> misc\_feature

<223> pDP5

20

<400> 3

```

atgtcttttg ttgatcacc tccagattgg ttggaagaag ttggtgaagg tcttcgagag      60
tttttgggcc ttgaagcggg cccaccgaaa ccaaaaccca atcagcagca tcaagatcaa      120
gcccgtggtc ttgtgctgcc tggttataac tatctcggac ccggaaacgg tctcgatoga      180
ggagagcctg tcaacagggc agacgaggtc gcgcgagagc acgacatctc gtacaacgag      240
cagcttgagg cgggagacaa cccctacctc aagtacaacc acgcccagcg cgagtttcag      300
gagaagctcg ccgacgacac atccttcggg gaaacctcg gaaaggcagt ctttcaggcc      360
aagaaaaggg ttctcgaacc ttttggcctg gttgaagagg gtgctaagac ggcccctacc      420
ggaaagcggg tagacgacca ctttcacaaa agaaagaagg ctccggaccga agaggactcc      480
aagccttcca cctcgtcaga cgcggaagct ggaccagcg gatcccagca gctgcaaate      540
ccagcccaac cagcctcaag tttgggagct gatacaatgt ctgcgggagg tggcggccca      600
ttgggcgaca ataaccaagg tgccgatgga gtgggcaatg cctcgggaga ttggcattgc      660
gattccacgt ggatggggga cagagtcgtc accaagtcca cccgaacctg ggtgctgccc      720
agctacaaca accaccagta ccgagagatc aaaagcggct ccgtcgacgg aagcaacgcc      780
aacgcctact ttggatacag caccocctgg gggactttg actttaaccg ctccacagc      840
cactggagcc ccgagactg gcaaagactc atcaacaact actggggctt cagaccccg      900
tccctcagag tcaaaatctt caacattcaa gtcaaagag tcaacggtgca ggactccacc      960
accaccatcg ccaacaacct cacctccacc gtccaagtgt ttacggacga cgactaccag     1020
ctgccctacg togtcgcaa cgggaccgag ggatgcctgc cggccttccc tccgcaggtc     1080
tttacgctgc cgcagtagcg ttacgcgacg ctgaaccgcg acaacacaga aaatcccacc     1140

```

ES 2 751 919 T3

```

gagaggagca gcttcttctg cctagagtac tttcccagca agatgctgag aacgggcaac 1200
aactttgagt ttacctacaa ctttgaggag gtgcccttcc actccagett cgctcccagt 1260
cagaacctgt tcaagctggc caaccgctg gtggaccagt acttgtagc cttogtgagc 1320
acaaataaca ctggcggagt ccagttcaac aagaacctgg cggggagata cgccaacacc 1380
tacaanaact ggttcccggg gcccatgggc cgaaccagg gctggaacct gggctccggg 1440
gtcaaccgcg ccagtgtcag cgccttcgcc acgaccaata ggatggagct cgagggcgcg 1500
agttaccagg tgccccgca gccgaacggc atgaccaaca acctccaggg cagcaacacc 1560
tatgccctgg agaactat gatcttcaac agccagccgg cgaaccggg caccaccgcc 1620
acgtacctcg agggcaacat gctcatcacc agcgagagcg agacgcagcc ggtgaaccgc 1680
gtggcgtaca acgtcggcgg gcagatggcc accaacaacc agagctccac cactgcccc 1740
gcgaccggca cgtacaacct ccaggaaatc gtgcccggca gcgtgtggat ggagagggac 1800
gtgtacctcc aaggacctat ctgggccaag atcccagaga cgggggcgca ctttcacccc 1860
tctccggcca tgggcggatt cggactcaa caccaccgc ccatgatgct catcaagaac 1920
acgcctgtgc ccgaaatat caccagcttc tcggacgtgc ccgtcagcag cttcatcacc 1980
cagtacagca ccgggcaggt caccgtggag atggagtggg agctcaagaa ggaaaactcc 2040
aagaggtgga acccagagat ccagtacaca acaactaca acgaccccc gtttgtggac 2100
tttccccgg acagaccgg ggaatacaga accaccagac ctatcggaac ccgatacctt 2160
acccgacccc ttaacccat tcatgtcgca taccctcaat aaa 2203

```

5 <210> 4  
 <211> 1876  
 <212> ADN  
 <213> virus adenoasociado 2

10 <220>  
 <221> CDS  
 <222> (11) .. (1876)

<400> 4

ES 2 751 919 T3

cgccagccgcc	atg	ccg	ggg	ttt	tac	gag	att	gtg	att	aag	gtc	ccc	agc		49	
	Met	Pro	Gly	Phe	Tyr	Glu	Ile	Val	Ile	Lys	Val	Pro	Ser			
	1				5					10						
gac	ctt	gac	gag	cat	ctg	ccc	ggc	att	tct	gac	agc	ttt	gtg	aac	tgg	97
Asp	Leu	Asp	Glu	His	Leu	Pro	Gly	Ile	Ser	Asp	Ser	Phe	Val	Asn	Trp	
	15				20					25						
gtg	gcc	gag	aag	gaa	tgg	gag	ttg	ccg	cca	gat	tct	gac	atg	gat	ctg	145
Val	Ala	Glu	Lys	Glu	Trp	Glu	Leu	Pro	Pro	Asp	Ser	Asp	Met	Asp	Leu	
	30				35				40						45	
aat	ctg	att	gag	cag	gca	ccc	ctg	acc	gtg	gcc	gag	aag	ctg	cag	cgc	193
Asn	Leu	Ile	Glu	Gln	Ala	Pro	Leu	Thr	Val	Ala	Glu	Lys	Leu	Gln	Arg	
				50					55					60		
gac	ttt	ctg	acg	gaa	tgg	cgc	cgt	gtg	agt	aag	gcc	ccg	gag	gcc	ctt	241
Asp	Phe	Leu	Thr	Glu	Trp	Arg	Arg	Val	Ser	Lys	Ala	Pro	Glu	Ala	Leu	
			65				70						75			
ttc	ttt	gtg	caa	ttt	gag	aag	gga	gag	agc	tac	ttc	cac	atg	cac	gtg	289
Phe	Phe	Val	Gln	Phe	Glu	Lys	Gly	Glu	Ser	Tyr	Phe	His	Met	His	Val	
		80					85					90				
ctc	gtg	gaa	acc	acc	ggg	gtg	aaa	tcc	atg	gtt	ttg	gga	cgt	ttc	ctg	337
Leu	Val	Glu	Thr	Thr	Gly	Val	Lys	Ser	Met	Val	Leu	Gly	Arg	Phe	Leu	
	95				100						105					
agt	cag	att	cgc	gaa	aaa	ctg	att	cag	aga	att	tac	cgc	ggg	atc	gag	385
Ser	Gln	Ile	Arg	Glu	Lys	Leu	Ile	Gln	Arg	Ile	Tyr	Arg	Gly	Ile	Glu	
	110				115					120					125	
ccg	act	ttg	cca	aac	tgg	ttc	gcg	gtc	aca	aag	acc	aga	aat	ggc	gcc	433
Pro	Thr	Leu	Pro	Asn	Trp	Phe	Ala	Val	Thr	Lys	Thr	Arg	Asn	Gly	Ala	
				130					135					140		
gga	ggc	ggg	aac	aag	gtg	gtg	gat	gag	tgc	tac	atc	ccc	aat	tac	ttg	481
Gly	Gly	Gly	Asn	Lys	Val	Val	Asp	Glu	Cys	Tyr	Ile	Pro	Asn	Tyr	Leu	
			145				150						155			
ctc	ccc	aaa	acc	cag	cct	gag	ctc	cag	tgg	gcg	tgg	act	aat	atg	gaa	529
Leu	Pro	Lys	Thr	Gln	Pro	Glu	Leu	Gln	Trp	Ala	Trp	Thr	Asn	Met	Glu	

ES 2 751 919 T3

160	165	170	
cag tat tta agc gcc tgt ttg aat ctc acg gag cgt aaa cgg ttg gtg			577
Gln Tyr Leu Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val			
175	180	185	
gcg cag cat ctg acg cac gtg tcg cag acg cag gag cag aac aaa gag			625
Ala Gln His Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu			
190	195	200	205
aat cag aat ccc aat tct gat gcg ccg gtg atc aga tca aaa act tca			673
Asn Gln Asn Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser			
210	215	220	
gcc agg tac atg gag ctg gtc ggg tgg ctc gtg gac aag ggg att acc			721
Ala Arg Tyr Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr			
225	230	235	
tcg gag aag cag tgg atc cag gag gac cag gcc tca tac atc tcc ttc			769
Ser Glu Lys Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe			
240	245	250	
aat gcg gcc tcc aac tcg cgg tcc caa atc aag gct gcc ttg gac aat			817
Asn Ala Ala Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn			
255	260	265	
gcg gga aag att atg agc ctg act aaa acc gcc ccc gac tac ctg gtg			865
Ala Gly Lys Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val			
270	275	280	285
ggc cag cag ccc gtg gag gac att tcc agc aat cgg att tat aaa att			913
Gly Gln Gln Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile			
290	295	300	
ttg gaa cta aac ggg tac gat ccc caa tat gcg gct tcc gtc ttt ctg			961
Leu Glu Leu Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu			
305	310	315	
gga tgg gcc acg aaa aag ttc ggc aag agg aac acc atc tgg ctg ttt			1009
Gly Trp Ala Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe			
320	325	330	
ggg cct gca act acc ggg aag acc aac atc gcg gag gcc ata gcc cac			1057
Gly Pro Ala Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His			
335	340	345	
act gtg ccc ttc tac ggg tgc gta aac tgg acc aat gag aac ttt ccc			1105
Thr Val Pro Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro			
350	355	360	365
ttc aac gac tgt gtc gac aag atg gtg atc tgg tgg gag gag ggg aag			1153
Phe Asn Asp Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys			
370	375	380	
atg acc gcc aag gtc gtg gag tcg gcc aaa gcc att ctc gga gga agc			1201
Met Thr Ala Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser			
385	390	395	
aag gtg cgc gtg gac cag aaa tgc aag tcc tcg gcc cag ata gac ccg			1249
Lys Val Arg Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro			
400	405	410	
act ccc gtg atc gtc acc tcc aac acc aac atg tgc gcc gtg att gac			1297

ES 2 751 919 T3

Thr	Pro	Val	Ile	Val	Thr	Ser	Asn	Thr	Asn	Met	Cys	Ala	Val	Ile	Asp		
415						420					425						
ggg	aac	tca	acg	acc	ttc	gaa	cac	cag	cag	ccg	ttg	caa	gac	cgg	atg	1345	
Gly	Asn	Ser	Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met		
430					435					440					445		
ttc	aaa	ttt	gaa	ctc	acc	cgc	cgt	ctg	gat	cat	gac	ttt	ggg	aag	gtc	1393	
Phe	Lys	Phe	Glu	Leu	Thr	Arg	Arg	Leu	Asp	His	Asp	Phe	Gly	Lys	Val		
				450					455						460		
acc	aag	cag	gaa	gtc	aaa	gac	ttt	ttc	cgg	tgg	gca	aag	gat	cac	gtg	1441	
Thr	Lys	Gln	Glu	Val	Lys	Asp	Phe	Phe	Arg	Trp	Ala	Lys	Asp	His	Val		
			465					470							475		
gtt	gag	gtg	gag	cat	gaa	ttc	tac	gtc	aaa	aag	ggt	gga	gcc	aag	aaa	1489	
Val	Glu	Val	Glu	His	Glu	Phe	Tyr	Val	Lys	Lys	Gly	Gly	Ala	Lys	Lys		
		480					485					490					
aga	ccc	gcc	ccc	agt	gac	gca	gat	ata	agt	gag	ccc	aaa	cgg	gtg	cgc	1537	
Arg	Pro	Ala	Pro	Ser	Asp	Ala	Asp	Ile	Ser	Glu	Pro	Lys	Arg	Val	Arg		
	495					500						505					
gag	tca	gtt	gcg	cag	cca	tcg	acg	tca	gac	gcg	gaa	gct	tcg	atc	aac	1585	
Glu	Ser	Val	Ala	Gln	Pro	Ser	Thr	Ser	Asp	Ala	Glu	Ala	Ser	Ile	Asn		
510					515						520				525		
tac	gca	gac	agg	tac	caa	aac	aaa	tgt	tct	cgt	cac	gtg	ggc	atg	aat	1633	
Tyr	Ala	Asp	Arg	Tyr	Gln	Asn	Lys	Cys	Ser	Arg	His	Val	Gly	Met	Asn		
				530						535					540		
ctg	atg	ctg	ttt	ccc	tgc	aga	caa	tgc	gag	aga	atg	aat	cag	aat	tca	1681	
Leu	Met	Leu	Phe	Pro	Cys	Arg	Gln	Cys	Glu	Arg	Met	Asn	Gln	Asn	Ser		
			545					550						555			
aat	atc	tgc	ttc	act	cac	gga	cag	aaa	gac	tgt	tta	gag	tgc	ttt	ccc	1729	
Asn	Ile	Cys	Phe	Thr	His	Gly	Gln	Lys	Asp	Cys	Leu	Glu	Cys	Phe	Pro		
		560						565					570				
gtg	tca	gaa	tct	caa	ccc	gtt	tct	gtc	gtc	aaa	aag	gcg	tat	cag	aaa	1777	
Val	Ser	Glu	Ser	Gln	Pro	Val	Ser	Val	Val	Lys	Lys	Ala	Tyr	Gln	Lys		
		575				580						585					
ctg	tgc	tac	att	cat	cat	atc	atg	gga	aag	gtg	cca	gac	gct	tgc	act	1825	
Leu	Cys	Tyr	Ile	His	His	Ile	Met	Gly	Lys	Val	Pro	Asp	Ala	Cys	Thr		
590					595					600					605		
gcc	tgc	gat	ctg	gtc	aat	gtg	gat	ttg	gat	gac	tgc	atc	ttt	gaa	caa	1873	
Ala	Cys	Asp	Leu	Val	Asn	Val	Asp	Leu	Asp	Asp	Cys	Ile	Phe	Glu	Gln		
				610					615					620			
taa																1876	

<210> 5

<211> 621

5 <212> PRT

<213> virus adenoasociado 2

<400> 5

ES 2 751 919 T3

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp  
1                   5                   10                   15

ES 2 751 919 T3

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu  
 20 25 30  
 Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile  
 35 40 45  
 Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu  
 50 55 60  
 Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val  
 65 70 75 80  
 Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu  
 85 90 95  
 Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile  
 100 105 110  
 Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu  
 115 120 125  
 Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly  
 130 135 140  
 Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys  
 145 150 155 160  
 Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu  
 165 170 175  
 Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His  
 180 185 190  
 Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn  
 195 200 205  
 Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr  
 210 215 220  
 Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys  
 225 230 235 240  
 Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala  
 245 250 255  
 Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys





ES 2 751 919 T3

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp  
 515 520 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu  
 530 535 540

Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys  
 545 550 555 560

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu  
 565 570 575

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr  
 580 585 590

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp  
 595 600 605

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln  
 610 615 620

- 5 <210> 6
- <211> 1194
- <212> ADN
- <213> Artificial
  
- 10 <220>
- <223> Rep52 AAV2 optimizada para AcMNPV
  
- <220>
- <221> misc\_feature
- <223> sequencia codificante de Rep52 optimizada para AcMNPV
  
- 15 <400> 6

ES 2 751 919 T3

atggaattgg tccggttggtt ggtggacaag ggtattacct cggagaagca atggatacaa 60  
 gaagatcaag cctcatacat ctcgtttaat gcggcatcca actcgcgtag ccaaatcaaa 120  
 gctgccttgg acaatgcggg caagattatg agcctgacta aaaccgcccc cgactacctg 180  
 gtgggccagc aaccctgga agacatttcc agcaatcgca tctataagat tttggagtta 240  
 aacggctacg atcctcaata tgcggcttcc gtatttttgg gctgggcgac gaaaaagttt 300  
 ggcaaaagaa acaccatttg gttgtttga cctgcaacta cgggaaaaac aaacatagcg 360  
 gaggccatag cccacactgt acctttttat ggctgcgtta actggaccaa tgagaacttt 420  
 ccattcaacg actgtgtcga caagatggtt atttgggtgg aggaaggcaa aatgaccgct 480  
 aaagtcgtgg agtcggccaa agcaatttta ggaggcagca aagtgcgctg agaccagaaa 540  
 tgcaaaagct ctgctcagat agaccgcgaca ccggtgatcg ttacaagcaa cacgaacatg 600  
 tgcgccgtga ttgacggtaa cagtacgaca ttcgaacacc aacaaccggt gcaagaccga 660  
  
 atgttcaaat ttgaattgac gcgccgactg gatcatgatt ttggcaaggt aacaaaacaa 720  
 gaagtcaaag acttctttcg ttgggcaaag gatcacgttg ttgaagtgga acatgaattt 780  
 tacgtcaaaa aaggtggtgc taagaaaaga cccgccccga gtgatgcaga tataagtgag 840  
 cccaaacgag tgagagaatc gtttgcgcag ccaagcacgt cagatgcgga agcttcgata 900  
 aactacgcag accgctacca aaacaaatgt tctcgtcacg taggcatgaa cttaatgttg 960  
 tttccctgca gacaatgtga gagaatgaat cagaatagta atatctgttt cactcacggc 1020  
 cagaaagact gtttagaatg ctttccggtg tcagaatctc aaccctttc tgtcgtaaaa 1080  
 aaggcgtatc aaaaattatg ctatattcat catatcatgg gaaaagtgcc agacgcttgt 1140  
 actgcctgcg atctggttaa tgtggatttg gatgactgta tctttgaaca ataa 1194

5 <210> 7  
 <211> 19  
 <212> ADN  
 <213> artificial

10 <220>  
 <223> cebador

<220>  
 <221> misc\_feature  
 <223> pr59

15 <400> 7  
 aatgggcggt aggcgtgta 19

20 <210> 8  
 <211> 22  
 <212> ADN  
 <213> artificial

<220>

ES 2 751 919 T3

<223> cebador  
 <220>  
 <221> misc\_feature  
 5 <223> pr60  
 <400> 8  
 aggcgatctg acggttcaact aa 22  
 10 <210> 9  
 <211> 21  
 <212> ADN  
 <213> artificial  
 15 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 20 <223> pr180  
 <400> 9  
 cgaaccgatg gctggactat c 21  
 25 <210> 10  
 <211> 23  
 <212> ADN  
 <213> artificial  
 30 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 35 <223> pr181  
 <400> 10  
 tgctgctaca agatttgca agt 23  
 40 <210> 11  
 <211> 20  
 <212> ADN  
 <213> artificial  
 45 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 50 <223> pr340  
 <400> 11  
 atacaaccgt tggttgcacg 20  
 55 <210> 12  
 <211> 18  
 <212> ADN  
 <213> artificial  
 60 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 65 <223> pr341

# ES 2 751 919 T3

<400> 12  
cgggacacgc catgtatt 18

5 <210> 13  
<211> 20  
<212> ADN  
<213> artificial

10 <220>  
<223> cebador

15 <220>  
<221> misc\_feature  
<223> pr402

<400> 13  
gggagtggcg gcgttgatt 20

20 <210> 14  
<211> 24  
<212> ADN  
<213> artificial

25 <220>  
<223> cebador

30 <220>  
<221> misc\_feature  
<223> pr403

<400> 14  
gcacagtca agcctcacag ccta 24

35 <210> 15  
<211> 22  
<212> ADN  
<213> artificial

40 <220>  
<223> cebador

45 <220>  
<221> misc\_feature  
<223> pr404

<400> 15  
caaacgtggt ttcgtgccc aa 22

50 <210> 16  
<211> 25  
<212> ADN  
<213> artificial

55 <220>  
<223> cebador

60 <220>  
<221> misc\_feature  
<223> pr405

<400> 16  
gatgcatgac tcacccaca cactt 25

65 <210> 17  
<211> 25  
<212> ADN

<213> artificial  
 <220>  
 <223> cebador  
 5  
 <220>  
 <221> misc\_feature  
 <223> pr406  
 10 <400> 17  
 acagccattg taatgagacg cacaa 25  
 <210> 18  
 <211> 25  
 15 <212> ADN  
 <213> artificial  
 <220>  
 <223> cebador  
 20  
 <220>  
 <221> misc\_feature  
 <223> pr407  
 25 <400> 18  
 cctagcgccc gatcagcaac tatat 25  
 <210> 19  
 <211> 25  
 30 <212> ADN  
 <213> artificial  
 <220>  
 <223> cebador  
 35  
 <220>  
 <221> misc\_feature  
 <223> pr408  
 40 <400> 19  
 taccgactct gctgaagagg aggaa 25  
 <210> 20  
 <211> 23  
 45 <212> ADN  
 <213> artificial  
 <220>  
 <223> cebador  
 50  
 <220>  
 <221> misc\_feature  
 <223> pr409  
 55 <400> 20  
 tgcgctctgt gcaaactcct tta 23  
 <210> 21  
 <211> 22  
 60 <212> ADN  
 <213> artificial  
 <220>  
 <223> cebador  
 65  
 <220>

ES 2 751 919 T3

<221> misc\_feature  
 <223> pr410  
  
 5 <400> 21  
 gattcgtcat ggccaccaca aa 22  
  
 <210> 22  
 <211> 23  
 <212> ADN  
 10 <213> artificial  
  
 <220>  
 <223> cebador  
  
 15 <220>  
 <221> misc\_feature  
 <223> pr411  
  
 20 <400> 22  
 ccaaagcgcc cggtgattat ttt 23  
  
 <210> 23  
 <211> 23  
 <212> ADN  
 25 <213> artificial  
  
 <220>  
 <223> cebador  
  
 30 <220>  
 <221> misc\_feature  
 <223> pr412  
  
 35 <400> 23  
 gcgtacttgc ggctgtcgtt gta 23  
  
 <210> 24  
 <211> 25  
 <212> ADN  
 40 <213> artificial  
  
 <220>  
 <223> cebador  
  
 45 <220>  
 <221> misc\_feature  
 <223> pr413  
  
 50 <400> 24  
 cgaggtcaag tcaaagggc aacat 25  
  
 <210> 25  
 <211> 24  
 <212> ADN  
 55 <213> artificial  
  
 <220>  
 <223> cebador  
  
 60 <220>  
 <221> misc\_feature  
 <223> pr426  
  
 65 <400> 25  
 gcatttcgca gctctccttc aatt 24

<210> 26  
 <211> 23  
 <212> ADN  
 <213> artificial  
 5  
 <220>  
 <223> cebador  
 <220>  
 10 <221> misc\_feature  
 <223> pr427  
 <400> 26  
 15 ctcaagcga gaacgcagca att 23  
 <210> 27  
 <211> 20  
 <212> ADN  
 <213> artificial  
 20  
 <220>  
 <223> cebador  
 <220>  
 25 <221> misc\_feature  
 <223> pr428  
 <400> 27  
 30 gtggcgttg ccgtggaaaa 20  
 <210> 28  
 <211> 23  
 <212> ADN  
 <213> artificial  
 35  
 <220>  
 <223> cebador  
 <220>  
 40 <221> misc\_feature  
 <223> pr429  
 <400> 28  
 45 tgcagctgtg cgtttgaat gaa 23  
 <210> 29  
 <211> 22  
 <212> ADN  
 <213> artificial  
 50  
 <220>  
 <223> cebador  
 <220>  
 55 <221> misc\_feature  
 <223> pr1018  
 <400> 29  
 60 ttgtatgtc aattttagc gc 22  
 <210> 30  
 <211> 22  
 <212> ADN  
 <213> artificial  
 65  
 <220>

ES 2 751 919 T3

<223> cebador  
 <220>  
 <221> misc\_feature  
 5 <223> pr1019  
 <400> 30  
 tgcataaaga cacagtacaa cg 22  
 10 <210> 31  
 <211> 26  
 <212> ADN  
 <213> Artificial  
 15 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 20 <223> pr1020  
 <400> 31  
 gacatagttc gttgaaaat tatccc 26  
 25 <210> 32  
 <211> 23  
 <212> ADN  
 <213> Artificial  
 30 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 35 <223> pr1021  
 <400> 32  
 aacgatcaag ctgtaataa acg 23  
 40 <210> 33  
 <211> 19  
 <212> ADN  
 <213> Artificial  
 45 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 50 <223> pr1024  
 <400> 33  
 cgcttcggcg tagtttacc 19  
 55 <210> 34  
 <211> 19  
 <212> ADN  
 <213> Artificial  
 60 <220>  
 <223> cebador  
 <220>  
 <221> misc\_feature  
 65 <223> pr1025



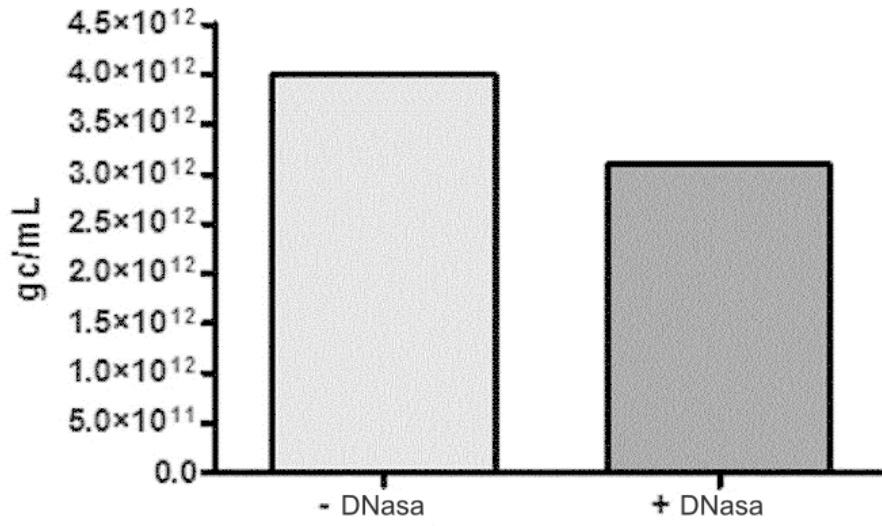
# ES 2 751 919 T3

<400> 34  
cgctataagc gcgggttac 19

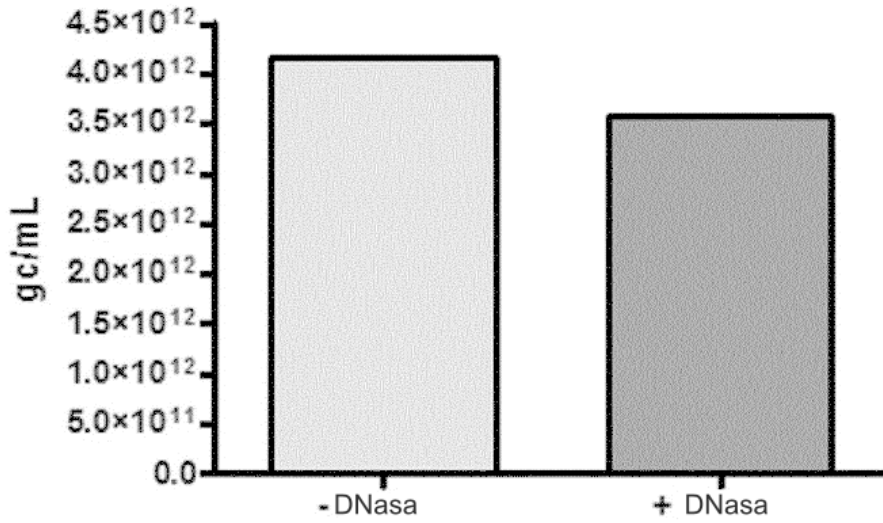
**REIVINDICACIONES**

- 5 1. Método para identificar y cuantificar una impureza de ácido nucleico sobrerrepresentada en una composición que comprende un vector parvoviral y donde el método comprende las etapas de:
- 10 a) someter la composición a secuenciación de ácido nucleico para obtener lecturas aleatorias de secuencias de nucleótidos, donde la secuenciación de ácido nucleico comprende la secuenciación de alto rendimiento;
- 15 b) comparar las lecturas aleatorias de la etapa a) con una secuencia de nucleótidos de un componente biológico usado en el proceso de producción de la composición por la cual una correspondencia entre una lectura aleatoria y una secuencia de nucleótidos de un componente biológico identifica una impureza de ácido nucleico;
- 20 c) determinar el número medio de lecturas por vector parvoviral; y
- d) determinar el número de lecturas por nucleótido de la impureza de ácido nucleico sobrerrepresentada, donde una impureza de ácido nucleico se identifica como una impureza sobrerrepresentada cuando la distribución de lecturas no es aleatoria y la impureza sobrerrepresentada comprende 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o 50 veces el número medio de lecturas del componente biológico, o cuando el número de lecturas por nucleótido de una impureza de ácido nucleico es de al menos el 0,001; 0,01; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1,0; 2,0; 3,0; 5,0; 7,0 o 10 % del número medio de lecturas por vector parvoviral.
- 25 2. Método según la reivindicación 1, donde el vector parvoviral es un vector de virus adenoasociado recombinante (rAAV).
3. Método según la reivindicación 1 o la reivindicación 2, donde la secuencia de nucleótidos de un componente biológico se selecciona a partir de un grupo que consiste en secuencias de nucleótidos de: una célula huésped, un plásmido, un vector diferente del vector parvoviral recombinante y un virus auxiliar, donde preferiblemente el vector es un vector baculoviral.
- 30 4. Método según la reivindicación 3, donde el virus auxiliar es un adenovirus recombinante y/o un virus del herpes simple recombinante.
- 35 5. Método según cualquiera de las reivindicaciones 1-4, donde la secuencia de nucleótidos del componente biológico comprende una secuencia de nucleótidos que codifica para Rep, Cap y/o un transgén, donde preferiblemente el componente biológico comprende una secuencia de nucleótidos que codifica para un transgén, donde más preferiblemente el componente biológico comprende una secuencia de nucleótidos que codifica para un transgén que está flanqueado por al menos una ITR parvoviral y donde, más preferiblemente, el componente biológico comprende una secuencia de nucleótidos que codifica para un transgén que está flanqueado por al menos una ITR parvoviral a cada lado.
- 40 6. Método según cualquiera de las reivindicaciones 1-5, donde la impureza de ácido nucleico sobrerrepresentada se cuantifica en una segunda composición o una composición adicional.
- 45 7. Método de determinación de si una composición que comprende un vector parvoviral es adecuada para el uso humano en un tratamiento clínico, donde el método comprende las etapas de:
- 50 i) cuantificar una impureza de ácido nucleico sobrerrepresentada en una composición de vector parvoviral según cualquiera de las reivindicaciones 1-6; y
- ii) determinar la composición como adecuada para uso humano en un tratamiento clínico si la impureza de ácido nucleico está al menos 10, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 10 000 o 100 000 veces menos presente que una secuencia de referencia y/o transgén tal y como ha sido determinado por la abundancia relativa de la impureza de ácido nucleico sobrerrepresentada.
8. Método según cualquiera de las reivindicaciones 1-7, donde la composición que comprende el vector parvoviral es una composición farmacéutica.

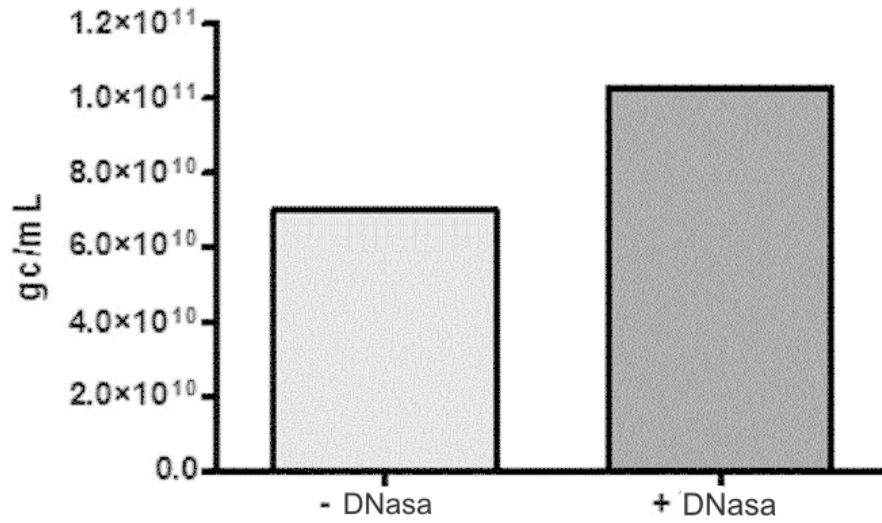
*Fig. 1A*



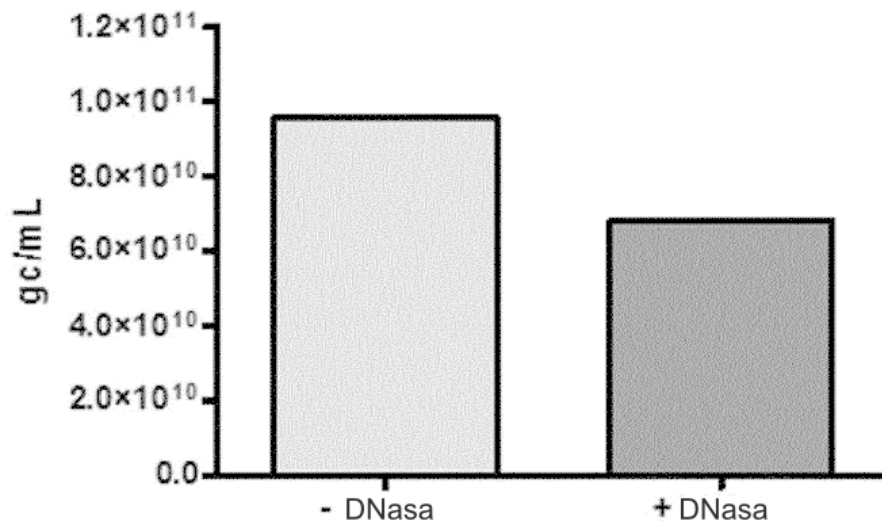
*Fig. 1B*



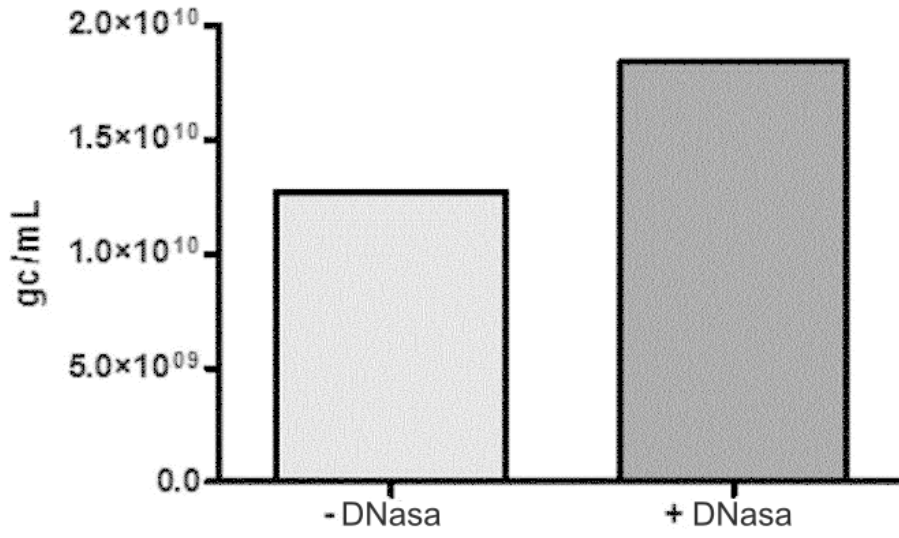
*Fig. 1C*



*Fig. 1D*



*Fig. 1E*



*Fig. 1F*

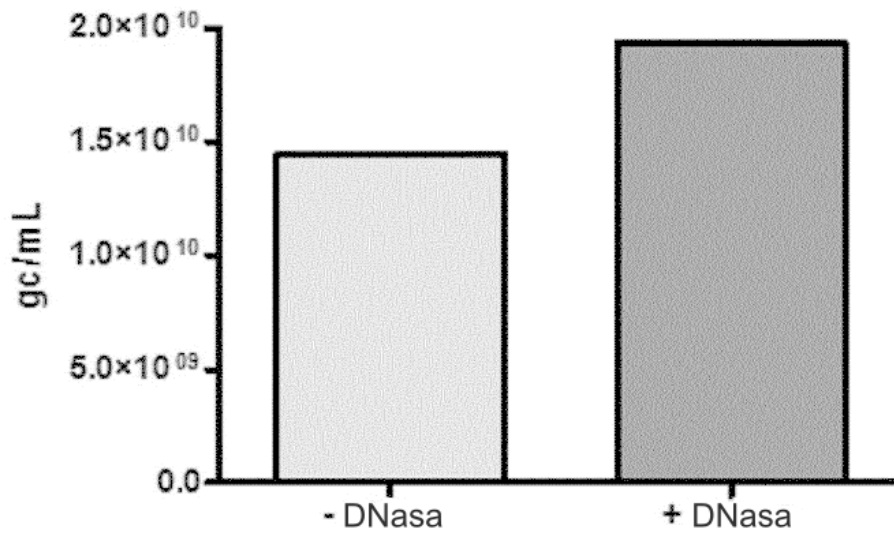
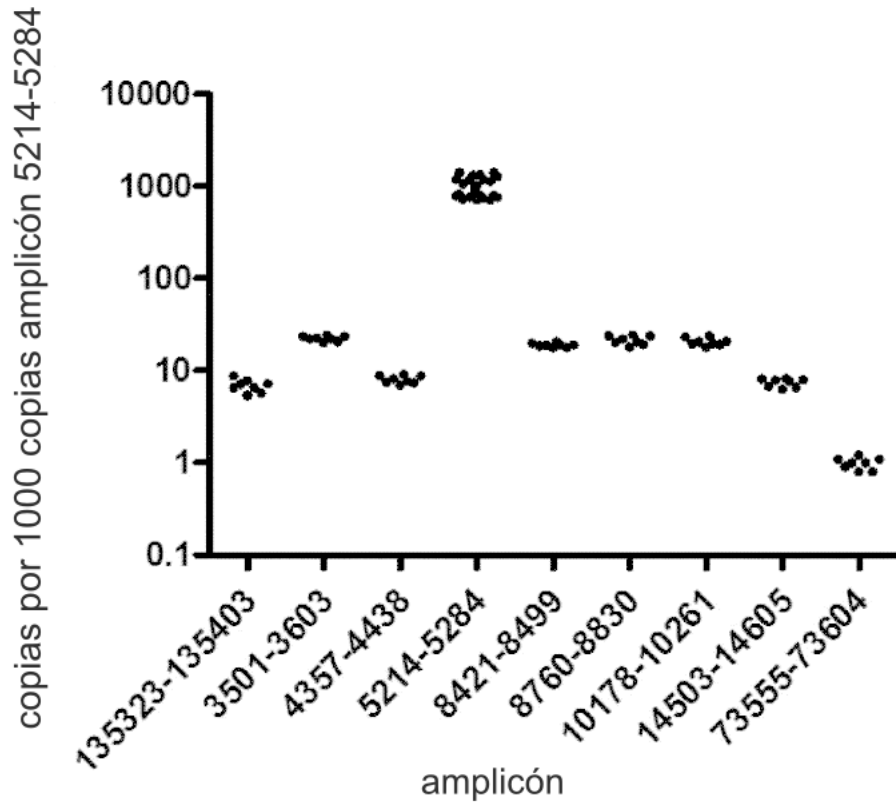




Fig. 3



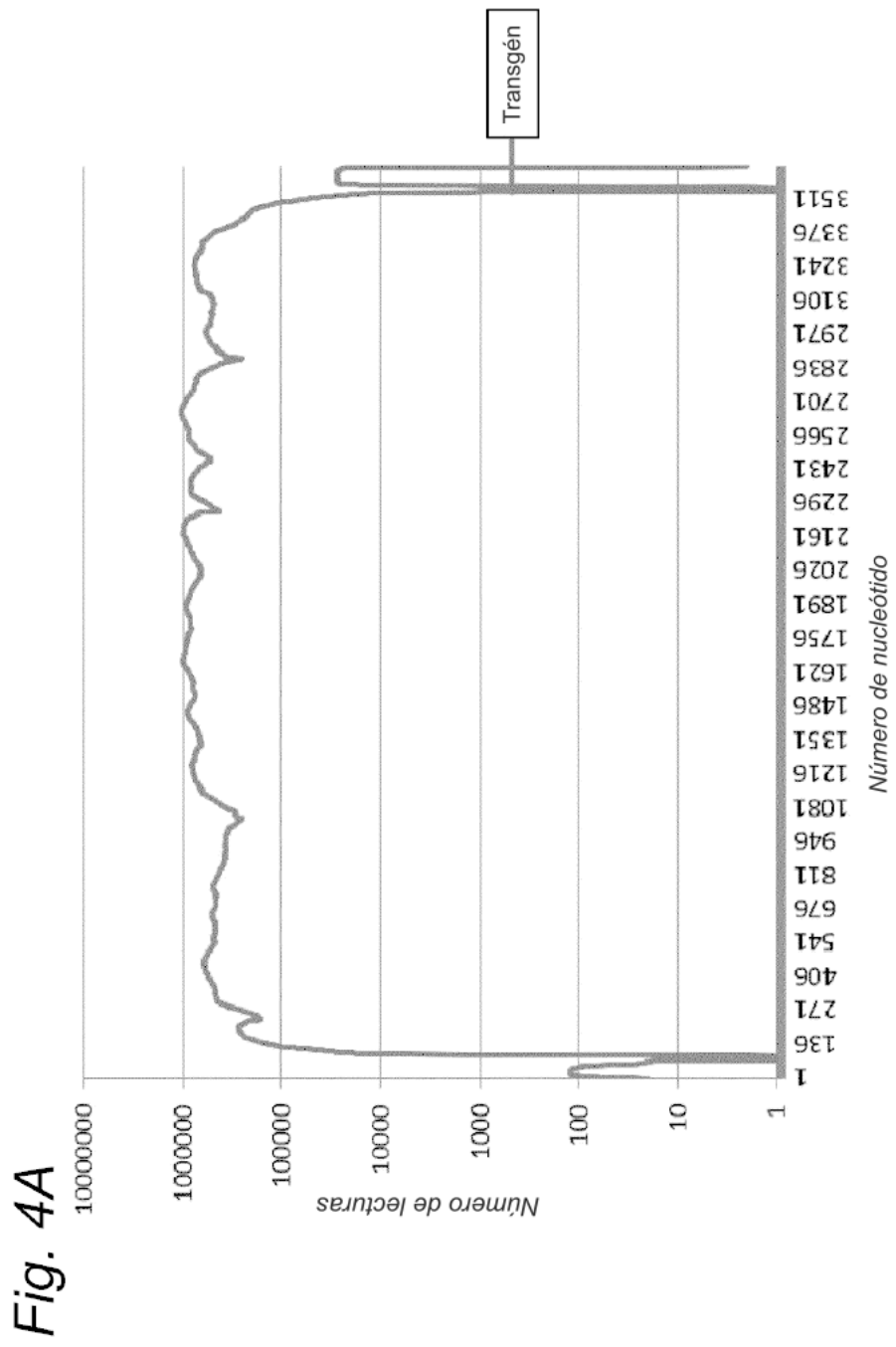
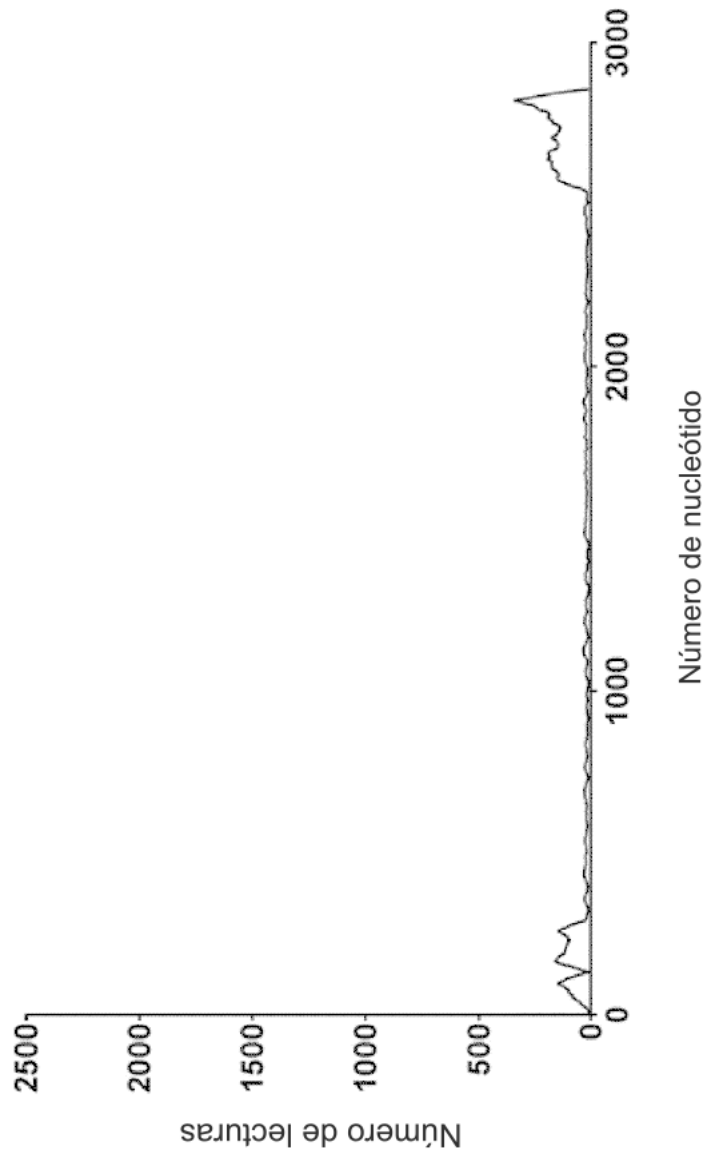
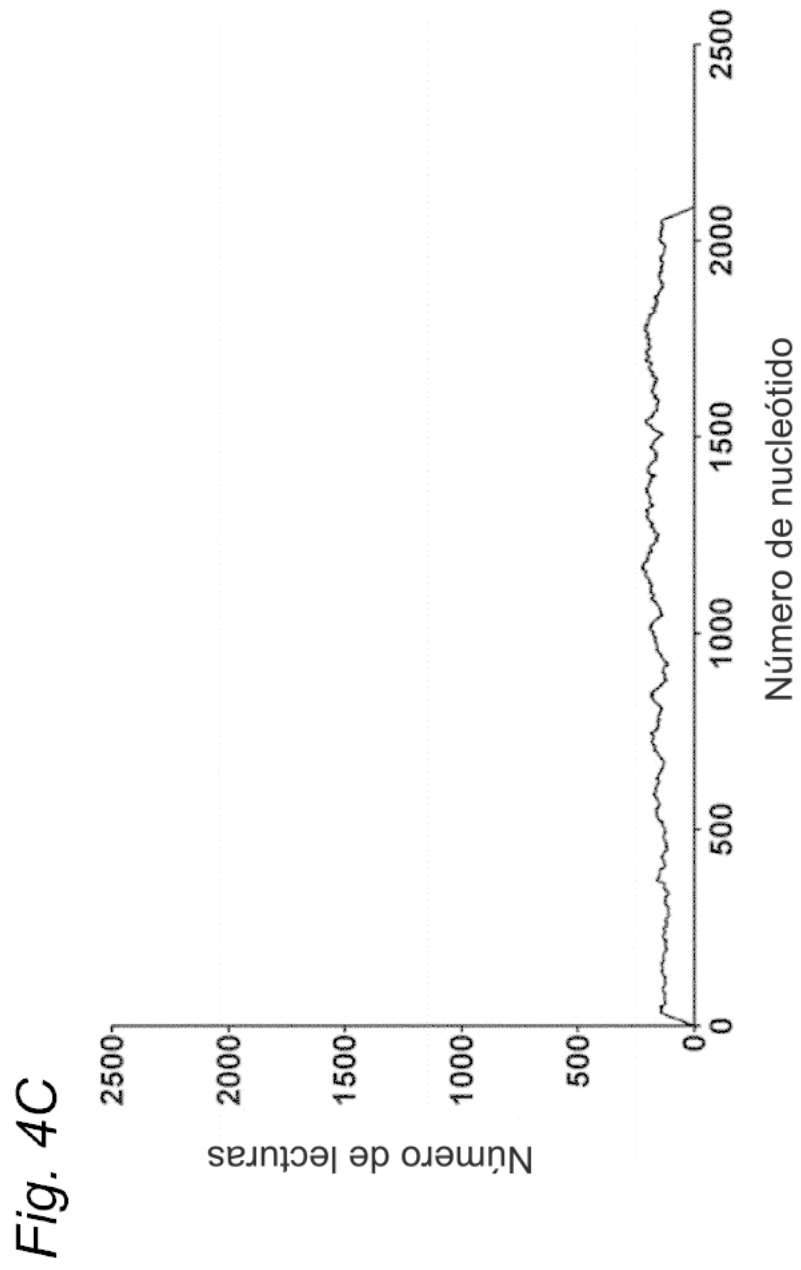




Fig. 4B





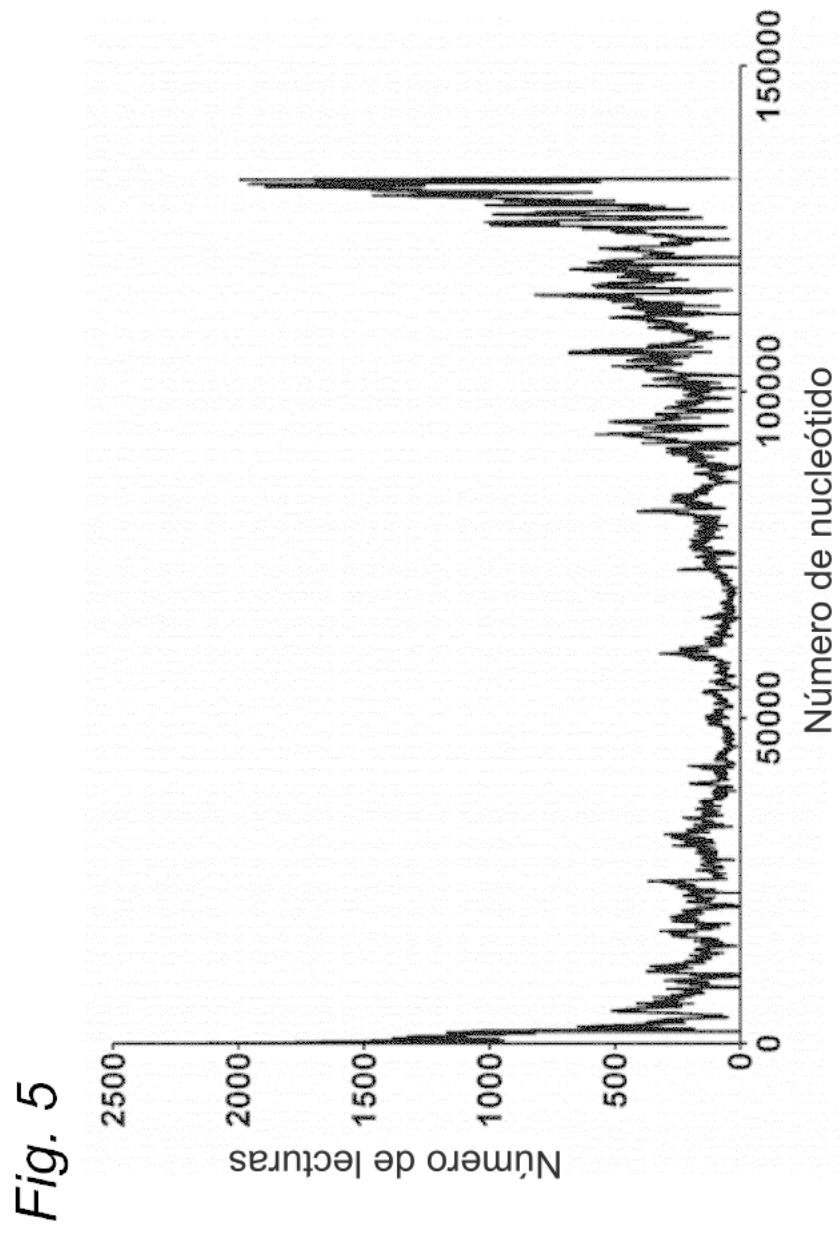


Fig. 6

