

MOLÉCULAS DE ÁCIDO NUCLEICO Y SUS USOS PARA LA TERAPIA GÉNICA NO VIRAL

SOLICITUDES RELACIONADAS

5 [0001] Esta solicitud reivindica prioridad de la Solicitud de Patente Provisional de Estados Unidos Núm. 62 /716.826, presentada el 9 de agosto de 2018, cuya descripción completa se incorpora a la presente como referencia.

REFERENCIA A LA LISTA DE SECUENCIAS PRESENTADA ELECTRÓNICAMENTE

[0002] El contenido de la lista de secuencias enviada electrónicamente en un archivo de texto ASCII (Nombre: SA9-465 PC_SL_ST25, txt; Tamaño: 466648 bytes; y Fecha de Creación: 8 de agosto de 2019) se incorpora a la presente memoria como referencia en su totalidad.

10 ANTECEDENTES DE LA INVENCION

[0003] La terapia génica ofrece el potencial de un medio duradero para tratar una variedad de enfermedades. En el pasado, muchos tratamientos de terapia génica se basaban típicamente en el uso de virus. Existen numerosos agentes virales que podrían ser seleccionados para este propósito, cada uno con propiedades distintas que los haría más o menos adecuados para la terapia génica. Zhou et al. Adv Drug
15 Deliv Rev. 106 (Pt-a): 3-26, 2016. Sin embargo, las propiedades no deseadas de algunos vectores virales, incluyendo sus perfiles inmunogénicos o su propensión a causar cáncer, han dado como resultado preocupaciones de seguridad clínica y, hasta recientemente, limitan su uso clínico a ciertas aplicaciones, por ejemplo, vacunas y estrategias oncolíticas. Cotter al. Front Biosci. 10: 1098-105 (2005).

[0004] El virus Adeno-asociado (AAV) es uno de los vectores de terapia génica más comúnmente
20 investigados. El AAV es una cubierta de proteína que rodea y protege un genoma de ADN de hebra sencilla pequeño de aproximadamente 4,8 kilobases (kb). Naso et al, BioDrugs, 31(4): 317-334, 2017. El AAV pertenece a la familia de parvovirus y depende de la co-infección con otros virus, principalmente adenovirus, con el fin de replicarse. *Id.* Su genoma de hebra sencilla contiene tres genes, Rep (Replicación), Cap (Cápside) y aap (Ensamblaje). *Id.* Estas secuencias codificantes están flanqueadas
25 por repeticiones terminales invertidas (ITR) que se requieren para replicación y empaquetamiento del genoma. *Id.* las dos ITR de AAV que actúan en cis tienen aproximadamente 145 nucleótidos de longitud con secuencias palindrómicas interrumpidas que pueden doblarse en estructuras de horquilla en forma de T que funcionan como cebadores durante el inicio de la replicación del ADN.

[0005] Sin embargo, el uso de AAV convencional como vector de suministro de genes tiene ciertos
30 inconvenientes. Uno de los principales inconvenientes está asociado con la capacidad de empaquetamiento viral limitada del AAV de aproximadamente 4,5 kb de ADN heterólogo. (Dong et al.

Hum Gene Ther. 7(17): 2101-12, 1996). Además, la administración de los vectores de AAV puede inducir una respuesta inmunitaria en seres humanos. Aunque se ha demostrado que el AAV es menos inmunogénico que algunos otros virus (es decir, adenovirus), las proteínas de la cápside pueden activar diversos componentes del sistema inmunitario humano. Véase Naso et al. 2017. AAV es un virus común en la población humana, y la mayoría de las personas se han expuesto a AAV, en consecuencia la mayoría de las personas ya han desarrollado una respuesta inmunitaria contra las variantes particulares a las que se habían expuesto previamente. Esta respuesta adaptativa pre-existente puede incluir anticuerpos neutralizantes (NAb) y células T que podrían disminuir la eficacia clínica de las re-infecciones subsiguientes con AAV y/o la eliminación de células que han sido transducidas, lo que puede descalificar a los pacientes con inmunidad anti-AAV pre-existente para el tratamiento de terapia génica basada en AAV. Además, la evidencia sugiere que los bucles de horquilla en forma de T de las ITR de AAV son susceptibles de inhibición por las proteínas/complejos de proteínas de la célula anfitriona que se unen a las estructuras de horquilla en forma de T de las ITR AAV. Véase, p. ej., Zhou et al. *Scientific Reports* 7: 5432 (14 de julio 14 de 2017).

[0006] Así, existe una necesidad en la técnica de expresar de manera eficaz y persistente las secuencias diana, p. ej., proteínas y/o miARN terapéuticos, en entornos *in vitro* e *in vivo*, evitando al mismo tiempo algunas de las consecuencias y limitaciones no pretendidas de la tecnología existente de vectores de AAV.

COMPENDIO DE LA DIVULGACIÓN

[0007] En ciertos aspectos, se proporciona una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga, en donde la primera ITR y/o la segunda ITR comprende una secuencia de nucleótidos de al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180,181,183, 184,185,186, 187 o 188, o un derivado funcional de la misma.

[0008] En ciertas realizaciones ilustrativas, la primera ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 180 y la segunda ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 181. En ciertas realizaciones ilustrativas, la primera ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 183 y la segunda ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 184. En ciertas realizaciones ilustrativas, la primera ITR comprende la secuencia de nucleótidos

expuesta en SEQ ID NO: 185 y la segunda ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 186. En ciertas realizaciones ilustrativas, la primera ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 187 y la segunda ITR comprende la secuencia de nucleótidos expuesta en SEQ ID NO: 188.

- 5 **[0009]** En ciertas realizaciones ilustrativas, la primera ITR y/o la segunda ITR consisten en una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188. En ciertas realizaciones ilustrativas, la primera ITR y la segunda ITR son complementos inversos entre sí.

- [0010]** En ciertas realizaciones ilustrativas, la molécula de ácido nucleico comprende adicionalmente un promotor. En ciertas realizaciones ilustrativas, el promotor es un promotor específico de tejido. En ciertas
10 realizaciones ilustrativas, el promotor impulsa la expresión de la secuencia de polinucleótidos heteróloga en un órgano seleccionado entre músculo, sistema nervioso central (SNC), ocular, hígado, corazón, riñón, páncreas, pulmones, piel, vejiga, tracto urinario, o cualquier combinación de los mismos. En ciertas realizaciones ilustrativas, el promotor impulsa la expresión de la secuencia de polinucleótidos heteróloga en hepatocitos, células endoteliales, células de músculo cardíaco, células de músculo esquelético, células
15 sinusoidales, neuronas aferentes, neuronas eferentes, interneuronas, células gliales, astrocitos, oligodendrocitos, microglia, células ependimarias, células epiteliales de pulmón, células de Schwann, células satélite, células fotorreceptoras, células del ganglio retiniano, o cualquier combinación de las mismas. En ciertas realizaciones ilustrativas, el promotor se coloca 5' con respecto a la secuencia de polinucleótidos heteróloga. En ciertas realizaciones ilustrativas, el promotor se selecciona del grupo que
20 consiste de un promotor de tiretina de ratón (mTTR), un promotor de factor VIII humano endógeno (F8), un promotor de alfa-1-antitripsina humana (hAAT), un promotor mínimo de albúmina humana, un promotor de albúmina de ratón, un promotor de tristetraprolina (TTP), un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV), al-antitripsina (AAT), creatina quinasa muscular (MCK), cadena pesada alfa de miosina (aMHC), mioglobina (MB), desmina (DES), SPc5-12, 2R5Sc5-12, dMCK,
25 tMCK, y un promotor de fosfoglicerato quinasa (PGK).

- [0011]** En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga comprende adicionalmente una secuencia intrónica. En ciertas realizaciones ilustrativas, la secuencia intrónica se coloca 5' con respecto a la secuencia de polinucleótidos heteróloga. En ciertas realizaciones ilustrativas, la secuencia intrónica se coloca 3' con respecto al promotor. En ciertas realizaciones ilustrativas, la
30 secuencia intrónica comprende una secuencia intrónica sintética. En ciertas realizaciones ilustrativas, la secuencia intrónica comprende SEQ ID NO: 115 o 192.

[0012] En ciertas realizaciones ilustrativas, el casete genético comprende adicionalmente un elemento regulador post-transcripcional. En ciertas realizaciones ilustrativas, el elemento regulador post-transcripcional se coloca 3' con respecto a la secuencia de polinucleótidos heteróloga. En ciertas realizaciones ilustrativas, el elemento regulador post-transcripcional comprende un elemento regulador post-transcripcional de virus de hepatitis de marmota mutado (WPRE), un sitio de unión a microARN, una secuencia de direccionamiento nuclear de ADN, o cualquier combinación de los mismos. En ciertas realizaciones ilustrativas, el sitio de unión a microARN comprende un sitio de unión a miR142-3 p.

[0013] En ciertas realizaciones ilustrativas, el casete genético comprende adicionalmente una secuencia de cola de poli (A) 3'UTR. En ciertas realizaciones ilustrativas, la secuencia de cola de poli (A) 3'UTR se selecciona del grupo que consiste de poli (A) de bGH, poli (A) de actina, poli (A) de hemoglobina, y cualquier combinación de las mismas. En ciertas realizaciones ilustrativas, la secuencia de cola de poli (A) 3'UTR comprende poli (A) de bGH.

[0014] En ciertas realizaciones ilustrativas, el casete genético comprende adicionalmente una secuencia potenciadora. En ciertas realizaciones ilustrativas, la secuencia potenciadora se coloca entre la primera ITR y la segunda ITR.

[0015] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico comprende de 5'a 3': la primera ITR, el casete genético, y la segunda ITR; en donde el casete genético comprende una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de polinucleótidos heteróloga, un elemento regulador post-transcripcional, y una secuencia de cola poli (A) 3'UTR. En ciertas realizaciones ilustrativas, el casete genético comprende de 5'a 3': una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de polinucleótidos heteróloga, un elemento regulador post-transcripcional y una secuencia de cola de poli (A) 3'UTR. En ciertas realizaciones ilustrativas, la secuencia promotora específica de tejido comprende un promotor TTT; el intrón es un intrón sintético; el elemento regulador post-transcripcional comprende WPRE; y la secuencia de cola de poli (A) 3'UTR comprende bGHpA.

[0016] En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de hebra sencilla. En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de doble hebra.

[0017] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo, o cualquier combinación de los mismos.

[0018] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un factor de crecimiento seleccionado del grupo que consiste de adrenomedulina (AM), angiopoyetina (Ang), factor de movilidad autocrina, una proteína morfogenética ósea (BMP) (p. ej. BMP2, BMP4, BMP5, BMP7), un miembro de la familia del factor neurotrófico ciliar (p. ej., factor neurotrófico ciliar (CNTF), factor inhibidor de leucemia (LIF), interleucina -6 (IL -6)), un factor estimulante de colonias (p. ej. factor estimulante de colonias de macrófagos (m-CSF), factor estimulante de colonias de granulocito (g-CSF), factor estimulante de colonias de granulocitos-macrófagos (GM-CSF)), un factor de crecimiento epidérmico (EGF), una efrina (p. ej., efrina A1, efrina A2, efrina A3, efrina-A4, efrina A5, efrina B1, efrina B2, efrina B3), eritropoyetina (EPO), un factor de crecimiento de fibroblastos (FGF) (p. ej. FGF1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF18, FGF19, FGF20, FGF21, FGF22, FGF23), somatotropina bovina fetal (FBS), un miembro de la familia de GDNF (p. ej. factor neurotrófico obtenido de líneas de células gliales (GDNF), neurturina, persefina, artemina), factor de diferenciación de crecimiento -9 (GDF9), factor de crecimiento de hepatocitos (HGF), factor de crecimiento obtenido de hepatoma (HDGF), insulina, factores de crecimiento de tipo insulínico (p. ej., factor de crecimiento de tipo insulínico-1 (IGF-1) o IGF-2, una interleucina (IL) (p. ej. IL -1, IL -2, IL -3, IL -4, IL -5, IL -6, IL -7), factor de crecimiento de queratinocitos (KGF), factor estimulante de la migración (MSF), proteína estimulante de macrófagos (MSP o proteína de tipo factor de crecimiento de hepatocitos (HGFLP)), miostatina (GDF-8), una neurregulina (p. ej., neurregulina 1 (NRG1), NRG2, NRG3, NRG4), una neurotrofina (p. ej. factor neurotrófico obtenido de cerebro (BDNF), factor de crecimiento nervioso (NGF), una neurotrofina-3 (NT-3), NT -4, factor de crecimiento de placenta (PGF), factor de crecimiento obtenido de plaquetas (PDGF), renalasa (RNLS), factor de crecimiento de células T (TCGF), trombopoyetina (TPO), un factor de crecimiento transformante (p. ej., factor de crecimiento transformante alfa (TGF- α), TGF- β , factor de necrosis tumoral-alfa (TNF- α), y factor de crecimiento endotelial vascular (VEGF), y cualquier combinación de los mismos.

[0019] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica una hormona.

[0020] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica una citocina.

[0021] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un anticuerpo o un fragmento del mismo.

[0022] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un gen seleccionado entre distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RS1, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida), y cualquier combinación de los mismos.

[0023] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un microARN (miARN). En ciertas realizaciones ilustrativas, el miARN regula la expresión de un gen diana seleccionado entre SOD1, HTT, RHO, y cualquier combinación de los mismos.

[0024] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga codifica un factor de coagulación seleccionado del grupo que consiste en factor I (FI), Factor II (FII), Factor III (FIII), Factor IV (FIV), factor V (FV), factor VI (FVI), factor VII (FVII), factor VIII (FVIII) factor IX (FIX), factor X (FX), factor XI (FXI), factor XII (FXII), factor XIII (FVIII), factor Von Willebrand (FVW), precalicreína, cininógeno de alto peso molecular, fibronectina, antitrombina III, cofactor II de heparina, proteína C, proteína S, proteína Z, inhibidor de proteasa relacionado con proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador de plasminógeno tisular (tPA), uroquinasa, inhibidor de activador de plasminógeno -1 (PAI -1), inhibidor del activador de plasminógeno-2 (PAI2), y cualquier combinación de los mismos.

[0025] En ciertas realizaciones ilustrativas, el factor de coagulación es FVIII. En ciertas realizaciones ilustrativas, el FVIII comprende FVIII maduro de longitud completa. En ciertas realizaciones ilustrativas, el FVIII comprende una secuencia de aminoácidos al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de aminoácidos que tiene SEQ ID NO: 106.

[0026] En ciertas realizaciones ilustrativas, el FVIII comprende el dominio A1, el dominio A2, el dominio A3, el dominio C1, el dominio C2, y un dominio parcial o distinto de B. En ciertas realizaciones ilustrativas, el FVIII comprende una secuencia de aminoácidos de menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de aminoácidos de SEQ ID NO: 109.

- [0027]** En ciertas realizaciones ilustrativas, el factor de coagulación comprende un radical heterólogo. En ciertas realizaciones ilustrativas, el radical heterólogo se selecciona del grupo que consiste en albúmina o un fragmento de la misma, una región Fe de inmunoglobulina, el péptido C-terminal (CTP) de la subunidad β de gonadotropina coriónica humana, una secuencia PAS, una secuencia HAP, una transferrina o un fragmento de la misma, un radical de unión a albúmina, un obtenido del mismo, o cualquier combinación de los mismos. En ciertas realizaciones ilustrativas, el radical heterólogo se conecta al extremo N-terminal o C-terminal del FVIII o se inserta entre dos aminoácidos en el FVIII. En ciertas realizaciones ilustrativas, el radical heterólogo se inserta entre dos aminoácidos en uno o más sitios de inserción seleccionados entre los sitios de inserción enumerados en la Tabla 4
- 5
- [0028]** En ciertas realizaciones ilustrativas, el FVIII comprende adicionalmente el dominio A1, el dominio A2, el dominio C1, el dominio C2, un dominio B opcional y un radical heterólogo, en donde el radical heterólogo se inserta inmediatamente aguas abajo del aminoácido 745 que corresponde al FVIII maduro (SEQ ID NO: 106).
- 10
- [0029]** En ciertas realizaciones ilustrativas, el FVIII comprende adicionalmente un compañero de unión a FcRn. En ciertas realizaciones ilustrativas, el compañero de unión a FcRn comprende una región Fc de un dominio constante de inmunoglobulina.
- 15
- [0030]** En ciertas realizaciones ilustrativas, la secuencia de ácido nucleico que codifica el FVIII tiene codones optimizados. En ciertas realizaciones ilustrativas, la secuencia de ácido nucleico que codifica el FVIII tiene codones optimizados para la expresión en un ser humano.
- 20
- [0031]** En ciertas realizaciones ilustrativas, la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% idéntica a una secuencia de nucleótidos de SEQ ID NO: 107.
- 25
- [0032]** En ciertas realizaciones ilustrativas, la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente
- 30

98%, al menos aproximadamente 99%, o aproximadamente 100% idéntica a la secuencia de nucleótidos de SEQ ID NO: 71.

[0033] En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga tiene codones optimizados. En ciertas realizaciones ilustrativas, la secuencia de polinucleótidos heteróloga tiene
5 codones optimizados para su expresión en un ser humano.

[0034] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se formula con un agente de suministro. En ciertas realizaciones ilustrativas, el agente de suministro comprende una nanopartícula lipídica. En ciertas realizaciones ilustrativas, el agente de suministro se selecciona del grupo que consiste en liposomas, moléculas poliméricas no-lipídicas y endosomas, y cualquier combinación de los mismos.

10 **[0035]** En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se formula para administración intravenosa, transdérmica, intradérmica, subcutánea, pulmonar u oral, o cualquier combinación de las mismas. En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se formula para la administración intravenosa.

15 **[0036]** En ciertos aspectos, se proporciona un vector que comprende una molécula de ácido nucleico como se describe en la presente memoria.

[0037] En ciertos aspectos, se proporciona una célula anfitriona que comprende una molécula de ácido nucleico como se describe en la presente memoria.

20 **[0038]** En ciertos aspectos, se proporciona una composición farmacéutica que comprende una molécula de ácido nucleico o un vector como se describe en la presente memoria, y un excipiente farmacéuticamente aceptable.

[0039] En ciertos aspectos, se proporciona una composición farmacéutica que comprende una célula anfitriona como se describe en la presente memoria, y un excipiente farmacéuticamente aceptable.

25 **[0040]** En ciertos aspectos, se proporciona un kit, que comprende una molécula de ácido nucleico como se describe en la presente memoria, e instrucciones para administrar la molécula de ácido nucleico a un sujeto que lo necesite.

[0041] En ciertos aspectos, se proporciona un sistema de baculovirus para la producción de una molécula de ácido nucleico como se describe en la presente memoria.

[0042] En ciertas realizaciones ilustrativas, una molécula de ácido nucleico como se describe en la presente memoria, se produce en células de insecto.

[0043] En ciertos aspectos, se proporciona un sistema de suministro de nanopartículas que comprende una molécula de ácido nucleico como se describe en la presente memoria.

[0044] En ciertos aspectos, se proporciona un método para producir un polipéptido, que comprende cultivar una célula anfitriona como se describe en la presente memoria en condiciones adecuadas y recuperar el polipéptido.

[0045] En ciertos aspectos, se proporciona un método para producir un polipéptido con actividad de coagulación, que comprende: cultivar una célula anfitriona como se describe en la presente memoria en condiciones adecuadas y recuperar el polipéptido con actividad de coagulación.

[0046] En ciertos aspectos, se proporciona un método para expresar una secuencia de polinucleótidos heteróloga en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico como se describe en la presente memoria, un vector como se describe en la presente memoria, o una composición farmacéutica como se describe en la presente memoria.

[0047] En ciertos aspectos, se proporciona un método para expresar un factor de coagulación en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico como se describe en la presente memoria, un vector como se describe en la presente memoria, un polipéptido como se describe en la presente memoria, o una composición farmacéutica como se describe en la presente memoria.

[0048] En ciertos aspectos, se proporciona un método para tratar una enfermedad o trastorno en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico como se describe en la presente memoria, un vector como se describe en la presente memoria, o una composición farmacéutica como se describe en la presente memoria.

[0049] En ciertos aspectos, se proporciona un método para tratar a un sujeto que tiene una deficiencia de factor de coagulación, que comprende administrar al sujeto una molécula de ácido nucleico como se describe en la presente memoria, un vector como se describe en la presente memoria, un polipéptido como se describe en la presente memoria, o una composición farmacéutica como se describe en la presente memoria.

[0050] En ciertos aspectos, se proporciona un método para tratar una deficiencia de factor de coagulación en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico como se describe en la presente memoria, un vector como se describe en la presente memoria, un polipéptido como se describe en la presente memoria, o una composición farmacéutica como se describe en la presente memoria.

[0051] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se administra intravenosamente, transdérmicamente, intradérmicamente, subcutáneamente, oralmente, pulmonarmente, o cualquier combinación de los mismos. En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se administra intravenosamente.

5 **[0052]** En ciertas realizaciones ilustrativas, el método comprende adicionalmente administrar al sujeto un segundo agente.

[0053] En ciertas realizaciones ilustrativas, el sujeto es un mamífero. En ciertas realizaciones ilustrativas, el sujeto es un ser humano.

[0054] En ciertas realizaciones ilustrativas, la administración de la molécula de ácido nucleico al sujeto da como resultado un aumento de la actividad de FVIII, con respecto a la actividad de FVIII en el sujeto antes de la administración, en donde la actividad de FVIII aumenta en al menos aproximadamente 2 veces, al menos aproximadamente 3 veces, al menos aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos aproximadamente 10 veces, al menos aproximadamente 11 veces, al menos aproximadamente 12 veces, al menos aproximadamente 13 veces, al menos aproximadamente 14 veces, al menos aproximadamente 15 veces, al menos aproximadamente 20 veces, al menos aproximadamente 25 veces, al menos aproximadamente 30 veces, al menos aproximadamente 35 veces, al menos aproximadamente 40 veces, al menos aproximadamente 50 veces, al menos aproximadamente 60 veces, al menos aproximadamente 70 veces, al menos aproximadamente 80 veces, al menos aproximadamente 90 veces, o al menos aproximadamente 100 veces.

[0055] En ciertas realizaciones ilustrativas, el sujeto tiene un trastorno hemorrágico. En ciertas realizaciones ilustrativas, el trastorno hemorrágico es la hemofilia. En ciertas realizaciones ilustrativas, el trastorno hemorrágico es hemofilia A.

25 **[0056]** En ciertos aspectos, se proporciona un método de tratamiento de n trastorno hemorrágico en un sujeto que lo necesita, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean una secuencia genética que comprende una secuencia de polinucleótidos heteróloga que codifica un factor de coagulación, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o

100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181 , 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

5 **[0057]** En ciertos aspectos, se proporciona un método de tratamiento de la hemofilia A en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica el factor VIII (FVIII), en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una
10 secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181 , 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

15 **[0058]** En ciertos aspectos, se proporciona un método de tratamiento de un trastorno metabólico del hígado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica una enzima metabólica asociada al hígado que es deficiente en el sujeto, en donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (no AAV).

20 **[0059]** En ciertas realizaciones ilustrativas, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181 , 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

25 **[0060]** En ciertos aspectos, se proporciona un método de tratamiento de un trastorno metabólico del hígado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica una enzima metabólica asociada al hígado que es deficiente en el sujeto, en donde la primera ITR y/o la segunda ITR
30 comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente

98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181 , 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0061] En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de hebra sencilla. En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de doble hebra.

[0062] En ciertas realizaciones ilustrativas, el trastorno metabólico el hígado se selecciona del grupo que consiste en fenilcetonuria (PKU), una enfermedad del ciclo de la urea, un trastorno de almacenamiento lisosómico, y un trastorno de almacenamiento de glucógeno. En ciertas realizaciones ilustrativas, el trastorno metabólico del hígado es la fenilcetonuria (PKU).

[0063] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se administra intravenosamente, transdérmicamente, intradérmicamente, subcutáneamente, oralmente, pulmonarmente, o cualquier combinación de los mismos. En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se administra intravenosamente.

[0064] En ciertas realizaciones ilustrativas, el método comprende adicionalmente administrar al sujeto un segundo agente.

[0065] En ciertas realizaciones ilustrativas, el sujeto es un mamífero. En ciertas realizaciones ilustrativas, el sujeto es un ser humano.

[0066] En ciertos aspectos, se proporciona un método de tratamiento de la fenilcetonuria (PKU) en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica la fenilalanina hidroxilasa, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181 , 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0067] En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de hebra sencilla. En ciertas realizaciones ilustrativas, el casete genético comprende un ácido nucleico de doble hebra.

- [0068] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico se formula con un agente de suministro. En ciertas realizaciones ilustrativas, el agente de suministro comprende una nanopartícula lipídica.
- 5 [0069] En ciertos aspectos, se proporciona un método de clonación de una molécula de ácido nucleico, que comprende insertar una molécula de ácido nucleico capaz de formar complejos con estructuras secundarias en un en un vector adecuado, e introducir el vector resultante en una cepa anfitriona bacteriana que comprende una alteración en el complejo SbcCD.
- 10 [0070] En ciertas realizaciones ilustrativas, la alteración en el complejo SbcCD comprende una alteración genética en el gen SbcC y/o el gen SbcD. En ciertas realizaciones ilustrativas, la alteración en el complejo SbcCD comprende una alteración genética en el gen SbcC. En ciertas realizaciones ilustrativas, la alteración en el complejo SbcCD comprende una alteración genética en el gen SbcD.
- [0071] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico comprende una primera repetición terminal invertida (ITR) y una segunda ITR, en donde la primera y/o segunda ITR son ITR de un virus no adenoasociado (no AAV).
- 15 [0072] En ciertas realizaciones ilustrativas, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180,
- 20 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.
- [0073] En ciertas realizaciones ilustrativas, la molécula de ácido nucleico comprende adicionalmente un casete genético, en donde el casete genético está flanqueado por la primera ITR y la segunda ITR.
- [0074] En ciertas realizaciones ilustrativas, el casete genético comprende una secuencia de polinucleótidos heteróloga.
- 25 [0075] En ciertas realizaciones ilustrativas, el vector adecuado es un vector de bajo número de copias. En ciertas realizaciones ilustrativas, el vector adecuado es pBR322.
- [0076] En ciertas realizaciones ilustrativas, la cepa anfitriona bacteriana es incapaz de resolver estructuras de ADN cruciformes.
- 30 [0077] En ciertas realizaciones ilustrativas, la cepa anfitriona bacteriana es PMC103, que comprende el genotipo *sbcC*, *recD*, *mcrA*, *AmcrBCF*. En ciertas realizaciones ilustrativas, la cepa anfitriona bacteriana

es PMC107, que comprende el genotipo *recBC*, *recJ*, *sbcBC*, *mcrA*, *AmcrBCF*. En ciertas realizaciones ilustrativas, la cepa anfitriona bacteriana es SURE, que comprende el genotipo *recB*, *recJ*, *sbcC*, *mcrA*, *AmcrBCF*, *umuC*, *uvrC*.

5 [0078] En ciertos aspectos, se proporciona un método para clonar una molécula de ácido nucleico, que comprende insertar una molécula de ácido nucleico susceptible de estructuras secundarias complejas en un vector adecuado, e introducir el vector resultante en una cepa anfitriona bacteriana que comprende una alteración en el complejo SbcCD, en donde la molécula de ácido nucleico comprende una primera repetición terminal invertida (ITR) y una segunda ITR, en donde la primera ITR y/o la segunda ITR comprende una secuencia de nucleótidos al menos aproximadamente 75% al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180,181,183, 184,185,186, 187 o 188, o un derivado funcional de la misma.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

15 [0079] Las FIG. 1A -1B son representaciones esquemáticas de un casete de expresión de un factor de coagulación de una sola hebra (p. ej., FVIII). Se muestran las ubicaciones de la ITR 5' de un virus no AAV (con bucle de horquilla al final de la estructura de ADNhs), la ITR 3' de un virus no AAV (con bucle de horquilla), una secuencia promotora (p. ej., TTPp o CAGp), y una secuencia de transgén, p. ej., la secuencia de Fviiicos6XTEN con un XTEN144 insertado dentro del dominio B. Los casetes de expresión ilustrativos también muestran elementos posibles adicionales, p. ej., una secuencia de intrón, una secuencia de WPREmut y una secuencia de bGHpA.

25 [0080] Las FIG. 1C-1F son representaciones esquemáticas de plásmidos utilizados para preparar casetes de expresión de factor de coagulación de hebra sencilla, tales como el casete mostrado en la FIG. 1A-1B, en donde las ITR del casete derivan de AAV2 (FIG. 1C), B19 (FIG. 1D), GPV (FIG. 1E), o son la secuencia de ITR de B19 de tipo salvaje (FIG. 1F). Una construcción de plásmido que comprende un casete de expresión de FVIIIhs como se muestra aquí se digirió con PvuII (en sitios PvuII) (FIG. 1C) o LguI (en sitios LguI) (FIG. 1D-F) para liberar de manera precisa la secuencia que comprende las ITR y el casete de expresión. El ADN de doble hebra se desnaturalizó térmicamente a 95° C para producir ADNhs y a continuación se incubó a 4° C para permitir la formación de la estructura de la ITR.

30 [0081] La FIG. 2A es un árbol filogenético que ilustra que las relaciones entre diversos miembros de la familia de parvovirus. B19, AAV -2 y GPV se marcan mediante recuadros contorneados.

[0082] La **FIG. 2B** es un dibujo esquemático de los diversos casetes, incluyendo las estructuras de horquilla.

[0083] Las **FIG. 3A y 3B** son alineamientos de las ITR de B19, GPV, y AAV2 (**FIG. 3A**) y B19 y GPV (**FIG. 3B**). El sombreado de color gris muestra homología.

5 [0084] Las **FIG. 4A-4C** muestran la actividad en plasma de FVIII después de la administración de ADN desnudo de FVIII-AAV de hebra sencilla (AAVhs-FVIII; **FIG. 1C**), ADNhs-B19-FVIII (**FIG. 1D**), o ADNhs-GPV-FVIII (**FIG. E**) mediante inyección hidrodinámica (HDI) en ratones Hem-a. Se midió la actividad de FVIII (como porcentaje de los niveles fisiológicos normales en seres humanos) en muestras de plasma a las 24 horas, 3 días, 2 semanas, 3 semanas, 1 mes, 2 meses, 3 meses, 4 meses, 5 meses y 6
10 meses en ratones tratados con una sola HDI de ADNhs a 50 µg/ratón (**FIG. 4C**), 20 µg/ratón (**FIG. 4A y 4B**), 10 µg/ratón (**FIG. 4A, 4B y 4C**), o 5 µg/ratón (**FIG. 4A**). Se proporcionó una HDI de 5 µg/ratón de ADN plasmídico como control (**FIG. 4A, 4B y 4C**).

[0085] La **FIG. 5** muestra la actividad de FVIII en plasma de ratón con hemofilia A después de una sola inyección hidrodinámica de cantidades molares iguales de ADN desnudo de hebra sencilla (AAVhs-FVIII, **FIG. 1A**), ADN de AAV-FVIII de doble hebra que contiene la secuencia de ITR (ADNdh), ADN de FVIII de doble hebra sin la secuencia de ITR (ADNdh sin ITR), o ADN de FVIII de doble hebra circularizado sin ITR o secuencias bacterianas (minicírculo). El ADNdh se generó por escisión
15 enzimática del plásmido AAV-FVIII (**FIG. 2C**) con PvuII pero no desnaturalizado por calor. El ADNdh no se generó por escisión enzimática del plásmido AAV-FVIII (**FIG. 2C**) con AflII y posteriormente
20 purificado. El ADN del minicírculo se generó por ligación del ADNdh sin ITR en sitios AflII. El plasma de ratón se recogió a lo largo de 3 meses o 4 meses y el FVIII se determinó por ensayo de actividad cromogénica.

[0086] La **FIG. 6** muestra la actividad de FVIII en plasma de ratón con hemofilia A después de una inyección hidrodinámica de 30 µg de ADN de FVIII desnudo de hebra sencilla (**FIG. 1A, FIG. 1D-1F**).
25 El plasma se recogió semanalmente durante 7 semanas y se determinó la actividad de FVIII mediante ensayo cromogénico. Después de 35 días (representado como una flecha de color negro), a los ratones que recibían ADNhs de FVIII-B19d135 y FVIII-GPVd162 se les volvieron a administrar 30 µg por medio de inyección hidrodinámica.

[0087] La **FIG. 7A** es una representación esquemática de un casete de expresión de fenilalanina hidroxilasa murina de hebra sencilla (p. ej., PAH). Se muestran las ubicaciones de la ITR 5' de un virus no AAV (con bucle de horquilla en el extremo de la estructura de ADNhs), ITR 3' de un virus no AAV (con bucle de horquilla), una secuencia del promotor (p. ej., CAGp), y una secuencia de transgén, p. ej.,
30

la secuencia 3xFLAG_mPAH. Los casetes de expresión ilustrativos también muestra posibles elementos adicionales, p. ej., la secuencia WPREmut, y la secuencia bGHpA.

[0088] Las FIG. 7B-7D muestran las concentraciones en plasma de fenilalanina (Phe) en ratones con fenilketonuria (PKU) antes (día 0) y después de una sola administración de ADN de hebra sencilla que contiene el ADNc de PAH murino y las ITR de virus no AAV B19d135 o GPVd162 por medio de inyección hidrodinámica. El plasma se recogió los días 3, 7, 14, 28, 42, y 56 después de la administración del ADNhs. Los niveles de fenilalanina residuales se muestran como la concentración en µg/ml (FIG. 7B-7C) o como el porcentaje antes de la administración (FIG. 7D). La línea horizontal representa los niveles de Phe en el momento inicial antes de la administración.

[0089] La FIG. 7E muestra una inmunotransferencia Western de productos lisados de hígado de ratones PKU tratados con ADNhs que contienen el transgén PAH murino y las ITR de B19d135 o GPVd165. Los hígados se recogieron el día 81 después del tratamiento y se extrajeron los productos lisados de proteína. Cada pocillo representa un único animal. La proteína PAH murina etiquetada con FLAG se detectó utilizando el anticuerpo anti-FLAG M2 y se incluyó un control de carga de GAPDH para la comparación.

[0090] Las FIG. 8A-B muestran los niveles de actividad de FVIII en sobrenadante de células Huh7 después de la transducción con ADN de FVIII-AAV (FIG. 1A-1C) encapsulado en nanopartículas de lípidos. El plásmido FVIII-AAV bajo el promotor CAGp (FIG. 1B) se encapsuló a tres razones de amina con respecto a fosfato (NP) y se aplicó a células Huh7 a diversas concentraciones determinadas por el ensayo PicoGreen (FIG. 8A). El plásmido, AAV-FVIII lineal de doble hebra (dh), y de hebra sencilla (hs) bajo el promotor TTPp (FIG 1A) también en nanopartículas de lípidos a dos razones de NP y se utilizó para transducir células Huh7 a diversas concentraciones de ADN (FIG. 8B). FVIII se midió mediante ensayo de actividad cromogénica en comparación con un patrón de plasma FACT humano.

DESCRIPCIÓN DETALLADA DE LA DIVULGACIÓN

[0091] La presente divulgación describe moléculas de ácido nucleico de tipo plásmido que comprenden una primera repetición terminal invertida (ITR), una segunda ITR, y un casete genético, p. ej., que codifica una secuencia diana (también referido como secuencia de polinucleótidos heteróloga), p. ej., una proteína terapéutica o miARN, en donde la primera ITR y/o la segunda ITR son ITR de un virus no adenoasociado (p. ej., la primera ITR y/o la segunda ITR son de un virus no AAV). En algunas realizaciones, el casete genético codifica una proteína terapéutica, p. ej., la secuencia diana codifica una proteína terapéutica. En algunas realizaciones, la proteína terapéutica comprende una proteína seleccionada entre un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo, o una combinación de los mismos. En algunas realizaciones, el

casete genético codifica distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1 , FXN (frataxina), GUCY2D, RS1 , CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa glucosidasa ácida), o cualquier combinación de los mismos.

- 5 **[0092]** En algunas realizaciones, la proteína terapéutica comprende un factor de coagulación. En una realización particular, la proteína terapéutica comprende una proteína FVIII o FIX.

[0093] En algunas realizaciones, el casete genético codifica un miARN. En ciertas realizaciones, el miARN regula por disminución la expresión de un gen diana seleccionado entre SOD1, HTT, RHO, o cualquier combinación de los mismos.

- 10 **[0094]** En ciertas realizaciones, el virus no AAV se selecciona del grupo que consiste en un miembro de la familia viral Parvoviridae y cualquier combinación de los mismos. La presente divulgación se dirige adicionalmente a métodos para expresar una proteína terapéutica, p. ej., un factor de coagulación, p. ej., un FVIII, en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR). , una segunda ITR y un casete genético, p. ej., que codifica una proteína terapéutica o un miARN, en donde la primer ITR y/o la
15 segunda ITR son una ITR de un virus no adenoasociado (no AAV). En determinadas realizaciones, la divulgación describe una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos, que tiene homología de secuencia con una secuencia de nucleótidos seleccionada entre SEQ ID NO: 1 13 y 120.

- 20 **[0095]** En ciertas realizaciones, la presente divulgación proporciona moléculas de ácido nucleico que comprenden una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga, en donde la primera y/o la segunda ITR derivan del parvovirus B19 o parvovirus de ganso (GPV).

- [0096]** Las construcciones ilustrativas de la divulgación se ilustran en las figuras adjuntas y la lista de
25 secuencias. Con el fin de proporcionar una comprensión clara de la memoria descriptiva y las reivindicaciones, a continuación se proporcionan las siguientes definiciones.

I. Definiciones

- [0097]** Se debe observar que el término "uno" o "una" entidad se refiere a una o más de esa entidad: por ejemplo, se entiende que "una secuencia de nucleótidos" representa una o más secuencias de nucleótidos.
30 De manera similar, se entiende que "una proteína terapéutica" y "un miARN" representan una o más

proteínas terapéuticas y uno o más miARN, respectivamente. Como tales, los términos "un" (o "una" o "uno"), "uno o más" y "al menos uno" se pueden utilizar indistintamente en la presente memoria.

[0098] El término "aproximadamente" se utiliza en la presente memoria para significar aproximadamente, groseramente, alrededor de o en las regiones de. Cuando el término

5 "aproximadamente" se utiliza junto con un intervalo numérico, modifica ese intervalo extendiendo los límites por encima y por debajo de los valores numéricos establecidos. En general, el término "aproximadamente" se utiliza en la presente memoria para modificar un valor numérico por encima y por debajo del valor establecido mediante una variación de 10 por ciento, hacia arriba o hacia abajo (mayor o menor).

10 **[0099]** También como se emplea en la presente memoria, "y/o" se refiere a, y abarca todas y cada una de las combinaciones posibles de uno o más de los elementos enumerados asociados, así como la falta de combinaciones cuando se interpreta en la alternativa ("o").

[0100] "Ácidos nucleicos", "moléculas de ácido nucleico", "nucleótidos", "secuencia de nucleótidos" y "polinucleótido" se utilizan indistintamente y se refieren a la forma polimérica de éster fosfato de

15 ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualquier fosfoéster análogo de los mismos, tal como fosforotioatos y tioésteres, en forma de hebra sencilla o de doble hélice. Las secuencias de ácidos nucleicos de hebra sencilla se refieren a ADN de hebra sencilla (ADNhs) o ARN de hebra sencilla (ARNhs). Son posibles las hélices de ADN-
20 ADN, ADN-ARN y ARN-ARN de doble hebra. El término molécula de ácido nucleico, y en particular molécula de ADN o ARN, se refiere únicamente a la estructura primaria y secundaria de la molécula, y no la limita a ninguna forma terciaria particular. Por tanto, este término incluye ADN de doble hebra que se encuentra, entre otros, en moléculas de ADN lineales o circulares (p. ej., fragmentos de restricción), plásmidos, ADN superenrollado y cromosomas. Al discutir la estructura de moléculas particulares de
25 ADN de doble hebra, las secuencias se pueden describir en la presente memoria de acuerdo con la convención normal de dar solo la secuencia en la dirección 5 'a 3' a lo largo de la hebra de ADN no transcrita (es decir, la hebra tiene una secuencia homóloga al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha sufrido una manipulación biológica molecular. El ADN incluye, pero no se limita a, ADNc, ADN genómico, ADN plasmídico, ADN sintético y ADN
30 semisintético. Una "composición de ácido nucleico" de la divulgación comprende uno o más ácidos nucleicos como se describe en la presente memoria.

[0101] Como se emplea en la presente memoria, una "repetición terminal invertida" (o "ITR") se refiere a una subsecuencia de ácido nucleico ubicada en el extremo 5' o 3' de una secuencia de ácido nucleico de hebra sencilla, que comprende un conjunto de nucleótidos (secuencia inicial) seguido aguas abajo por su complemento inverso, es decir, secuencia palindrómica. La secuencia de nucleótidos intermedia entre la secuencia inicial y el complemento inverso puede tener cualquier longitud, incluyendo cero. En una realización, la ITR útil para la presente divulgación comprende una o más "secuencias palindrómicas". Una ITR puede tener varias funciones. En algunas realizaciones, una ITR descrita en la presente memoria forma una estructura de horquilla. En algunas realizaciones, la ITR forma una estructura de horquilla en forma de T. En algunas realizaciones, la ITR forma una estructura de horquilla sin forma de T, p. ej., una estructura de horquilla en forma de U. En algunas realizaciones, la ITR promueve la supervivencia a largo plazo de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR promueve la supervivencia permanente de la molécula de ácido nucleico en el núcleo de una célula (p. ej., durante toda la vida útil de la célula). En algunas realizaciones, la ITR promueve la estabilidad de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR promueve la retención de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR promueve la persistencia de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR inhibe o previene la degradación de la molécula de ácido nucleico en el núcleo de una célula.

[0102] En una realización, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-600 nucleótidos, aproximadamente 2-550 nucleótidos, aproximadamente 2-500 nucleótidos, aproximadamente 2-450 nucleótidos, aproximadamente 2-400 nucleótidos, aproximadamente 2-350 nucleótidos, aproximadamente 2-300 nucleótidos, o aproximadamente 2-250 nucleótidos. En algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 5-600 nucleótidos, aproximadamente 10-600 nucleótidos, aproximadamente 15-600 nucleótidos, aproximadamente 20-600 nucleótidos, aproximadamente 25-600 nucleótidos, aproximadamente 30-600 nucleótidos, aproximadamente 35-600 nucleótidos, aproximadamente 40-600 nucleótidos, aproximadamente 45-600 nucleótidos, aproximadamente 50-600 nucleótidos, aproximadamente 60-600 nucleótidos, aproximadamente 70-600 nucleótidos, aproximadamente 80-600 nucleótidos, aproximadamente 90-600 nucleótidos, aproximadamente 100-600 nucleótidos, aproximadamente 150-600 nucleótidos, aproximadamente 200-600 nucleótidos, aproximadamente 300-600 nucleótidos, aproximadamente 350-600 nucleótidos, aproximadamente 400-600 nucleótidos, aproximadamente 450-600 nucleótidos, aproximadamente 500-600 nucleótidos, o aproximadamente 550-600 nucleótidos. En algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 5-550 nucleótidos, aproximadamente 5 a 500 nucleótidos, aproximadamente 5-450

nucleótidos, aproximadamente 5 a 400 nucleótidos, aproximadamente 5-350 nucleótidos, aproximadamente 5 a 300 nucleótidos, o aproximadamente 5-250 nucleótidos. En algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 10-550 nucleótidos, aproximadamente 15-500 nucleótidos, aproximadamente 20-450 nucleótidos, aproximadamente 25-400

5 nucleótidos, aproximadamente 30-350 nucleótidos, aproximadamente 35-300 nucleótidos, o aproximadamente 40-250 nucleótidos. En ciertas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 225 nucleótidos, aproximadamente 250 nucleótidos, aproximadamente 275 nucleótidos, aproximadamente 300 nucleótidos, aproximadamente 325 nucleótidos, aproximadamente 350 nucleótidos, aproximadamente 375 nucleótidos, aproximadamente

10 400 nucleótidos, aproximadamente 425 nucleótidos, aproximadamente 450 nucleótidos, aproximadamente 475 nucleótidos, aproximadamente 500 nucleótidos, aproximadamente 525 nucleótidos, aproximadamente 550 nucleótidos, aproximadamente 575 nucleótidos, o aproximadamente 600 nucleótidos. En realizaciones particulares, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 400 nucleótidos.

15 **[0103]** En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-200 nucleótidos, aproximadamente 5-200 nucleótidos, aproximadamente 10-200 nucleótidos, aproximadamente 20-200 nucleótidos, aproximadamente 30-200 nucleótidos, aproximadamente 40-200 nucleótidos, aproximadamente 50-200 nucleótidos, aproximadamente 60-200 nucleótidos, aproximadamente 70-200 nucleótidos, aproximadamente 80-200 nucleótidos,

20 aproximadamente 90-200 nucleótidos, aproximadamente 100-200 nucleótidos, aproximadamente 125-200 nucleótidos, aproximadamente 150-200 nucleótidos, o aproximadamente 175-200 nucleótidos. En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-150 nucleótidos, aproximadamente 5-150 nucleótidos, aproximadamente 10-150 nucleótidos, aproximadamente 20-150 nucleótidos, aproximadamente 30-150 nucleótidos, aproximadamente 40-150

25 nucleótidos, aproximadamente 50-150 nucleótidos, aproximadamente 75-150 nucleótidos, aproximadamente 100-150 nucleótidos, o aproximadamente 125-150 nucleótidos. En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-100 nucleótidos, aproximadamente 5-100 nucleótidos, aproximadamente 10-100 nucleótidos, aproximadamente 20-100 nucleótidos, aproximadamente 30-100 nucleótidos, aproximadamente 40-100 nucleótidos,

30 aproximadamente 50-100 nucleótidos, o aproximadamente 75-100 nucleótidos. En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-50 nucleótidos, aproximadamente 10-50 nucleótidos, aproximadamente 20-50 nucleótidos, aproximadamente 30-50 nucleótidos, aproximadamente 40-50 nucleótidos, aproximadamente 3-30 nucleótidos, aproximadamente 4-20 nucleótidos, o aproximadamente 5-10 nucleótidos. En otra realización, la secuencia inicial y/o el

complemento inverso consisten en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos, nueve nucleótidos, diez nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, 17 nucleótidos, 18 nucleótidos, 19 nucleótidos, o 20 nucleótidos. En otras realizaciones, un nucleótido intermedio entre la secuencia inicial y el complemento inverso es (p. ej., consiste en) 0 nucleótido, 1 nucleótido, dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos, nueve nucleótidos, 10 nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, 17 nucleótidos, 18 nucleótidos, 19 nucleótidos, o 20 nucleótidos.

[0104] Por lo tanto, una "ITR" como se emplea en la presente memoria puede plegarse sobre sí misma y formar un segmento de doble hebra. Por ejemplo, la secuencia GATCXXXXGATC comprende una secuencia inicial de GATC y su complemento (3'CTAG5 ') cuando se pliega para formar una doble hélice. En algunas realizaciones, la ITR comprende una secuencia palindrómica continua (por ejemplo, GATCGATC) entre la secuencia inicial y el complemento inverso. En algunas realizaciones, la ITR comprende una secuencia palindrómica interrumpida (por ejemplo, GATCXXXXGATC) entre la secuencia inicial y el complemento inverso. En algunas realizaciones, las secciones complementarias de la secuencia palindrómica continua o interrumpida interactúan entre sí para formar una estructura de "bucle en horquilla". Como se emplea en la presente memoria, una estructura de "bucle en horquilla" se produce cuando al menos dos secuencias complementarias en una molécula de nucleótidos de hebra sencilla se emparejan por bases para formar una sección de doble hebra. En algunas realizaciones, solo una parte de la ITR forma un bucle en horquilla. En otras realizaciones, toda la ITR forma un bucle en horquilla.

[0105] En la presente divulgación, al menos una ITR es una ITR de un virus no adenoasociado (no AAV). En ciertas realizaciones, la ITR es una ITR de un miembro no AAV de la familia viral *Parvoviridae*. En algunas realizaciones, la ITR es una ITR de un miembro no AAV del género *Dependovirus* o del género *Erythrovirus*. En realizaciones particulares, la ITR es una ITR de un parvovirus de ganso (GPV), un parvovirus de pato real (MDPV) o un parvovirus B19 de eritrovirus (también conocido como parvovirus B19, eritroparvovirus 1 de primates, virus B19 y eritrovirus). En ciertas realizaciones, una ITR de las dos ITR es una ITR de un AAV. En otras realizaciones, una ITR de las dos ITR en la construcción es una ITR de un serotipo AAV seleccionado entre los serotipos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 y cualquier combinación de los mismos. En una realización particular, la ITR se obtiene del serotipo 2 de AAV, p. ej., una ITR de serotipo 2 de AAV.

[0106] En ciertos aspectos de la presente divulgación, la molécula de ácido nucleico comprende dos ITR, una ITR 5' y una ITR 3', en donde la ITR 5' está ubicada en el extremo 5' de la molécula de ácido nucleico, y la ITR 3' está ubicada en el extremo 3' de la molécula de ácido nucleico. La ITR 5' y la ITR 3' se pueden obtener del mismo virus o de virus diferentes. En ciertas realizaciones, la ITR 5' se obtiene de un AAV y la ITR 3' no se obtiene de un virus AAV (p. ej., un virus no AAV). En algunas realizaciones, la ITR 3' se obtiene de un AAV y la ITR 5' no se obtiene de un virus AAV (p. ej., un virus no AAV). En otras realizaciones, la ITR 5' no se obtiene de un virus AAV (p. ej., un virus no AAV), y la ITR 3' se obtiene del mismo o diferente virus no AAV.

[0107] El término "parvovirus", como se emplea en la presente memoria, abarca la familia *Parvoviridae*, que incluye, pero no se limita a, parvovirus que se replican de forma autónoma y Dependovirus. Los parvovirus autónomos incluyen, por ejemplo, miembros de los géneros *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus* y *Penstildensovirus*.

[0108] Los parvovirus autónomos ilustrativos incluyen, pero no se limitan a parvovirus porcino, virus mínimo de ratón, parvovirus canino, virus de la enteritis de visón, parvovirus bovino, parvovirus de pollo, virus de panleucopenia felina, parvovirus felino, parvovirus de ganso, parvovirus H1, parvovirus de pato real, parvovirus de serpiente y virus B19. Otros parvovirus autónomos son conocidos por los expertos en la técnica. Véase, p. ej., FIELDS et al VIROLOGY, volumen 2, capítulo 69 (4ª ed., Lippincott-Raven Publishers).

[0109] El término "no AAV" como se emplea en la presente abarca ácidos nucleicos, proteínas y virus de la familia *Parvoviridae* excluyendo cualquiera de los virus adenoasociados (AAV) de la familia *Parvoviridae*. El término "No AAV" incluye, pero no se limita a, miembros de replicación autónoma de los géneros *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, y *Penstildensovirus*.

[0110] Como se emplea en la presente memoria, el término "virus adenoasociado" (AAV), incluye pero no se limita a, AAV tipo 1, AAV tipo 2, AAV tipo 3 (incluyendo los tipos 3A y 3B) AAV tipo 4, AAV tipo 5, AAV tipo 6, AAV tipo 7, AAV tipo 8, AAV tipo 9, AAV tipo 10, AAV tipo 11, AAV tipo 12, AAV tipo 13, AAV de serpiente, AAV aviar, AAV bovino, AAV canino, AAV equino, AAV ovino, AAV de cabra, AAV de camarón, aquellos serotipos de AAV y clados descritos por Gao et al. (J Virol. 78: 6381 (2004)) y Moris et al. (Virol. 33:375 (2004)), y cualquier otro AAV ahora conocido o

descubierto posteriormente. Véase, p. ej., FIELDS et al VIROLOGY, volumen 2, capítulo 69 (4ª ed., Lippincott-Raven Publishers).

[0111] El término obtenido de, como se emplea en la presente memoria, se refiere a un componente que se aísla de o se elabora utilizando una molécula o organismo especificados, o información (p. ej.,
5 secuencia de aminoácidos o ácido nucleico) de la molécula o organismo especificados. Por ejemplo, una secuencia de ácido nucleico (p. ej., ITR) que se obtiene de una segunda secuencia de ácido nucleico (p. ej., ITR) puede incluir una secuencia de nucleótidos que es idéntica o sustancialmente similar a la secuencia de nucleótidos de la segunda secuencia de ácido nucleico. En el caso de nucleótidos o polipéptidos, las especies \$derivadas se pueden obtener, por ejemplo, por mutagénesis natural,
10 mutagénesis dirigida artificial o mutagénesis aleatoria artificial. La mutagénesis utilizada para obtener nucleótidos o polipéptidos puede ser dirigida intencionadamente o intencionadamente aleatoria, o una mezcla de cada una. La mutagénesis de un nucleótido o polipéptido para crear un nucleótido o polipéptido diferentes obtenidos del primero puede ser un evento aleatorio (p. ej., causado por la infidelidad de la polimerasa) y la identificación del nucleótido o polipéptido obtenidos se puede realizar
15 mediante métodos de escrutinio apropiados, p. ej., como se describe en la presente memoria. La mutagénesis de un polipéptido implica típicamente la manipulación del polinucleótido que codifica el polipéptido. En algunas realizaciones, una secuencia de nucleótidos o aminoácidos que se obtiene de una segunda secuencia de nucleótidos o aminoácidos tiene una identidad de secuencia de al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%,
20 al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al
25 menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% con la segunda secuencia de nucleótidos o aminoácidos, respectivamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En otras realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 90% idéntica a la ITR de un virus no
30 AAV (o AAV ITR, respectivamente), en donde la ITR de un virus no AAV (o AAV) conserva la propiedad funcional de la ITR de un virus no AAV (o ITR de AAV, respectivamente). En algunas realizaciones, un ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 80% idéntica a la ITR de no AAV (o ITR de AAV, respectivamente), en donde la ITR de no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente). En algunas realizaciones,

una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 70% idéntica a la ITR de no AAV (o ITR de AAV, respectivamente), en donde la ITR de no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente). En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 60% idéntica a la ITR de no AAV (o ITR de AAV, respectivamente), en donde la ITR de no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente). En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 50% idéntica a la ITR de no AAV (o ITR de AAV, respectivamente), en donde la ITR de no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente).

- 10 **[0112]** En ciertas realizaciones, una ITR obtenida de un no AAV (o AAV) comprende o consiste en un fragmento de la ITR de no AAV (o AAV). En algunas realizaciones, la ITR obtenida de una ITR de no AAV (o AAV) comprende o consiste en un fragmento de la ITR del no AAV (o AAV), en donde el fragmento comprende al menos aproximadamente 5 nucleótidos, al menos aproximadamente 10 nucleótidos, al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 25 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 45 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 55 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 65 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 75 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 85 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 95 nucleótidos, al menos aproximadamente 100 nucleótidos, al menos aproximadamente 125 nucleótidos, al menos aproximadamente 150 nucleótidos, al menos aproximadamente 175 nucleótidos, al menos aproximadamente 200 nucleótidos, al menos aproximadamente 225 nucleótidos, al menos aproximadamente 250 nucleótidos, al menos aproximadamente 275 nucleótidos, al menos aproximadamente 300 nucleótidos, al menos aproximadamente 325 nucleótidos, al menos aproximadamente 350 nucleótidos, al menos aproximadamente 375 nucleótidos, al menos aproximadamente 400 nucleótidos, al menos aproximadamente 425 nucleótidos, al menos aproximadamente 450 nucleótidos, al menos aproximadamente 475 nucleótidos, al menos aproximadamente 500 nucleótidos, al menos aproximadamente 525 nucleótidos, al menos aproximadamente 550 nucleótidos, al menos aproximadamente 575 nucleótidos, o al menos aproximadamente 600 nucleótidos; en donde la ITR obtenida de una ITR de un virus no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente). En ciertas realizaciones, la ITR obtenida de una ITR de un virus

no AAV (o AAV) comprende o consiste en un fragmento de la ITR del no AAV (o AAV), en donde el fragmento comprende al menos aproximadamente 129 nucleótidos, y en donde la ITR obtenida de una ITR de un virus no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente). En ciertas realizaciones, la ITR obtenida de una ITR de un virus no AAV (o AAV) comprende o consiste en un fragmento de la ITR del no AAV (o AAV), en donde el fragmento comprende al menos aproximadamente 102 nucleótidos, y en donde la ITR obtenida de una ITR de un virus no AAV (o AAV) conserva una propiedad funcional de la ITR de no AAV (o ITR de AAV, respectivamente).

[0113] En algunas realizaciones, la ITR obtenida de una ITR de un virus no AAV (o AAV) comprende o consiste en un fragmento de la ITR del no AAV (o AAV), en donde el fragmento comprende al menos aproximadamente 5%, al menos aproximadamente 10%, al menos aproximadamente 15%, al menos aproximadamente 20%, al menos aproximadamente 25%, al menos aproximadamente 30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 45%, al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de la longitud de la ITR del no AAV (o AAV).

[0114] En ciertas realizaciones, una secuencia de nucleótidos o aminoácidos que se obtiene de una segunda secuencia de nucleótidos o aminoácidos tiene una identidad de secuencia de al menos 50%, al menos 51 %, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61 %, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72 %, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93 %, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% con una porción homóloga de la segunda secuencia de nucleótidos o aminoácidos, respectivamente, cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En otras realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 90% idéntica a una porción homóloga de la ITR de no AAV (o ITR de AAV respectivamente),

cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 80% idéntica a una porción homóloga de la ITR de no AAV (o ITR de AAV respectivamente), cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 70% idéntica a una porción homóloga de la ITR de no AAV (o ITR de AAV respectivamente), cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 60% idéntica a una porción homóloga de la ITR de no AAV (o ITR de AAV respectivamente), cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, una ITR obtenida de una ITR de un virus no AAV (o AAV) es al menos 50% idéntica a una porción homóloga de la ITR de no AAV (o ITR de AAV respectivamente), cuando se alinean adecuadamente, en donde la primera secuencia de nucleótidos o aminoácidos conserva la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos.

[0115] Un vector o molécula de ácido nucleico "libre de cápside" o "sin cápside" se refiere a una construcción de vector libre de una cápside. En algunas realizaciones, el vector o molécula de ácido nucleico sin cápside no contiene secuencias que codifican, p. ej. , una proteína Rep de AAV.

[0116] Como se emplea en la presente memoria, una "región codificante" o "secuencia codificante" es una porción de un polinucleótido que consiste en codones traducibles a aminoácidos. Aunque un "codón de parada" (TAG, TGA, o TAA) no es típicamente traducible a aminoácido, se puede considerar que es parte de una región codificante, pero cualquiera de las secuencias flanqueantes, por ejemplo promotores, sitios de unión al ribosoma, terminadores transcripcionales, intrones, y similares, no son parte de una región codificante. Los límites de una región codificante están determinados típicamente por un codón de inicio en el extremo 5', que codifica el extremo amino del polipéptido resultante, y un codón de parada de la traducción en el extremo 3', que codifica el extremo carboxilo del polipéptido resultante. Pueden estar presentes dos o más regiones codificantes en una sola construcción de polinucleótidos, p. ej., en un único vector, o en construcciones de polinucleótidos separadas, p. ej., en vectores separados (diferentes). Se deduce, en ese caso, que un solo vector puede contener únicamente una sola región codificante, o comprender dos o más regiones codificantes.

[0117] Ciertas proteínas secretadas por células de mamífero están asociadas con un péptido señal secretor que se escinde de la proteína madura una vez que la exportación de la cadena de proteína en crecimiento a través del retículo endoplásmico rugoso se ha iniciado. Los expertos en la técnica son conscientes de que los péptidos señal se fusionan generalmente al extremo N del polipéptido, y son escindidos del polipéptido completo o de "longitud completa" para producir una forma secretada o "madura" del polipéptido. En ciertas realizaciones, se puede utilizar un péptido señal nativo o un derivado funcional de esa secuencia que conserva la capacidad para dirigir la secreción del polipéptido que está asociado operablemente con él. Alternativamente, se puede utilizar un péptido señal de mamífero heterólogo, p. ej., un péptido señal de activador de plasminógeno tisular humano (TPA) o de β -glucuronidasa de ratón, o un derivado funcional del mismo.

[0118] El término "aguas abajo" se refiere a una secuencia de nucleótidos que está ubicada 3' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos aguas abajo se refieren a secuencias que siguen al punto de partida de la transcripción. Por ejemplo, el codón de inicio de la traducción de un gen está ubicado aguas abajo del sitio de inicio de la transcripción.

[0119] El término "aguas arriba" se refiere a una secuencia de nucleótidos que está ubicada 5' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos aguas arriba se refieren a secuencias que están ubicadas en el lado 5' de una región codificante o punto de partida de la transcripción. Por ejemplo, la mayoría de los promotores están ubicados aguas arriba del sitio de inicio de la transcripción.

[0120] Como se emplea en la presente memoria, el término "región reguladora de genes" o "región reguladora" se refiere a secuencias de nucleótidos ubicadas aguas arriba (secuencias no codificantes 5'), dentro o aguas abajo (secuencias no codificantes 3') de una región codificante y que influyen en la transcripción, el procesamiento del ARN, la estabilidad o la traducción de la región codificante asociada. Las regiones reguladoras pueden incluir promotores, secuencias líder de traducción, intrones, secuencias de reconocimiento de poliadenilación, sitios de procesamiento de ARN, sitios de unión efectores o estructuras de tallo-bucle. Si una región codificante está destinada a la expresión en una célula eucariótica, normalmente se ubicarán una señal de poliadenilación y una secuencia de terminación de la transcripción en 3' con respecto a la secuencia codificante

[0121] Un polinucleótido que codifica un producto, p. ej., un miARN o un producto génico (p. ej., un polipéptido tal como una proteína terapéutica), puede incluir un promotor y/u otros elementos de control de la expresión (p. ej., transcripción o traducción) asociados operablemente con una o más regiones

codificantes. En una asociación operable, una región codificante para un producto génico, p. ej., un polipéptido, se asocia con una o más regiones reguladoras de tal manera que coloque la expresión del producto génico bajo la influencia o el control de la región o regiones reguladoras. Por ejemplo, una región codificante y un promotor están "operablemente asociados" si la inducción de la función del promotor da como resultado la transcripción del ARNm que codifica el producto génico codificado por la región codificante, y si la naturaleza de la conexión entre el promotor y la región codificante no interfiere en la capacidad del promotor para dirigir la expresión del producto génico o interfiere en la capacidad de transcripción del molde de ADN. Otros elementos de control de la expresión, además de un promotor, por ejemplo potenciadores, operadores, represores y señales de terminación de la transcripción, también pueden asociarse operablemente con una región codificante para dirigir la expresión del producto génico.

[0122] Las "secuencias de control de la transcripción" se refieren a secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una secuencia codificante en una célula anfitriona. Los expertos en la técnica conocen una variedad de regiones de control de la transcripción. Estas incluyen, sin limitación, regiones de control de la transcripción que funcionan en células de vertebrados, tales como, pero no limitados a, segmentos promotores y potenciadores de citomegalovirus (el promotor temprano inmediato, junto con el intrón-A), virus simio 40 (el promotor temprano) y retrovirus (tal como el virus del sarcoma de Rous). Otras regiones de control de la transcripción incluyen las obtenidas de genes de vertebrados tales como actina, proteína de choque térmico, hormona del crecimiento bovina y β -globina de conejo así como otras secuencias capaces de controlar la expresión génica en células eucarióticas. Las regiones de control de la transcripción adecuadas adicionales incluyen promotores y potenciadores específicos de tejido, así como promotores inducibles por linfocinas (p. ej., promotores inducibles por interferones o interleucinas).

[0123] De manera similar, los expertos en la técnica conocen una variedad de elementos de control de la traducción. Estos incluyen, pero no se limitan a, sitios de unión al ribosoma, codones de inicio y terminación de la traducción, y elementos derivados de picornavirus (concretamente un sitio interno de entrada al ribosoma, o IRES, también referido como secuencia CITE).

[0124] El término "expresión" como se emplea en la presente memoria se refiere a un proceso mediante el cual un polinucleótido produce un producto génico, por ejemplo, un ARN o un polipéptido. Incluye, sin limitación, la transcripción del polinucleótido a ARN mensajero (ARNm), ARN de transferencia (ARNt), ARN en horquilla pequeña (ARNhp), ARN interferencia pequeño (ARNip) o cualquier otro producto de ARN, y la traducción de un ARNm a un polipéptido. La expresión produce un "producto génico". Como se emplea en la presente memoria, un producto génico puede ser un ácido nucleico, p. ej., un ARN mensajero producido por transcripción de un gen, o un polipéptido que se traduce a partir de un

transcrito. Los productos génicos descritos en la presente memoria incluyen adicionalmente ácidos nucleicos con modificaciones post-transcripcionales, p. ej., poliadenilación o corte y empalme, o polipéptidos con modificaciones post-traduccionales, p. ej., metilación, glicosilación, adición de lípidos, asociación con otras subunidades de proteínas, o escisión proteolítica. El término "rendimiento", como se emplea en la presente memoria, se refiere a la cantidad de un polipéptido producido por la expresión de un gen.

[0125] Un "vector" se refiere a cualquier vehículo para la clonación y/o transferencia de un ácido nucleico a una célula anfitriona. Un vector puede ser un replicón al que se puede unir otro segmento de ácido nucleico para provocar la replicación del segmento unido. Un "replicón" se refiere a cualquier elemento genético (p. ej., plásmido, fago, cósmido, cromosoma, virus) que funciona como una unidad autónoma de replicación *in vivo*, es decir, capaz de replicar bajo su propio control. El término "vector" incluye vehículos para introducir el ácido nucleico en una célula *in vitro*, *ex vivo* o *in vivo*. Se conoce y se utiliza en la técnica un gran número de vectores que incluyen, por ejemplo, plásmidos, virus eucarióticos modificados o virus bacterianos modificados. La inserción de un polinucleótido en un vector adecuado se puede lograr ligando fragmentos de polinucleótidos apropiados en un vector seleccionado que tenga extremos cohesivos complementarios.

[0126] Los vectores se pueden diseñar para codificar marcadores o informadores seleccionables que proporcionen la selección o identificación de células que han incorporado el vector. La expresión de marcadores o informadores seleccionables permite la identificación y/o selección de células anfitrionas que incorporan y expresan otras regiones codificantes contenidas en el vector. Los ejemplos de genes marcadores seleccionables conocidos y utilizados en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomycin, gentamicina, kanamicina, higromicina, herbicida bialafos, sulfonamida y similares; y genes que se utilizan como marcadores fenotípicos, es decir, genes reguladores de antocianina, gen de isopentanol transferasa y similares. Los ejemplos de informadores conocidos y utilizados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), β -galactosidasa (LacZ), β -glucuronidasa (Gus) y similares. Los marcadores seleccionables también pueden considerarse informadores.

[0127] El término "célula anfitriona" como se emplea en la presente memoria se refiere, por ejemplo, a microorganismos, células de levadura, células de insectos y células de mamíferos, que pueden ser, o han sido, utilizados como receptores de ADNhs o vectores. El término incluye la progenie de la célula original que ha sido transducida. Por tanto, una "célula anfitriona" como se emplea en la presente memoria generalmente se refiere a una célula que ha sido transducida con una secuencia de ADN exógena. Se entiende que la progenie de una única célula parental puede no ser necesariamente

completamente idéntica en cuanto a morfología o complemento genómico o de ADN total al parental original, debido a una mutación natural, accidental o deliberada. En algunas realizaciones, la célula anfitriona puede ser una célula anfitriona in vitro.

[0128] El término "marcador seleccionable" se refiere a un factor de identificación, normalmente un gen de resistencia a antibióticos o agentes químicos, que se puede seleccionar basándose en el efecto del gen marcador, es decir, resistencia a un antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes y similares, en los que el efecto se utiliza para rastrear la herencia de un ácido nucleico de interés y/o para identificar una célula u organismo que ha heredado el ácido nucleico de interés. Los ejemplos de genes marcadores seleccionables conocidos y utilizados en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomicina, gentamicina, kanamicina, higromicina, herbicida bialafos, sulfonamida y similares; y genes que se utilizan como marcadores fenotípicos, es decir, genes reguladores de antocianina, gen de isopentanol transferasa y similares.

[0129] El término "gen informador" se refiere a un ácido nucleico que codifica un factor de identificación que puede identificarse basándose en el efecto del gen informador, en donde el efecto se utiliza para rastrear la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés y/o para medir la inducción o transcripción de la expresión génica. Los ejemplos de genes informadores conocidos y utilizados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), β -galactosidasa (LacZ), β -glucuronidasa (Gus) y similares. Los genes marcadores seleccionables también pueden considerarse genes informadores.

[0130] "Promotor" y "secuencia promotora" se utilizan indistintamente y se refieren a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una secuencia codificante se ubica en 3' con respecto a una secuencia promotora. Los promotores se pueden obtener en su totalidad de un gen nativo, o estar compuestos de diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintéticos. Los expertos en la técnica entenderán que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes fases de desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas. Los promotores que hacen que un gen se exprese en la mayoría de los tipos de células en la mayoría de los casos se denominan comúnmente "promotores constitutivos". Los promotores que hacen que un gen se exprese en un tipo celular específico se denominan comúnmente "promotores específicos de células" o "promotores específicos de tejidos". Los promotores que hacen que un gen se exprese en una fase específica de desarrollo o diferenciación celular se denominan

comúnmente "promotores específicos del desarrollo" o "promotores específicos de la diferenciación celular". Los promotores que se inducen y hacen que un gen se exprese después de la exposición o el tratamiento de la célula con un agente, molécula biológica, agente químico, ligando, luz o similar que induce el promotor se denominan comúnmente "promotores inducibles" o "promotores regulables".

- 5 Adicionalmente, se reconoce que, dado que en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido completamente, fragmentos de ADN de diferentes longitudes pueden tener una actividad promotora idéntica.

[0131] La secuencia del promotor está típicamente unida en su extremo 3' por el sitio de inicio de la transcripción y se extiende aguas arriba (dirección 5') para incluir el número mínimo de bases o
10 elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro de la secuencia del promotor se encontrará un sitio de inicio de la transcripción (convenientemente definido, por ejemplo, mediante mapeo con nucleasa S1), así como dominios de unión a proteínas (secuencias consenso) responsables de la unión de la ARN polimerasa.

[0132] En algunas realizaciones, la molécula de ácido nucleico comprende un promotor específico de
15 tejido. En determinadas realizaciones, el promotor específico de tejido dirige la expresión de la proteína terapéutica, p. ej., el factor de coagulación, en el hígado, p. ej., en hepatocitos y/o células endoteliales. En particular, en las realizaciones, el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor de factor VIII humano endógeno (F8), un promotor de alfa-1-antitripsina humana (hAAT), un promotor mínimo de albúmina humana, un promotor de albúmina de
20 ratón, un promotor de tristetraprolina (TTP), un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV), un promotor de fosfoglicerato quinasa (PGK) y cualquier combinación de los mismos. En algunas realizaciones, el promotor se selecciona entre un promotor específico de hígado (p. ej., α 1-antitripsina (AAT)), un promotor específico de músculo (p. ej., creatina quinasa muscular (MCK), cadena pesada alfa de miosina (α MHC), mioglobina (MB) y desmina (DES)), un promotor sintético (p.
25 ej., SPc5-12, 2R5Sc5-12, dMCK y tMCK) y cualquier combinación de los mismos. En una realización particular, el promotor comprende un promotor de TTP

[0133] Los términos "endonucleasa de restricción" y "enzima de restricción" se utilizan indistintamente y se refieren a una enzima que se une y corta dentro de una secuencia de nucleótidos específica dentro del ADN de doble hebra.

30 **[0134]** El término "plásmido" se refiere a un elemento extra-cromosómico que a menudo porta un gen que no es parte del metabolismo central de la célula, y normalmente está en forma de moléculas de ADN de doble hebra circulares. Tales elementos pueden ser secuencias de replicación autónoma, secuencias

integrantes del genoma, secuencias de fagos o nucleótidos, ADN o ARN de hebra sencilla o doble lineal, circular, o superenrollado, derivados de cualquier fuente, en los que numerosas secuencias de nucleótidos se han unido o recombinado en una única construcción, que es capaz de introducir en una célula un fragmento de promotor y una secuencia de ADN para un producto génico seleccionado junto con una secuencia no traducida 3' apropiada.

[0135] Los vectores virales eucarióticos que se pueden utilizar incluyen, pero no se limitan a, vectores de adenovirus, vectores de retrovirus, vectores de virus adenoasociados, poxvirus, p. ej., vectores de virus vaccinia, vectores de baculovirus o vectores de herpesvirus. Los vectores no virales incluyen plásmidos, liposomas, lípidos cargados eléctricamente (citofectinas), complejos de ADN-proteína y biopolímeros.

[0136] Un "vector de clonación" se refiere a un "replicón", que es una unidad de longitud de un ácido nucleico que replica secuencialmente y que comprende un origen de replicación, tal como un plásmido, fago o cósmido, al que se puede unir otro segmento de ácido nucleico puede adjuntarse para provocar la replicación del segmento adjunto. Ciertos vectores de clonación son capaces de replicar en un tipo de célula, p. ej., bacterias, y de expresarse en otro, p. ej., células eucarióticas. Los vectores de clonación típicamente comprenden una o más secuencias que se pueden utilizar para la selección de células que comprenden el vector y/o uno o más sitios de clonación múltiples para la inserción de secuencias de ácido nucleico de interés.

[0137] El término "vector de expresión" se refiere a un vehículo diseñado para permitir la expresión de una secuencia de ácido nucleico insertada después de la inserción en una célula anfitriona. La secuencia de ácido nucleico insertada se coloca en asociación operable con regiones reguladoras como se describió anteriormente.

[0138] Los vectores se introducen en las células anfitrionas mediante métodos bien conocidos en la técnica, p. ej., transfección, electroporación, microinyección, transducción, fusión celular, DEAE dextrano, precipitación con fosfato cálcico, lipofección (fusión de lisosomas), uso de una pistola de genes, o un transportador de vector de ADN. "Cultivo", "cultivar" y "cultivando", como se emplean en la presente memoria, significan incubar células en condiciones *in vitro* que permiten el crecimiento o división celular o mantener las células en un estado vivo. "Células cultivadas", como se emplea en la presente memoria, significa células que se propagan *in vitro*.

[0139] Como se emplea en la presente memoria, se pretende que el término "polipéptido" abarque un "polipéptido" singular así como también "polipéptidos" en plural, y se refiere a una molécula compuesta de monómeros (aminoácidos) conectados linealmente por enlaces amida (también conocidos como enlaces peptídicos). El término "polipéptido" se refiere a cualquier cadena o cadenas de dos o más

aminoácidos y no se refiere a una longitud específica del producto. Por tanto, péptidos, dipéptidos, tripéptidos, oligopéptidos, "proteína", "cadena de aminoácidos" o cualquier otro término utilizado para referirse a una cadena o cadenas de dos o más aminoácidos, se incluyen dentro de la definición de "polipéptido", y el término "polipéptido" se puede utilizar en lugar de, o de manera indistinta con cualquiera de estos términos. También se pretende que el término "polipéptido" haga referencia a los productos de modificaciones posteriores a la expresión del polipéptido, incluyendo, sin limitación, glicosilación, acetilación, fosforilación, amidación, derivatización por grupos protectores/bloqueantes conocidos, escisión proteolítica o modificación por aminoácidos de origen no natural. Un polipéptido se puede obtener de una fuente biológica natural o se puede producir por tecnología recombinante, pero no necesariamente se traduce a partir de una secuencia de ácido nucleico designada. Se puede generar de cualquier manera, incluso mediante síntesis química.

[0140] El término "aminoácido" incluye alanina (Ala o A); arginina (Arg o R); asparragina (Asn o N); ácido aspártico (Asp o D); cisteína (Cys o C); glutamina (Gln o Q); ácido glutámico (Glu o E); glicina (Gly o G); histidina (His o H); isoleucina (Ile o I); leucina (Leu o L); lisina (Lys o K); metionina (Met o M); fenilalanina (Phe o F); prolina (Pro o P); serina (Ser o S); treonina (Thr o T); triptófano (Trp o W); tirosina (Tyr o Y); y valina (Val o V). Los aminoácidos no tradicionales también están dentro del alcance de la divulgación e incluyen norleucina, ornitina, norvalina, homoserina y otros análogos de restos de aminoácidos tales como los descritos en Ellman et al. Meth. Enzym. 202:301-336 (1991). Para generar tales restos de aminoácidos de origen no natural, se pueden utilizar los procedimientos de Noren et al. Science 244: 182 (1989) y Ellman et al., Supra. Brevemente, estos procedimientos implican activar químicamente un ARNt supresor con un resto de aminoácido de origen no natural seguido de transcripción y traducción in vitro del ARN. La introducción del aminoácido no tradicional también se puede lograr utilizando químicas de péptidos conocidas en la técnica. Como se emplea en la presente memoria, el término "aminoácido polar" incluye aminoácidos que tienen carga neta cero, pero que tienen cargas parciales distintas de cero en diferentes porciones de sus cadenas laterales (p. ej., M, F, W, S, Y, N, Q, C). Estos aminoácidos pueden participar en interacciones hidrófobas e interacciones electrostáticas. Como se emplea en la presente memoria, el término "aminoácido cargado" incluye aminoácidos que pueden tener una carga neta distinta de cero en sus cadenas laterales (p. ej., R, K, H, E, D). Estos aminoácidos pueden participar en interacciones hidrófobas e interacciones electrostáticas.

[0141] También se incluyen en la presente divulgación fragmentos o variantes de polipéptidos y cualquier combinación de los mismos. Los términos "fragmento" o "variante" cuando se refieren a dominios de unión de polipéptidos o moléculas de unión de la presente divulgación incluyen cualquier polipéptido que conserve al menos algunas de las propiedades (p. ej., afinidad de unión a FcRn para un

dominio de unión a FcRn o variante de Fc, actividad de coagulación para una variante de FVIII, o actividad de unión a FVIII para el fragmento VWF) del polipéptido de referencia. Los fragmentos de polipéptidos incluyen fragmentos proteolíticos, así como fragmentos de delección, además de fragmentos de anticuerpos específicos discutidos en otra parte de la presente memoria, pero no incluyen el polipéptido de longitud completa natural (o polipéptido maduro). Las variantes de dominios de unión a polipéptidos o moléculas de unión de la presente divulgación incluyen fragmentos como se describió anteriormente, y también polipéptidos con secuencias de aminoácidos alteradas debido a sustituciones, delecciones o inserciones de aminoácidos. Las variantes pueden ser de origen natural o no natural. Se pueden producir variantes de origen no natural utilizando técnicas de mutagénesis conocidas en la técnica. Los polipéptidos variantes pueden comprender sustituciones, delecciones o adiciones de aminoácidos conservativas o no conservativas.

[0142] Una "sustitución de aminoácidos conservativa" es aquella en la que el resto de aminoácido se reemplaza por un resto de aminoácido que tiene una cadena lateral similar. Se han definido en la técnica familias de restos de aminoácidos que tienen cadenas laterales similares, que incluyen cadenas laterales alcalinas (p. ej., lisina, arginina, histidina), cadenas laterales ácidas (p. ej., ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (p. ej., glicina, asparragina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (p. ej., alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (p. ej., treonina, valina, isoleucina) y cadenas laterales aromáticas (p. ej., tirosina, fenilalanina, triptófano, histidina). Por lo tanto, si un aminoácido en un polipéptido se reemplaza por otro aminoácido de la misma familia de cadenas laterales, la sustitución se considera conservativa. En otra realización, una serie de aminoácidos se puede reemplazar de forma conservativa por una serie estructuralmente similar que difiere en el orden y/o composición de los miembros de la familia de cadenas laterales.

[0143] El término "porcentaje de identidad", como se conoce en la técnica, es una relación entre dos o más secuencias polipeptídicas o dos o más secuencias polinucleotídicas, según se determina comparando las secuencias. En la técnica, "identidad" también significa el grado de relación de secuencia entre secuencias de polipéptidos o polinucleótidos, según sea el caso, según lo determinado por la coincidencia entre series de tales secuencias. La "identidad" se puede calcular fácilmente mediante métodos conocidos, incluidos, entre otros, los descritos en: *Computational Molecular Biology* (Lesk, A. Mm, ed.) Oxford University Press, Nueva York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, Nueva York (1993); *Computer Analysis of Secuencia Data, Part I* (Griffin, A. M., y Griffin, H. G., eds.) Humana Press, Nueva Jersey (1994); *Secuencia Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); y *Secuencia Analysis Primer* (Gribskov, M. y

Devereux, J., eds.) Stockton Press, New York (1991). Los métodos preferidos para determinar la identidad están diseñados para proporcionar la mejor coincidencia entre las secuencias sometidas a ensayo. Los métodos para determinar la identidad están codificados en programas informáticos disponibles públicamente. Se pueden realizar alineamientos de secuencia y cálculos de identidad porcentual utilizando soportes lógicos de análisis de secuencia tales como el programa Megalign del paquete de computación bioinformática LASERGENE (DNASTAR Inc., Madison, WI), el paquete de programas GCG (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403 (1990)) y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE.UU.). Dentro del contexto de esta solicitud, se entenderá que cuando se utilice un soporte lógico de análisis de secuencias para el análisis, los resultados del análisis se basarán en los "valores predeterminados" del programa referenciado, a menos que se especifique lo contrario. Como se emplea en la presente memoria, "valores predeterminados" significará cualquier conjunto de valores o parámetros que se cargan originalmente con el soporte lógico cuando se inicializan por primera vez. Con el fin de determinar el porcentaje de identidad entre una proteína terapéutica, p. ej., un factor de coagulación, la secuencia de la divulgación y una secuencia de referencia, solamente se utilizan los nucleótidos de la secuencia de referencia correspondientes a los nucleótidos de la proteína terapéutica, p. ej., el factor de coagulación, secuencia de la divulgación para calcular el porcentaje de identidad. Por ejemplo, cuando se compara una secuencia de nucleótidos de FVIII de longitud completa que contiene el dominio B con una secuencia de nucleótidos de FVIII con el dominio B suprimido optimizado (BDD) de la divulgación, la porción del alineamiento que incluye el dominio A1, A2, A3, C1 y C2 se utilizará para calcular el porcentaje de identidad. Los nucleótidos en la porción de la secuencia de FVIII de longitud completa que codifica el dominio B (que dará como resultado un gran "espacio" en el alineamiento) no se contarán como un emparejamiento erróneo. Además, al determinar el porcentaje de identidad entre una secuencia de FVIII BDD optimizada de la divulgación, o una porción designada de la misma (p. ej., nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3), y una secuencia de referencia, el porcentaje de identidad se calculará por alineamiento dividiendo el número de nucleótidos coincidentes por el número total de nucleótidos en la secuencia completa de la secuencia de FVIII BDD optimizada, o una porción designada de la misma, como se indica en la presente memoria.

[0144] Como se emplea en la presente memoria, los nucleótidos correspondientes a los nucleótidos de una secuencia particular de la divulgación se identifican mediante el alineamiento de la secuencia de la divulgación para maximizar la identidad con una secuencia de referencia. El número utilizado para identificar un aminoácido equivalente en una secuencia de referencia se basa en el número utilizado para identificar el aminoácido correspondiente en la secuencia de la divulgación.

[0145] Una proteína de "fusión" o "quimérica" comprende una primera secuencia de aminoácidos conectada a una segunda secuencia de aminoácidos con la que no está conectada naturalmente en la naturaleza. Las secuencias de aminoácidos que normalmente existen en proteínas separadas se pueden juntar en el polipéptido de fusión, o las secuencias de aminoácidos que normalmente existen en la misma proteína se pueden colocar en una nueva disposición en el polipéptido de fusión, p. ej., fusión de un dominio del factor VIII de la divulgación con un dominio Fc de Ig. Una proteína de fusión se crea, por ejemplo, mediante síntesis química, o creando y traduciendo un polinucleótido en el que las regiones peptídicas están codificadas en la relación deseada. Una proteína quimérica puede comprender adicionalmente una segunda secuencia de aminoácidos asociada con la primera secuencia de aminoácidos por un enlace no peptídico covalente, o no covalente.

[0146] Como se emplea en la presente memoria, el término "sitio de inserción" se refiere a una posición en un polipéptido, o fragmento, variante o obtenido del mismo, que está inmediatamente aguas arriba de la posición en la que se puede insertar un radical heterólogo. Un "sitio de inserción" se especifica como un número, siendo el número el número del aminoácido en una secuencia de referencia. Por ejemplo, un "sitio de inserción" en FVIII se refiere al número de la secuencia de aminoácidos en el FVIII nativo maduro (SEQ ID NO: 15) al que corresponde el sitio de inserción, que es inmediatamente N-terminal a la posición de la inserción. Por ejemplo, la frase "a3 comprende un radical heterólogo en un sitio de inserción que corresponde al aminoácido 1656 de SEQ ID NO: 15", indica que el radical heterólogo está ubicado entre dos aminoácidos correspondientes al aminoácido 1656 y al aminoácido 1657 de SEQ ID NO: 15.

[0147] La frase "inmediatamente aguas abajo de un aminoácido", como se emplea en la presente memoria, se refiere a la posición justo al lado del grupo carboxilo terminal del aminoácido. De manera similar, la frase "inmediatamente aguas arriba de un aminoácido" se refiere a la posición justo al lado del grupo amino terminal del aminoácido.

[0148] Los términos "insertado", "se inserta", "insertado en" o términos gramaticalmente relacionados, como se emplean en la presente memoria, se refieren a la posición de un radical heterólogo en un polipéptido, p. ej., un factor de coagulación, en relación con la posición análoga el polipéptido parental. Por ejemplo, en cierta realización, "insertado" y similares se refieren a la posición de un radical heterólogo en un polipéptido de FVIII recombinante, en relación con la posición análoga en el FVIII humano maduro nativo. Como se emplean en la presente memoria, los términos se refieren a las características del polipéptido y no indican, implican o infieren ningún método o procedimiento mediante el cual se elaboró el polipéptido.

[0149] Como se emplea en la presente memoria, el término "semivida" se refiere a una semivida biológica de un polipéptido particular *in vivo*. La semivida se puede representar por el tiempo necesario para que la mitad de la cantidad administrada a un sujeto se elimine de la circulación y/o de otros tejidos del animal. Cuando se construye una curva de aclaramiento de un polipéptido dado en función del tiempo, la curva suele ser bifásica con una fase α rápida y una fase β más larga. La fase α representa típicamente un equilibrio del polipéptido Fc administrado entre el espacio intravascular y extravascular y está determinada, en parte, por el tamaño del polipéptido. La fase β representa típicamente el catabolismo del polipéptido en el espacio intravascular. En algunas realizaciones, la proteína terapéutica, p. ej., el factor de coagulación, p. ej., FVIII, y las proteínas quiméricas que las comprenden son monofásicas y, por lo tanto, no tienen una fase alfa, sino solo una fase beta. Por lo tanto, en ciertas realizaciones, el término semivida como se emplea en la presente memoria se refiere a la semivida del polipéptido en la fase β .

[0150] El término "conectado" como se emplea en la presente memoria se refiere a una primera secuencia de aminoácidos o secuencia de nucleótidos unida covalente o no covalentemente a una segunda secuencia de aminoácidos o secuencia de nucleótidos, respectivamente. La primera secuencia de aminoácidos o nucleótidos se puede unir directamente o yuxtaponer a la segunda secuencia de aminoácidos o nucleótidos o, alternativamente, una secuencia intermedia puede unir covalentemente la primera secuencia a la segunda secuencia. El término "conectado" significa no solo una fusión de una primera secuencia de aminoácidos a una segunda secuencia de aminoácidos en el extremo C-terminal o en el extremo N-terminal, sino que también incluye la inserción de la primera secuencia de aminoácidos completa (o el segundo aminoácido secuencia) en dos aminoácidos cualesquiera en la segunda secuencia de aminoácidos (o la primera secuencia de aminoácidos, respectivamente). En una realización, la primera secuencia de aminoácidos se puede conectar a una segunda secuencia de aminoácidos mediante un enlace peptídico o un enlazador. La primera secuencia de nucleótidos se puede unir a una segunda secuencia de nucleótidos mediante un enlace fosfodiéster o un enlazador. El conector puede ser un péptido o polipéptido (para cadenas polipeptídicas) o un nucleótido o una cadena nucleotídica (para cadenas nucleotídicas) o cualquier radical químico (para cadenas polipeptídicas y polinucleotídicas). El término "conectado" también se indica mediante un guión (-).

[0151] Hemostasia, como se emplea en la presente memoria, significa detener o ralentizar el sangrado o hemorragia; o detener o ralentizar el flujo sanguíneo a través de un vaso sanguíneo o parte del organismo.

[0152] Trastorno hemostático, como se emplea en la presente memoria, significa una afección genéticamente heredada o adquirida caracterizada por una tendencia a la hemorragia, ya sea espontáneamente o como resultado de un trauma, debido a una capacidad alterada o a una incapacidad

para formar un coágulo de fibrina. Los ejemplos de tales trastornos incluyen las hemofilias. Las tres formas principales son la hemofilia A (deficiencia de factor VIII), hemofilia B (deficiencia de factor IX o "enfermedad de Christmas") y hemofilia C (deficiencia de factor XI, tendencia leve a hemorragias).

Otros trastornos hemostáticos incluyen, p. ej., enfermedad de von Willebrand, deficiencia de factor XI (deficiencia de PTA), deficiencia de factor XII, deficiencias o anomalías estructurales en fibrinógeno, protrombina, factor V, factor VII, factor X o factor XIII, síndrome de Bernard-Soulier, que es un defecto o deficiencia de GPIb. GPIb, el receptor de vWF, puede ser defectuoso y provocar la falta de formación de coágulos primarios (hemostasia primaria) y el aumento de la tendencia al sangrado) y la trombastenia de Glanzman y Naegeli (trombastenia de Glanzmann). En la insuficiencia hepática (formas aguda y crónica), hay una producción insuficiente de factores de coagulación por parte del hígado; esto puede aumentar el riesgo de sangrado.

[0153] Las moléculas de ácido nucleico aisladas, los polipéptidos aislados o los vectores que comprenden la molécula de ácido nucleico aislada de la divulgación se pueden utilizar de forma profiláctica. Como se emplea en la presente memoria, el término "tratamiento profiláctico" se refiere a la administración de una molécula antes de un episodio de sangrado. En una realización, el sujeto que necesita un agente hemostático general se está sometiendo a cirugía o está a punto de someterse a ella. Se puede administrar un polinucleótido, polipéptido o vector de la divulgación antes o después de la cirugía como agente profiláctico. El polinucleótido, polipéptido o vector de la divulgación se puede administrar durante o después de la cirugía para controlar un episodio de sangrado agudo. La cirugía puede incluir, pero no se limita a, trasplante de hígado, resección de hígado, procedimientos dentales o trasplante de células madre.

[0154] Las moléculas de ácido nucleico aisladas, polipéptidos aislados o vectores de la divulgación también se utilizan para el tratamiento a demanda. El término "tratamiento a demanda" se refiere a la administración de una molécula de ácido nucleico aislada, un polipéptido aislado o un vector en respuesta a los síntomas de un episodio de sangrado o antes de una actividad que pueda causar sangrado. En un aspecto, el tratamiento a demanda se puede administrar a un sujeto cuando comienza el sangrado, tal como después de una lesión, o cuando se espera sangrado, tal como antes de una cirugía. En otro aspecto, el tratamiento a demanda se puede administrar antes de actividades que aumentan el riesgo de sangrado, tales como los deportes de contacto.

[0155] Como se emplea en la presente memoria, el término "sangrado agudo" se refiere a un episodio de sangrado independientemente de la causa subyacente. Por ejemplo, un sujeto puede tener traumatismo, uremia, un trastorno de sangrado hereditario (p. ej., deficiencia de factor VII), un trastorno plaquetario o resistencia debida al desarrollo de anticuerpos contra factores de coagulación.

[0156] Tratar, tratamiento, tratando, como se emplean en la presente memoria, se refieren, por ejemplo, a la reducción de la gravedad de una enfermedad o afección; la reducción de la duración del curso de una enfermedad; la mejora de uno o más síntomas asociados con una enfermedad o afección; la provisión de efectos beneficiosos a un sujeto con una enfermedad o afección, sin necesariamente curar la enfermedad o afección, o la profilaxis de uno o más síntomas asociados con una enfermedad o afección. En una realización, el término "tratando" o "tratamiento" significan mantener, p. ej., un nivel mínimo de FVIII de al menos aproximadamente 1 UI/dL, 2 UI/dL, 3 UI/dL, 4 UI/dL, 5 UI/dL, 6 UI/dL, 7 UI/dL, 8 UI/dL, 9 UI/dL, 10 UI/dL, 11 UI/dL, 12 UI/dL, 13 UI/dL, 14 UI/dL, 15 UI/dL, 16 UI/dL, 17 UI/dL, 18 UI/dL, 19 UI/dL, 20 UI/dL, 25 UI/dL, 30 UI/dL, 35 UI/dL, 40 UI/dL, 45 UI/dL, 50 UI/dL, 55 UI/dL, 60 UI/dL, 65 UI/dL, 70 UI/dL, 75 UI/dL, 80 UI/dL, 85 UI/dL, 90 UI/dL, 95 UI/dL, 100 UI/dL, 105 UI/dL, 110 UI/dL, 115 UI/dL, 120 UI/dL, 125 UI/dL, 130 UI/dL, 135 UI/dL, 140 UI/dL, 145 UI/dL o 150 UI/dL en un sujeto mediante la administración de una molécula de ácido nucleico aislada, un polipéptido aislado o un vector de la divulgación. En otra realización, tratando o tratamiento significa mantener un nivel mínimo de FVIII entre aproximadamente 1 y aproximadamente 150 UI/dL, aproximadamente 1 y aproximadamente 125 UI/dL, aproximadamente 1 y aproximadamente 100 UI/dL, aproximadamente 1 y aproximadamente 90 UI/dL, aproximadamente 1 y aproximadamente 85 UI/dL, aproximadamente 1 y aproximadamente 80 UI/dL, aproximadamente 1 y aproximadamente 75 UI/dL, aproximadamente 1 y aproximadamente 70 UI/dL, aproximadamente 1 y aproximadamente 65 UI/dL, aproximadamente 1 y aproximadamente 60 UI/dL, aproximadamente 1 y aproximadamente 55 UI/dL, aproximadamente 1 y aproximadamente 50 UI/dL, aproximadamente 1 y aproximadamente 45 UI/dL, aproximadamente 1 y aproximadamente 40 UI/dL, aproximadamente 1 y aproximadamente 35 UI/dL, aproximadamente 1 y aproximadamente 30 UI/dL, aproximadamente 1 y aproximadamente 25 UI/dL, aproximadamente 25 y aproximadamente 125 UI/dL, aproximadamente 50 y aproximadamente 100 UI/dL, aproximadamente 50 y aproximadamente 75 UI/dL, aproximadamente 75 y aproximadamente 100 UI/dL, aproximadamente 1 y aproximadamente 20 UI/dL, aproximadamente 2 y aproximadamente 20 UI/dL, aproximadamente 3 y aproximadamente 20 UI/dL, aproximadamente 4 y aproximadamente 20 UI/dL, aproximadamente 5 y aproximadamente 20 UI/dL, aproximadamente 6 y aproximadamente 20 UI/dL, aproximadamente 7 y aproximadamente 20 UI/dL, aproximadamente 8 y aproximadamente 20 UI/dL, aproximadamente 9 y aproximadamente 20 UI/dL, o aproximadamente 10 y aproximadamente 20 UI/dL. El tratamiento de o tratar una enfermedad o afección también pueden incluir el mantenimiento de la actividad de FVIII en un sujeto a un nivel comparable con al menos aproximadamente 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9. %, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, 125%, 130%, 135%, 140%, 145%, o 150% de la actividad de FVIII en un sujeto no hemofílico. El nivel mínimo requerido

para el tratamiento se puede medir mediante uno o más métodos conocidos y se puede ajustar (aumentar o disminuir) para cada persona.

[0157] "Administrar", como se emplea en la presente memoria, significa proporcionar una molécula de ácido nucleico farmacéuticamente aceptable, un polipéptido expresado a partir del mismo o un vector que comprende la molécula de ácido nucleico de la divulgación a un sujeto mediante una vía farmacéuticamente aceptable. Las vías de administración pueden ser intravenosas, p. ej., inyección intravenosa e infusión intravenosa. Las vías de administración adicionales incluyen, p. ej., administración subcutánea, intramuscular, oral, nasal y pulmonar. Las moléculas de ácido nucleico, polipéptidos y vectores se pueden administrar como parte de una composición farmacéutica que comprende al menos un excipiente.

[0158] El término "farmacéuticamente aceptable" como se emplea en la presente memoria se refiere a entidades moleculares y composiciones que son fisiológicamente tolerables y no producen típicamente toxicidad o una reacción adversa alérgica o similar, tal como malestar gástrico, mareos y similares, cuando se administran a un ser humano. Opcionalmente, como se emplea en la presente memoria, el término "farmacéuticamente aceptable" significa aprobado por una agencia reguladora del gobierno federal o estatal o enumerado en la Farmacopea de Estados Unidos u otra farmacopea generalmente reconocida para uso en animales, y más particularmente en seres humanos.

[0159] Como se emplea en la presente memoria, la frase "sujeto que lo necesita" incluye sujetos, tales como sujetos mamíferos, que se beneficiarían de la administración de una molécula de ácido nucleico, polipéptido o vector de la divulgación, p. ej., para mejorar la hemostasia. En una realización, los sujetos incluyen, pero no se limitan a, individuos con hemofilia. En otra realización, los sujetos incluyen, pero no se limitan a, individuos que han desarrollado un inhibidor de la proteína terapéutica, p. ej., el factor de coagulación, p. ej., FVIII, y por tanto necesitan una terapia de desvío o puente ("bypass"). El sujeto puede ser un adulto o un menor de edad (p. ej., menor de 12 años).

[0160] Como se emplea en la presente memoria, el término "proteína terapéutica" se refiere a cualquier polipéptido conocido en la técnica que se puede administrar a un sujeto. En algunas realizaciones, la proteína terapéutica comprende una proteína seleccionada entre un factor de coagulación, un factor de crecimiento, un anticuerpo, un fragmento funcional del mismo o una combinación de los mismos. Como se emplea en la presente memoria, el término "factor de coagulación" se refiere a moléculas, o análogos de las mismas, de origen natural o producidas de manera recombinante que previenen o disminuyen la duración de un episodio de sangrado en un sujeto. En otras palabras, significa moléculas que tienen actividad pro-coagulante, es decir, son responsables de la conversión de fibrinógeno en una malla de

fibrina insoluble que hace que la sangre se coagule o forme coágulos. "Factor de coagulación" como se emplea en la presente memoria incluye un factor de coagulación activado, su zimógeno, o un factor de coagulación activable. Un "factor de coagulación activable" es un factor de coagulación en una forma inactiva (p. ej., en su forma de zimógeno) que es capaz de convertirse en una forma activa. El término "factor de coagulación" incluye pero no se limita a factor I (FI), factor II (FII), factor V (FV), FVII, FVIII, FIX, factor X (FX), factor XI (FXI), factor XII. (FXII), factor XIII (FXIII), factor de von Willebrand (VWF), precalicreína, cininógeno de alto peso molecular, fibronectina, antitrombina III, cofactor II de heparina, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionado con la proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador del plasminógeno tisular (tPA), uroquinasa, inhibidor del activador del plasminógeno-1 (PAI-1), inhibidor del activador del plasminógeno-2 (PAI2), zimógenos de los mismos, formas activadas de los mismos o cualquier combinación de los mismos .

[0161] La actividad coagulación, como se emplea en la presente memoria, significa la capacidad de participar en una cascada de reacciones bioquímicas que culmina en la formación de un coágulo de fibrina y/o reduce la gravedad, duración o frecuencia de un episodio de hemorragia o sangrado.

[0162] Un "factor de crecimiento", como se emplea en la presente memoria, incluye cualquier factor de crecimiento conocido en la técnica, incluidas las citocinas y hormonas. En algunas realizaciones, el factor de crecimiento se selecciona entre adrenomedulina (AM), angiopoyetina (Ang), factor de motilidad autocrina, una proteína morfogenética ósea (BMP) (p. ej., BMP2, BMP4, BMP5, BMP7), un miembro de la familia del factor neurotrófico ciliar (p. ej., , factor neurotrófico ciliar (CNTF), factor inhibidor de leucemia (LIF), interleucina-6 (IL-6)), factor estimulante de colonias (p. ej., factor estimulante de colonias de macrófagos (m-CSF), factor estimulante de colonias de granulocitos (G-CSF), factor estimulante de colonias de granulocitos macrófagos (GM-CSF)), un factor de crecimiento epidérmico (EGF), una efrina (p. ej., efrina A1, efrina A2, efrina A3, efrina A4, efrina A5, efrina B1, efrina B2, efrina B3), eritropoyetina (EPO), un factor de crecimiento de fibroblastos (FGF) (p. ej., FGF1, FGF2, FGF3, FGF4 , FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF18, FGF19, FGF20, FGF21, FGF22, FGF23) somatotropina fetal bovina (FBS), un miembro de la familia de GDNF (p. ej., factor neurotrófico obtenido de la línea de células gliales (GDNF), neurturina, persefina, artemina), factor de diferenciación del crecimiento-9 (GDF9), factor de crecimiento de hepatocitos (HGF), factor de crecimiento obtenido de hepatoma (HDGF), insulina, un factor de crecimiento de tipo insulínico (p. ej., factor de crecimiento de tipo insulínico 1 (IGF-1) o IGF-2, una interleucina (IL) (p. ej., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL -7), factor de crecimiento de queratinocitos (KGF), factor de estimulación de la migración (MSF), proteína estimulante de macrófagos

(MSP o proteína de tipo factor de crecimiento de hepatocitos (HGFLP)), miostatina (GDF-8), una neurregulina (p. ej., Neurregulina 1 (NRG1), NRG2, NRG3, NRG4), una neurotrofina (p. ej., factor neurotrófico obtenido del cerebro (BDNF), factor de crecimiento nervioso (NGF), una neurotrofina-3 (NT-3), NT-4, factor de crecimiento placentario (PGF), factor de crecimiento obtenido de plaquetas (PDGF), renalasa (RNLS), factor de crecimiento de células T (TCGF), trombopoyetina (TPO), un factor de crecimiento transformante (p. ej., factor de crecimiento transformante alfa (TGF- α), TGF- β , factor de necrosis tumoral alfa (TNF- α) y factor de crecimiento endotelial vascular (VEGF).

[0163] En algunas realizaciones, la proteína terapéutica está codificada por un gen seleccionado entre distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RS1, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida) o cualquier combinación de los mismos.

[0164] Como se emplean en la presente memoria, los términos "heterólogo" o "exógeno" se refieren a moléculas que normalmente no se encuentran en un contexto dado, p. ej., en una célula o en un polipéptido. Por ejemplo, una molécula exógena o heteróloga se puede introducir en una célula y solo está presente después de la manipulación de la célula, p. ej., mediante transfección u otras formas de ingeniería genética o una secuencia de aminoácidos heteróloga puede estar presente en una proteína en la que se encuentra presente de forma natural.

[0165] Como se emplea en la presente memoria, el término "secuencia de nucleótidos heteróloga" se refiere a una secuencia de nucleótidos que no se encuentra naturalmente con una secuencia de polinucleótidos dada. En una realización, la secuencia de nucleótidos heteróloga codifica un polipéptido capaz de prolongar la semivida de la proteína terapéutica, p. ej., el factor de coagulación, p. ej., FVIII. En otra realización, la secuencia de nucleótidos heteróloga codifica un polipéptido que aumenta el radio hidrodinámico de la proteína terapéutica, p. ej., el factor de coagulación, p. ej., FVIII. En otras realizaciones, la secuencia de nucleótidos heteróloga codifica un polipéptido que mejora una o más propiedades farmacocinéticas de la proteína terapéutica sin afectar significativamente a su actividad o función biológica (p. ej., una actividad procoagulante). En algunas realizaciones, la proteína terapéutica está unida o conectada al polipéptido codificado por la secuencia de nucleótidos heteróloga mediante un conector. Los ejemplos no limitantes de radicales polipeptídicos codificados por secuencias de nucleótidos heterólogas incluyen una región constante de inmunoglobulina o una porción de la misma, albúmina o un fragmento de la misma, un radical que se une a la albúmina, una transferrina, los polipéptidos PAS de la Solicitud de Patente de Estados Unidos Núm. 20100292130, una secuencia HAP, transferrina o un fragmento de la misma, el péptido C-terminal (CTP) de la subunidad β de la

gonadotropina coriónica humana, molécula pequeña de unión a albúmina, una secuencia XTEN, radicales de unión a FcRn (p. ej., regiones Fc completas o porciones de las mismas que se unen a FcRn), regiones Fc de cadena sencilla (regiones ScFc, p. ej., como se describe en los documentos US 2008/0260738, WO 2008/012543 o WO 2008/1439545), conectores de poliglicina, conectores de poliserina, péptidos y polipéptidos cortos de 6-40 aminoácidos de dos tipos de aminoácidos seleccionados entre glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P) con diversos grados de estructura secundaria desde menos de 50% hasta más de 50%, entre otros, o dos o más combinaciones de los mismos. En algunas realizaciones, el polipéptido codificado por la secuencia de nucleótidos heteróloga está conectado a un radical no polipeptídico. Los ejemplos no limitantes de radicales no polipeptídicos incluyen polietilenglicol (PEG), moléculas pequeñas que se unen a albúmina, ácido polisiálico, hidroxietil almidón (HES), un obtenido del mismo, o cualquier combinación de los mismos.

[0166] Como se emplea en la presente memoria, el término "región Fc" se define como la porción de un polipéptido que corresponde a la región Fc de la Ig nativa, es decir, formada por la asociación dimérica de los respectivos dominios Fc de sus dos cadenas pesadas. Una región Fc nativa forma un homodímero con otra región Fc. Por el contrario, el término "región Fc fusionada genéticamente" o "región Fc monocatenaria" (región scFc), como se emplea en la presente memoria, se refiere a una región Fc dimérica sintética que comprende dominios Fc genéticamente conectados dentro de una única cadena polipeptídica (es decir, codificado en una única secuencia genética contigua).

[0167] En una realización, la "región Fc" se refiere a la porción de una única cadena pesada de Ig que comienza en la región bisagra justo aguas arriba del sitio de escisión de papaína (p. ej., resto 216 en IgG, tomando el primer resto de de la región constante de la cadena pesada como 114) y termina en el extremo C-terminal del anticuerpo. Por consiguiente, un dominio Fc completo comprende al menos un dominio bisagra, un dominio CH2 y un dominio CH3.

[0168] La región Fc de una región constante de Ig, dependiendo del isotipo de Ig, puede incluir los dominios CH2, CH3 y CH4, así como la región bisagra. Las proteínas quiméricas que comprenden una región Fc de una Ig confieren varias propiedades deseables a una proteína quimérica, incluida una mayor estabilidad, una mayor semivida en suero (véase Capon et al., 1989, *Nature* 337: 525) así como la unión a receptores Fc tales como el receptor de Fc neonatal (FcRn) (Patentes de Estados Unidos Núm. 6.086.875, 6.485.726, 6,030,613; documentos WO 03/077834; US2003-0235536A1), que se incorporan a la presente memoria como referencia en su totalidad.

[0169] Una "secuencia de nucleótidos de referencia", cuando se emplea en la presente memoria como una comparación con una secuencia de nucleótidos de la divulgación, es una secuencia de polinucleótidos esencialmente idéntica a la secuencia de nucleótidos de la divulgación, excepto que la secuencia no está optimizada. Por ejemplo, la secuencia de nucleótidos de referencia para una molécula de ácido nucleico que consiste en el codón optimizado FVIII BDD de SEQ ID NO: 1 y una secuencia de nucleótidos heteróloga que codifica una región Fc de cadena sencilla unida a SEQ ID NO: 1 en su extremo 3' es una molécula de ácido nucleico que consiste en el FVIII BDD original (o "parental") de SEQ ID NO: 16 y la secuencia de nucleótidos heteróloga idéntica que codifica una región Fc de cadena sencilla conectada a SEQ ID NO: 16 en su extremo 3'.

[0170] Como se emplea en la presente memoria, el término "optimizado", con respecto a las secuencias de nucleótidos, se refiere a una secuencia de polinucleótidos que codifica un polipéptido, en donde la secuencia de polinucleótidos se ha mutado para mejorar una propiedad de esa secuencia de polinucleótidos. En algunas realizaciones, la optimización se realiza para aumentar los niveles de transcripción, aumentar los niveles de traducción, aumentar los niveles de ARNm en estado estacionario, aumentar o disminuir la unión de proteínas reguladoras tales como los factores de transcripción generales, aumentar o disminuir el corte y empalme o aumentar el rendimiento del polipéptido producido por la secuencia de polinucleótidos. Los ejemplos de cambios que se pueden realizar en una secuencia de polinucleótidos para optimizarla incluyen optimización de codones, optimización del contenido de G/C, eliminación de secuencias repetidas, eliminación de elementos ricos en AT, eliminación de sitios de corte y empalme crípticos, eliminación de elementos que actúan en cis que reprimen la transcripción o la traducción, adición o eliminación de secuencias de poli-T o poli-A, adición de secuencias entorno al sitio de inicio de la transcripción que aumentan la transcripción, tales como secuencias consenso de Kozak, eliminación de secuencias que podrían formar estructuras de tallo bucle, eliminación de secuencias desestabilizadoras, y dos o más combinaciones de los mismos.

II. Moléculas de Ácido Nucleico

[0171] La presente divulgación está dirigida a una molécula de ácido nucleico sin cápside, de tipo plásmido, que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica o un gen que puede modular la expresión de una proteína diana, p. ej., un miARN. . Una cápside, la cubierta proteica de un virus, encierra el material genético del virus. Se sabe que las cápsides ayudan a las funciones del virión protegiendo el genoma viral, suministrando el genoma a un anfitrión e interactuando con el anfitrión. No obstante, las cápsides virales pueden ser un factor que limite la capacidad de empaquetamiento de los vectores y/o induzca respuestas inmunitarias, especialmente cuando se utilizan en terapia génica.

[0172] Los vectores de AAV han surgido como uno de los tipos más comunes de vectores de terapia génica. Sin embargo, la presencia de la cápside limita la utilidad de un vector de AAV en terapia génica. En particular, la cápside en sí misma puede limitar el tamaño del transgén que se incluye en el vector a tan solo menos de 4,5 kb. Varias proteínas terapéuticas que pueden ser útiles en una terapia génica pueden superar fácilmente este tamaño incluso antes de que se añadan elementos reguladores.

[0173] Además, las proteínas que forman la cápside pueden servir como antígenos que pueden ser elegidos como diana por el sistema inmunológico de un sujeto. El AAV es muy común en la población general, y la mayoría de las personas han estado expuestas a un AAV a lo largo de toda su vida. Como resultado, la mayoría de los posibles receptores de terapia génica probablemente ya hayan desarrollado una respuesta inmunitaria a un AAV y, por lo tanto, es más probable que rechacen la terapia.

[0174] Ciertos aspectos de la presente divulgación tienen como objetivo superar estas deficiencias de los vectores de AAV. En particular, ciertos aspectos de la presente divulgación están dirigidos a una molécula de ácido nucleico, que comprende una primera ITR, una segunda ITR y un casete genético, p. ej., que codifica una proteína terapéutica y/o un miARN. En algunas realizaciones, la primera ITR y la segunda ITR flanquean un casete genético que comprende una la secuencia de polinucleótidos heteróloga. En algunas realizaciones, la molécula de ácido nucleico no comprende un gen que codifique una proteína de la cápside, una proteína de replicación y/o una proteína de ensamblaje. En algunas realizaciones, el casete genético codifica una proteína terapéutica. En algunas realizaciones, la proteína terapéutica comprende un factor de coagulación. En algunas realizaciones, el casete genético codifica un miARN. En ciertas realizaciones, el casete genético se coloca entre la primera ITR y la segunda ITR. En algunas realizaciones, la molécula de ácido nucleico comprende adicionalmente una o más regiones no codificantes. En determinadas realizaciones, las una o más regiones no codificantes comprenden una secuencia promotora, un intrón, un elemento regulador post-transcripcional, una secuencia de poli (A) 3' UTR o cualquier combinación de los mismos.

[0175] En una realización, el casete genético es un ácido nucleico de hebra sencilla. En otra realización, el casete genético es un ácido nucleico de doble hebra.

[0176] En una realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV (p. ej., una ITR de B19 o GPV);
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;

- (d) un nucleótido que codifica un miARN o una proteína terapéutica, p. ej., un factor de coagulación;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA;
- (g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV (p. ej., una ITR de B19 o GPV).

[0177] En una realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV,
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) un nucleótido que codifica un miARN, en donde el miARN regula por disminución la expresión de un gen diana seleccionado entre SOD1, HTT, RHO y cualquier combinación de los mismos;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA;
- (g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV

[0178] En una realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV,
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) un nucleótido que codifica distrofina ligada al cromosoma X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RS1, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida) o cualquier combinación de los mismos;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA;

- (g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* que no es AAV

[0179] En una realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR que es una ITR de un AAV, por ejemplo, un genoma de serotipo 2 de AAV;
- (b) una secuencia promotora específica de tejido, por ejemplo, promotor TTP;
- (c) un intrón, por ejemplo, un intrón sintético;
- (d) un nucleótido que codifica FVIII; en el que el nucleótido tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con una secuencia de nucleótidos seleccionada entre SEQ ID NO: 1-14 o SEQ ID NO: 71, en donde el FVIII codificado por el nucleótido conserva una actividad de FVIII;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y
- 10 (g) una segunda ITR que es una ITR de un AAV, p. ej., un genoma de serotipo 2 de AAV.

[0180] En una realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR que es una ITR de un AAV, p. ej., un genoma de serotipo 2 de AAV;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- 15 (d) un nucleótido que codifica un miARN, en donde el miARN regula por disminución la expresión de un gen diana, p. ej., SOD1, HTT, RHO y cualquier combinación de los mismos;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y
- (g) una segunda ITR que es una ITR de un AAV, p. ej., un genoma de serotipo 2 de AAV.

[0181] En una realización, la molécula de ácido nucleico comprende:

- 20 (a) una primera ITR que es una ITR de un AAV, por ejemplo, un genoma de serotipo 2 de AAV;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) un nucleótido que codifica distrofina ligada al cromosoma X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RS1, CFH, HTRA, ARMS,
- 25 CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida) o cualquier combinación de los mismos;

- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y
- (g) una segunda ITR que es una ITR de un AAV, p. ej., un genoma de serotipo 2 de AAV.

[0182] En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR;
- 5 (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) un nucleótido que codifica un miARN o una proteína terapéutica, p. ej., factor de coagulación;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y
- 10 (g) una segunda ITR,

en donde una de la primera ITR o la segunda ITR es una ITR de un miembro de la familia *Parvoviridae* que no es AAV y la otra ITR es una ITR de un AAV, p. ej., un genoma de serotipo 2 de AAV.

[0183] En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que lleva la secuencia de ITR 5' de AAV2 expuesta en SEQ ID NO: 111;
- 15 (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y/o
- 20 (g) una ITR 3' que lleva la secuencia ITR 3' de AAV2 expuesta en SEQ ID NO: 124.

[0184] En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que lleva la secuencia de ITR 5' de AAV2 expuesta en SEQ ID NO: 111;
- (b) una secuencia promotora específica de tejido, p. ej., promotor CAG;
- (c) un intrón, p. ej., un intrón sintético;

- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli (A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que lleva la secuencia ITR 3' de AAV2 expuesta en SEQ ID NO: 193.

5 **[0185]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR;
- (b) una secuencia promotora específica de tejido, promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) un nucleótido que codifica un miARN o una proteína terapéutica, p. ej., factor de coagulación;

- 10 (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y
- (g) una segunda ITR,

en donde la primera ITR es una ITR sintética, la segunda ITR es una ITR sintética, o tanto la primera ITR como la segunda ITR son ITR sintéticas.

15 **[0186]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR de B19;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica seleccionada
- 20 del grupo que consiste en un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo y una combinación de los mismos;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una segunda ITR de B19.

25 **[0187]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR de GPV;
 - (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
 - (c) un intrón, p. ej., un intrón sintético;
 - (d) una la secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica seleccionada
- 5 del grupo que consiste en un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo y una combinación de los mismos
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
 - (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
 - (g) una segunda ITR de GPV.
- 10 **[0188]** En otra realización, la molécula de ácido nucleico comprende:
- (a) una primera ITR de B19;
 - (b) una secuencia promotora ubicua, p. ej., promotor CAG;
 - (c) un intrón, p. ej., un intrón sintético;
 - (d) una la secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica seleccionada
- 15 del grupo que consiste en un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo y una combinación de los mismos
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
 - (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
 - (g) una segunda ITR de B19.
- 20 **[0189]** En otra realización, la molécula de ácido nucleico comprende:
- (a) una primera ITR de GPV;
 - (b) una secuencia promotora ubicua, p. ej., promotor CAG;
 - (c) un intrón, p. ej., un intrón sintético;
 - (d) una la secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica seleccionada del
- 25 grupo que consiste en un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento del mismo y una combinación de los mismos

- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una segunda ITR de GPV.

[0190] En otra realización, la molécula de ácido nucleico comprende:

- 5 (a) una primera ITR de B19;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica fenilalanina hidroxilasa (PAH);
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- 10 (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una segunda ITR de B19.

[0191] En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR de GPV;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- 15 (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica fenilalanina hidroxilasa (PAH);
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una segunda ITR de GPV.

20 **[0192]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que porta la secuencia ITR 5' de B19d135 expuesta en SEQ ID NO: 180;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;

- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia ITR 3' de B19d135 expuesta en SEQ ID NO: 181.

[0193] En otra realización, la molécula de ácido nucleico comprende:

- 5 (a) una ITR 5' que porta la secuencia de ITR 5' de GPVd162 expuesta en SEQ ID NO: 183;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- 10 (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de GPVd162 expuesta en SEQ ID NO: 184.

[0194] En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que porta la secuencia de ITR 5' de B19 completo expuesta en SEQ ID NO: 185;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- 15 (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de B19 completo expuesta en SEQ ID NO: 186.

20 **[0195]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que porta la secuencia de ITR 5' de GPV completo expuesta en SEQ ID NO: 187;
- (b) una secuencia promotora específica de tejido, p. ej., promotor TTP;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica FVIII, p. ej., FVIIIco6XTEN;

- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de GPV expuesta en SEQ ID NO: 188.

[0196] En otra realización, la molécula de ácido nucleico comprende:

- 5 (a) una ITR 5' que porta la secuencia de ITR 5' de B19d135 expuesta en SEQ ID NO: 180;
- (b) una secuencia promotora específica de tejido, p. ej., promotor CAG;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica PAH;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- 10 (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de B19d135 expuesta en SEQ ID NO: 181.

[0197] En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que porta la secuencia de ITR 5' de GPVd162 expuesta en SEQ ID NO: 183;
- (b) una secuencia promotora específica de tejido, p. ej., promotor CAG;
- 15 (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica PAH;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de GPVd162 expuesta en SEQ ID NO: 184.

20 **[0198]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una ITR 5' que porta la secuencia de ITR 5' de B19 completo expuesta en SEQ ID NO: 185;
- (b) una secuencia promotora específica de tejido, p. ej., promotor CAG;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica PAH;

- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de B19 completo expuesta en SEQ ID NO: 186.

[0199] En otra realización, la molécula de ácido nucleico comprende:

- 5 (a) una ITR 5' que porta la secuencia de ITR 5' de GPV completo expuesta en SEQ ID NO: 187;
- (b) una secuencia promotora específica de tejido, p. ej., promotor CAG;
- (c) un intrón, p. ej., un intrón sintético;
- (d) una la secuencia de polinucleótidos heteróloga que codifica PAH;
- (e) un elemento regulador post-transcripcional, p. ej., WPRE;
- 10 (f) una secuencia de cola de poli(A) 3' UTR, p. ej., bGHpA; y/o
- (g) una ITR 3' que porta la secuencia de ITR 3' de GPV expuesta en SEQ ID NO: 188.

A. Repeticiones Terminales Invertidas

- [0200]** Ciertos aspectos de la presente divulgación están dirigidos a una molécula de ácido nucleico que comprende una primera ITR, p. ej., una ITR 5', y una segunda ITR, p. ej., una ITR 3'. Típicamente, las
- 15 ITR están implicadas en la replicación y el rescate o escisión del ADN de parvovirus (p. ej., AAV) de plásmidos procarióticos (Samulski et al., 1983, 1987; Senapathy et al., 1984; Gottlieb y Muzyczka, 1988). Además, las ITR parecen ser las secuencias mínimas requeridas para la integración proviral de AAV y para el empaquetamiento del ADN de AAV en viriones (McLaughlin et al., 1988; Samulski et al., 1989). Estos elementos son esenciales para la multiplicación eficaz de un genoma de parvovirus. Se
- 20 plantea la hipótesis de que los elementos definatorios mínimos indispensables para la función de las ITR son un sitio de unión a Rep (p. ej., RBS; GCGCGCTCGCTCGCTC (SEQ ID NO: 104) para AAV2) y un sitio de resolución terminal (p. ej., TRS; AGTTGG (SEQ ID NO: 105) para AAV2) más una secuencia palindrómica variable que permite la formación de horquillas. Las regiones de nucleótidos palindrómicas normalmente funcionan juntas en *cis* como orígenes de la replicación del ADN y como señales de
- 25 empaquetamiento del virus. Las secuencias complementarias en las ITR se pliegan en una estructura de horquilla durante la replicación del ADN. En algunas realizaciones, las ITR se pliegan en una estructura en forma de horquilla en T. En otras realizaciones, las ITR se pliegan en estructuras de horquilla que no tienen forma de T, por ejemplo, en una estructura de horquilla en forma de U. Los datos sugieren que las estructuras de horquilla en forma de T de las ITR de AAV pueden inhibir la expresión de un transgén

flanqueado por las ITR. Véase, p. ej., Zhou et al., Scientific Reports 7: 5432 (14 de julio de 2017). Al utilizar una ITR que no forma estructuras de horquilla en forma de T, se puede evitar esta forma de inhibición. Por lo tanto, en ciertos aspectos, un polinucleótido que comprende una ITR que no es de AAV tiene una mejor expresión transgénica en comparación con un polinucleótido que comprende una ITR de AAV que forma una horquilla en forma de T.

[0201] En algunas realizaciones, la ITR comprende una ITR de origen natural, por ejemplo, la ITR comprende toda o una parte de una ITR de parvovirus. En algunas realizaciones, la ITR comprende una secuencia sintética. En una realización, la primera ITR o la segunda ITR comprenden una secuencia sintética. En otra realización, cada una de la primera ITR y la segunda ITR comprende una secuencia sintética. En algunas realizaciones, la primera ITR o la segunda ITR comprenden una secuencia de origen natural. En otra realización, cada uno de la primera ITR y la segunda ITR comprende una secuencia de origen natural.

[0202] En algunas realizaciones, la ITR comprende o consiste en una parte de una ITR de origen natural, p. ej., una ITR truncada. En algunas realizaciones, la ITR comprende o consiste en un fragmento de una ITR de origen natural, en donde el fragmento comprende al menos aproximadamente 5 nucleótidos, al menos aproximadamente 10 nucleótidos, al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 25 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 45 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 55 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 65 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 75 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 85 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 95 nucleótidos, al menos aproximadamente 100 nucleótidos, al menos aproximadamente 125 nucleótidos, al menos aproximadamente 150 nucleótidos, al menos aproximadamente 175 nucleótidos, al menos aproximadamente 200 nucleótidos, al menos aproximadamente 225 nucleótidos, al menos aproximadamente 250 nucleótidos, al menos aproximadamente 275 nucleótidos, al menos aproximadamente 300 nucleótidos, al menos aproximadamente 325 nucleótidos, al menos aproximadamente 350 nucleótidos, al menos aproximadamente 375 nucleótidos, al menos aproximadamente 400 nucleótidos, al menos aproximadamente 425 nucleótidos, al menos aproximadamente 450 nucleótidos, al menos aproximadamente 475 nucleótidos, al menos aproximadamente 500 nucleótidos, al menos aproximadamente 525 nucleótidos, al menos

aproximadamente 550 nucleótidos, al menos aproximadamente 575 nucleótidos o al menos aproximadamente 600 nucleótidos; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En ciertas realizaciones, la ITR comprende o consiste en un fragmento de una ITR de origen natural, en donde el fragmento comprende al menos aproximadamente 129 nucleótidos; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En determinadas realizaciones, la ITR comprende o consiste en un fragmento de una ITR de origen natural, en donde el fragmento comprende al menos aproximadamente 102 nucleótidos; en donde la ITR conserva una propiedad funcional de la ITR de origen natural.

[0203] En algunas realizaciones, la ITR comprende o consiste en una porción de una ITR de origen natural, en donde el fragmento comprende al menos aproximadamente 5%, al menos aproximadamente 10%, al menos aproximadamente 15%, al menos aproximadamente 20%, al menos aproximadamente 25%, al menos aproximadamente 30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 45%, al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de la longitud de la ITR de origen natural; en donde el fragmento conserva una propiedad funcional de la ITR de origen natural.

[0204] En ciertas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62 %, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69 %, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76 %, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83 %, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En otras realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 90% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad

de secuencia de al menos 80% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 70% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 60% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 50% con una porción homóloga de una ITR de origen natural, cuando se alinean adecuadamente; en donde la ITR conserva una propiedad funcional de la ITR de origen natural.

[0205] En algunas realizaciones, la ITR comprende una ITR de un genoma de AAV. En algunas realizaciones, la ITR es una ITR de un genoma de AAV seleccionado entre AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 AAV11, y cualquier combinación de los mismos. En una realización particular, la ITR es una ITR del genoma de AAV2. En otra realización, la ITR es una secuencia sintética diseñada para que incluya en sus extremos 5' y 3' ITR obtenidas de uno o más genomas de AAV.

[0206] En algunas realizaciones, la ITR no se obtiene del genoma de AAV. En algunas realizaciones, la ITR es una ITR de un virus no AAV. En algunas realizaciones, la ITR es una ITR de un genoma de no AAV de la familia viral *Parvoviridae* seleccionada entre, pero sin limitarse, al grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquier combinación de los mismos. En ciertas realizaciones, la ITR se obtiene del parvovirus Erythrovirus B19 (virus humano). En otra realización, la ITR se obtiene de una cepa de parvovirus de pato real (MDPV). En ciertas realizaciones, la cepa de MDPV está atenuada, p. ej. , la cepa FZ91-30 de MDPV. En otras realizaciones, la cepa de MDPV es patógena, p. ej., la cepa YY de MDPV. En algunas realizaciones, la ITR se obtiene de un parvovirus porcino, p. ej. , parvovirus porcino U44978. En algunas realizaciones, la ITR se obtiene de un virus mínimo de ratón, p. ej., virus mínimo de ratón U34256. En algunas realizaciones, la ITR se obtiene de parvovirus canino, p. ej. , parvovirus canino M19296. En algunas realizaciones, la ITR se obtiene de un virus de enteritis de visón, p. ej., virus de enteritis de visón D00765. En algunas realizaciones, la ITR se obtiene de un *Dependoparvovirus*. En una realización, el *Dependoparvovirus* es una cepa de *Dependovirus*, parvovirus

de ganso (GPV). En una realización específica la cepa de GPV está atenuada, p. ej., cepa 82-0321V de GPV. En otra realización específica, la cepa de GPV es patógena, p. ej., cepa B de GPV.

[0207] La primera ITR y la segunda ITR de la molécula de ácido nucleico se pueden obtener del mismo genoma, p. ej., del genoma del mismo virus, o de diferentes genomas, p. ej., de los genomas de dos o más genomas de virus diferentes. . En ciertas realizaciones, la primera ITR y la segunda ITR se obtienen del mismo genoma de AAV. En una realización específica, las dos ITR presentes en la molécula de ácido nucleico de la invención son las mismas y, en particular, pueden ser ITR de AAV2. En otras realizaciones, la primera ITR se obtiene de un genoma de AAV y la segunda ITR no se obtiene de un genoma de AAV (p. ej., un genoma que no es de AAV). En otras realizaciones, la primera ITR no se obtiene de un genoma de AAV (p. ej., un genoma que no es de AAV) y la segunda ITR se obtiene de un genoma de AAV. En otras realizaciones más, ni la primera ITR ni la segunda ITR se obtienen de un genoma de AAV (p. ej., un genoma que no es de AAV). En una realización particular, la primera ITR y la segunda ITR son idénticas.

[0208] En algunas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquier combinación de los mismos. En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV y la primera ITR se obtiene de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquier combinación de los mismos. En otras realizaciones, la primera ITR y la segunda ITR se obtienen de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquier combinación de los mismos, en donde la primera ITR y la segunda ITR se obtienen del mismo genoma. En otras realizaciones, la primera ITR y la segunda ITR se obtienen de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquier combinación de los mismos, en donde la primera ITR y la segunda ITR se obtienen de los diferentes genomas.

[0209] En algunas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene del Parvovirus Erythrovirus B19 (virus humano). En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene del Parvovirus Erythrovirus B19 (virus humano).

- 5 **[0210]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten toda o una porción de una ITR obtenida de B19. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos
- 10 aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 167, 168, 169, 170, y 171 , en donde la primera ITR y/o la segunda ITR conservan una propiedad funcional de la ITR de B19 de la cual se obtienen. En algunas
- 15 realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos
- 20 aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 167, 168, 169, 170, y 171 , en donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una estructura de horquilla en forma de T.
- 25 **[0211]** En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada entre SEQ ID NO: 167, 168, 169, 170, y 171. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 167. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 168. En algunas realizaciones, la
- 30 primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 169. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 170. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 171.

Tabla 1. Secuencias ITR de Parvovirus de Muestra

Parvovirus	ID ITR	Descripción	Longitud (nt)	Secuencia
B19	wt	Gene Bank: KY940273.1	383	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGC GGGACTTCCGGTACAAGATGGCGGACAATTACGT CATTTCCCTGTGACGTCATTTCCCTGTGACGTC TCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGC TCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACT AAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCA CGTCAACCCCAAGCGCTGGCCCAGAGCCAACCT AATTCCGGAAGTCCCGCCACCGGAAGTGACGTC ACAGGAAATGACGTCACAGGAAATGACGTAATTG TCCGCCATCTTGTACCGGAAGTCCCGCCTACCGG CGGCGACCGGCGGCATCTGATTTGGTGTCTTCTT TTAAATTTT (SEQ ID NO: 167)
	d135	Excluye los primeros 135 nucleótidos	248	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACAC TAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGC ACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCC TAATTCCGGAAGTCCCGCCACCGGAAGTGACGT CACAGGAAATGACGTCACAGGAAATGACGTAATT GTCCGCCATCTTGTACCGGAAGTCCCGCCTACCG GCGGCGACCGGCGGCATCTGATTTGGTGTCTTCT TTTAAATTTT (SEQ ID NO: 168)
	v1	Secuencia mínima basada en la comparación con AAV2	129	CGGCGCGCCGCTTGATCTTAGTGGCACGTCAACC AGCGCTGGCCCAGAGCCAACCTAATTCCGGAAG TCCTCAGTCCGCCATCTTGCCCGCCTACCGGCGG CGACCGGCGGCATCATTGGTGTCTT (SEQ ID NO: 169)
	v2	Excluye los primeros 135 nucleótidos y los 135 nucleótidos complementarios en el palíndromo	113	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACAC TAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGC ACGTCAACCCCAAGCGCTGGCCCAGAGTGTCTTC TTTTAAATTTT (SEQ ID NO: 170)
	v3	Secuencia mínima basada en la comparación con GPV	340	CAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCG GGACTTCCGGTACAAGATGGCGGACAATTACGTC ATTTCCCTGTGACGTATTTCCCTGTGACGTACTTCC GGTGGCGGGACTTCCGGAATTTGGCTCTGGGCC AGCTTGCTTGGGGTTGCCTTGACCAAGCGCGCGC CGCTTGATCACCCCAAGCGCTGGCCCAGAGCCAC CTAACCAGGAAGTCCCCCACCAGGAAGTGACGTCA CAGGAAAGACGTCACAGGAAGTAATTGTCCGCCA TCTTGTACCGGAAGTCCCGCACCGGCGGCGACCG GCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT (SEQ ID NO: 171)
GPV	wt	Gene Bank: U25749.1	444	CTCATTGGAGGGTTTCGTTTCGAACCAGCCAA TCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCA CATGCTTCCGGTGACGCACATCCGGTGACGTAGT TCCGGTCACGTGCTTCCTGTACGTGTTTCCGGT

				CACGTGACTTCCGGTCATGTGACTTCCGGTGACG TGTTTCCGGCTGTTAGGTTGACCACGCGCATGCC GCGCGGTCAGCCCAATAGTTAAGCCGGAACACG TCACCGGAAGTCACATGACCGGAAGTCACGTGAC CGGAAACACGTGACAGGAAGCACGTGACCGGAAC TACGTACCGGATGTGCGTCACCGGAAGCATGTG ACCGGAACTTGCGTCACTTCCCCCTCCCCTGATT GGCTGGTTTGAACGAACGAACCCCTCCAATGAGAC TCAAGGACAAGAGGATATTTTGC GCGCCAGGAAG TG (SEQ ID NO: 172)
	d162	Excluye los primeros 162 nucleótidos	282	CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACG CGCATGCCGCGCGGTCAGCCCAATAGTTAAGCCG GAAACACGTACCGGAAGTCACATGACCGGAAGT CACGTGACCGGAAACACGTGACAGGAAGCACGTG ACCGGAACTACGTACCGGATGTGCGTCACCGGA AGCATGTGACCGGAACTTGCGTCACTTCCCCCTC CCCTGATTGGCTGGTTTGAACGAACGAACCCCTCC AATGAGACTCAAGGACAAGAGGATATTTTGC GCG CCAGGAAGTG (SEQ ID NO: 173)
	v1	Secuencia mínima basada en la comparación con AAV2	145	TTGACCACGCGCATGCCGCGCGGTCAGCCCAATA GTTAAGCCGGGTGACCACACGTGACAGGAAGCAC GGGATGTGCGTCACCGGAAGCAGTGACCGGGCTG GTTTGAACGAACGAACCCCTCCAACCTCAAGGACAA GAGGATATT (SEQ ID NO: 174)
	v2	Excluye los primeros 162 nucleótidos y los 162 nucleótidos complemen- tarios correspon- dientes en el palíndromo	120	CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACG CGCATGCCGCGCGGTCAGCCCAATAGTTAAGCCG GAAACACGTACCGGACTCAAGGACAAGAGGATAT TTTGC GCGCCAGGAAGTG (SEQ ID NO: 175)
	v3	Secuencia mínima basada en la comparación con B19	102	GGGAACAATCAGGGGAAGTGACCGGTGACGTCAT GTAACCTTGCGTCACTTCCCGTTTGAACGAACGAA CGAGACTCAAGGACAAGAGGCGCGCCAGGAAGTG (SEQ ID NO: 176)
AAV2	wt	Gene Bank: NC_001401.2	145	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCA CTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCC GGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGA GCGCGCAGAGAGGGAGTGGCCAACTCCATCACTA GGGGTTTCCT (SEQ ID NO: 177)
	GTx	Utilizada en vectores GTx	130	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGG CAAAGCCCCGGGCGTCGGGCGACCTTTGGTTCGCCC GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGA GTGGCCAACTCCATCACTAGGGGTTTCCT (SEQ ID NO: 178)

[0212] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 167. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 167. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 168. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 168. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 169, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 167, conserva una propiedad funcional de la ITR de B19 de la cual se obtiene. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 169, y en donde la secuencia de nucleótidos es a al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO:

167, en donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

[0213] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en toda o una porción de una ITR obtenida de B19. En algunas realizaciones, la segunda ITR es un complemento inverso de la primera ITR. En algunas realizaciones, la primera ITR es un complemento inverso de la segunda ITR. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 180, 181, 185, y 186, o un derivado funcional de la misma. En algunas realizaciones, el derivado funcional conserva una propiedad funcional de la ITR de B19 de la cual se obtiene. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 180, 181, 185, y 186, o un derivado funcional de la misma. En algunas realizaciones, el derivado funcional es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

[0214] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 180. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 180. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos

aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 181. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 181. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 185. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 185. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 186. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 186.

[0215] En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada entre SEQ ID NO: 180, 181, 185, y 186. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 180. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 181. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 185. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 186. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 180. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 181. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 185. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 186.

[0216] En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 180, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 181. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 181, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 180. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 185, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 186. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 186, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 185.

[0217] En algunas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de GPV. En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene de GPV.

[0218] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 172. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 172. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 173. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 173. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en toda o una porción de una ITR obtenida de GPV. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos

aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO:

- 5 172, 173, 174, 175, y 176, en donde la primera ITR y/o la segunda ITR conservan una propiedad funcional de la ITR de GPV de la cual se obtiene. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en toda o una porción de una ITR obtenida de GPV. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos
- 10 aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO:
- 15 172, 173, 174, 175, y 176, en donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada entre SEQ ID NO: 172, 173, 174, 175, y 176. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos
- 20 expuesta en SEQ ID NO: 172. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 173. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 174. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 175. En algunas realizaciones, la primera ITR y/o la
- 25 segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 176.

- [0219] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 174, y en donde la secuencia de nucleótidos es a al menos aproximadamente
- 30 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 172, en donde la primera ITR y/o la segunda ITR conservan una

propiedad funcional de la ITR de GPV de la cual se obtienen. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 174, y en donde la secuencia de nucleótidos es a al menos aproximadamente 50%, al menos aproximadamente 55%, al menos

5 aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 172, en
10 donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

[0220] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 176, y en donde la secuencia de nucleótidos es al menos aproximadamente
15 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia
20 de nucleótidos expuesta en SEQ ID NO: 172, en donde la primera ITR y/o la segunda ITR conservan una propiedad funcional de la ITR de GPV de la cual se obtienen. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima expuesta en SEQ ID NO: 176, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos
25 aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 172, en
30 donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

[0221] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en toda o una porción de una ITR obtenida de GPV. En algunas realizaciones, la segunda ITR es un complemento

inverso de la primera ITR. En algunas realizaciones, la primera ITR es un complemento inverso de la segunda ITR. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 183, 184, 187 y 188, o un derivado funcional de la misma. En algunas realizaciones, el derivado funcional conserva una propiedad funcional de la ITR de GPV de la que se obtiene. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada entre SEQ ID NO: 183, 184, 187 y 188, o un derivado funcional de la misma. En algunas realizaciones, el derivado funcional es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

[0222] En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 183. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 183. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100%

idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 184. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 184. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 187. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 187. En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a la secuencia de nucleótidos expuesta en SEQ ID NO: 188. En ciertas realizaciones, la primera ITR y/o la segunda ITR consisten en SEQ ID NO: 188.

[0223] En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada entre SEQ ID NO: 183, 184, 187 y 188. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 183. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 184. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 187. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 188. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 183. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 184. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 187. En algunas realizaciones, la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 188.

[0224] En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 183, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 184. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 184, y la segunda ITR comprende o consiste en la

secuencia de nucleótidos expuesta en SEQ ID NO: 183. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 187, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 188. En algunas realizaciones, la primera ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 188, y la segunda ITR comprende o consiste en la secuencia de nucleótidos expuesta en SEQ ID NO: 187.

[0225] En ciertas realizaciones, una de la primera ITR o la segunda ITR comprende o consiste en toda o una porción de una ITR obtenida de AAV2. En algunas realizaciones, la primera ITR o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 177 o 178, en donde la primera ITR y/o la segunda ITR conservan una propiedad funcional de la ITR de AAV2 de la cual se obtienen. En algunas realizaciones, la primera ITR o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 177 o 178, en donde la primera ITR y/o la segunda ITR son capaces de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos expuesta en SEQ ID NO: 177 o 178. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 177. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos expuesta en SEQ ID NO: 178.

[0226] En algunas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de una cepa de parvovirus de pato real (MDPV). En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene de una cepa de parvovirus de pato real

(MDPV). En ciertas realizaciones, la cepa de MDPV está atenuada, p. ej., la cepa FZ91-30 de MDPV. En otras realizaciones, la cepa de MDPV es patógena, p. ej., la cepa YY de MDPV.

[0227] En algunas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de un *Dependoparvovirus*. En algunas realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene de un *Dependoparvovirus*. En otras realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de una cepa del parvovirus de ganso *Dependovirus* (GPV). En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene de una cepa de GPV *Dependovirus*. En ciertas realizaciones, la cepa de GPV está atenuada, p. ej., la cepa 82-0321V de GPV. En otras realizaciones, la cepa de GPV es patógena, p. ej., cepa B de GPV.

[0228] En ciertas realizaciones, la primera ITR se obtiene de un genoma de AAV, y la segunda ITR se obtiene de un genoma seleccionado del grupo que consiste en parvovirus porcino, p. ej., la cepa U44978 de parvovirus porcino; virus mínimo de ratón, p. ej., la cepa U34256 de virus mínimo de ratón; parvovirus canino, p. ej., la cepa M19296 de parvovirus canino; virus de la enteritis de visón, p. ej., la cepa D00765; del virus de enteritis del visón y cualquier combinación de los mismos. En otras realizaciones, la segunda ITR se obtiene de un genoma de AAV, y la primera ITR se obtiene de un genoma seleccionado del grupo que consiste en parvovirus porcino, p. ej., la cepa U44978 de parvovirus porcino; virus mínimo de ratón, p. ej., la cepa U34256 de virus mínimo de ratón; parvovirus canino, p. ej., la cepa M 19296 de parvovirus canino; virus de la enteritis del visón, p. ej., la cepa D00765 de virus de la enteritis del visón; y cualquier combinación de los mismos.

[0229] En otra realización particular, la ITR es una secuencia sintética diseñada genéticamente para que incluya en sus extremos 5' y 3' ITR que no se obtienen de un genoma de AAV. En otra realización particular, la ITR es una secuencia sintética diseñada genéticamente para que incluya en sus extremos 5' y 3' ITR obtenidas de uno o más genomas que no son AAV. Las dos ITR presentes en la molécula de ácido nucleico de la invención pueden ser genomas que no son de AAV iguales o diferentes. En particular, las ITR se pueden obtener del mismo genoma que no es AAV. En una realización específica, las dos ITR presentes en la molécula de ácido nucleico de la invención son las mismas y, en particular, pueden ser ITR de AAV2.

[0230] En algunas realizaciones, la secuencia de ITR comprende una o más secuencias palindrómicas. Una secuencia palindrómica de una ITR descrita en la presente memoria incluye, pero no se limita a, secuencias palindrómicas nativas (es decir, secuencias que se encuentran en la naturaleza), secuencias sintéticas (es decir, secuencias que no se encuentran en la naturaleza), tales como secuencias

pseudopalindrómicas y combinaciones o formas modificadas de las mismos. Una "secuencia pseudopalindrómica" es una secuencia de ADN palindrómica, que incluye una secuencia palindrómica imperfecta, que comparte menos de 80%, incluso menos de 70%, 60%, 50%, 40%, 30%, 20%, 10% o 5%, o ninguna, identidad de secuencia de ácido nucleico con secuencias en la secuencia palindrómica nativa de AAV o que no es de AAV que forman una estructura secundaria. Las secuencias palindrómicas nativas se pueden obtener o derivar de cualquier genoma descrito en la presente memoria.

[0231] La secuencia palindrómica puede ser continua o interrumpida. En algunas realizaciones, la secuencia palindrómica se interrumpe, en donde la secuencia palindrómica comprende una inserción de una segunda secuencia. En algunas realizaciones, la segunda secuencia comprende un promotor, un potenciador, un sitio de integración para una integrasa (p. ej., sitios para recombinasa Cre o Flp), un marco de lectura abierto para un producto génico o una combinación de los mismos.

[0232] En algunas realizaciones, las ITR forman estructuras de bucle en horquilla. En una realización, la primera ITR forma una estructura de horquilla. En otra realización, la segunda ITR forma una estructura de horquilla. En otra realización más, tanto la primera ITR como la segunda ITR forman estructuras de horquilla. En algunas realizaciones, la primera ITR y/o la segunda ITR no forman una estructura de horquilla en forma de T. En ciertas realizaciones, la primera ITR y/o la segunda ITR forman una estructura de horquilla en forma de T. En algunas realizaciones, la estructura de horquilla sin forma de T comprende una estructura de horquilla en forma de U.

[0233] En algunas realizaciones, una ITR en una molécula de ácido nucleico descrita en la presente memoria puede ser una ITR activada transcripcionalmente. Una ITR activada transcripcionalmente puede comprender toda o una parte de una ITR de tipo salvaje que se ha activado transcripcionalmente mediante la inclusión de al menos un elemento transcripcionalmente activo. Varios tipos de elementos transcripcionalmente activos son adecuados para su uso en este contexto. En algunas realizaciones, el elemento transcripcionalmente activo es un elemento transcripcionalmente activo constitutivo. Los elementos transcripcionalmente activos constitutivos proporcionan un nivel continuo de transcripción génica y se prefieren cuando se desea que el transgén se exprese de manera continua. En otras realizaciones, el elemento transcripcionalmente activo es un elemento transcripcionalmente activo inducible. Los elementos transcripcionalmente activos inducibles generalmente muestran baja actividad en ausencia de un inductor (o condición inductora) y se regulan por incremento en presencia del inductor (o cambian a una condición inductora). Los elementos transcripcionalmente activos inducibles pueden ser preferidos cuando la expresión se desea solo en ciertos momentos o en ciertos lugares, o cuando es deseable valorar el nivel de expresión utilizando un agente inductor. Los elementos transcripcionalmente

activos también pueden ser específicos de tejido; es decir, muestran actividad solo en ciertos tejidos o tipos de células.

[0234] Los elementos transcripcionalmente activos se pueden incorporar a una ITR de diversas formas.

En algunas realizaciones, un elemento transcripcionalmente activo se incorpora en 5' a cualquier porción

de una ITR o en 3' a cualquier porción de una ITR. En otras realizaciones, un elemento

transcripcionalmente activo de una ITR activada transcripcionalmente se encuentra entre dos secuencias

de ITR. Si el elemento transcripcionalmente activo comprende dos o más elementos que deben estar

separados, esos elementos pueden alternar con porciones de la ITR. En algunas realizaciones, una

estructura de horquilla de una ITR se elimina y se reemplaza con repeticiones invertidas de un elemento

transcripcional. Esta última disposición crearía una horquilla que imitaría la estructura de la porción

suprimida. También pueden estar presentes múltiples elementos transcripcionalmente activos en tándem

en una ITR activada transcripcionalmente, y estos pueden ser adyacentes o estar espaciados. Además, se

pueden introducir sitios de unión a proteínas (p. ej., Sitios de unión a Rep) en elementos

transcripcionalmente activos de las ITR activadas transcripcionalmente. Un elemento

transcripcionalmente activo puede comprender cualquier secuencia que permita la transcripción

controlada de ADN mediante ARN polimerasa para formar ARN y puede comprender, por ejemplo, un

elemento transcripcionalmente activo, como se define a continuación.

[0235] Las ITR activadas transcripcionalmente proporcionan funciones tanto de activación

transcripcional como de ITR a la molécula de ácido nucleico en una longitud de secuencia de nucleótidos

relativamente limitada que maximiza eficazmente la longitud de un transgén que se puede transportar y

expresar a partir de la molécula de ácido nucleico. La incorporación de un elemento transcripcionalmente

activo a una ITR se puede lograr de diversas formas. Una comparación de la secuencia de la ITR y los

requisitos de secuencia del elemento transcripcionalmente activo puede proporcionar información sobre

las formas de codificar el elemento dentro de una ITR. Por ejemplo, la actividad transcripcional se puede

añadir a una ITR mediante la introducción de cambios específicos en la secuencia de la ITR que replica

los elementos funcionales del elemento transcripcionalmente activo. Existen varios mecanismos en la

técnica para añadir, suprimir y/o cambiar de manera eficaz secuencias de nucleótidos particulares en

sitios específicos (véase, por ejemplo, Deng y Nickoloff (1992) Anal. Biochem. 200:81-88). Otra forma

de crear ITR activadas transcripcionalmente implica la introducción de un sitio de restricción en una

ubicación deseada en la ITR. Adicionalmente, se pueden incorporar múltiples elementos

transcripcionalmente activos a una ITR activada transcripcionalmente, utilizando métodos conocidos en la técnica.

[0236] A modo de ilustración, las ITR activadas transcripcionalmente se pueden generar mediante la inclusión de uno o más elementos transcripcionalmente activos tales como: caja TATA, caja GC, caja CCAAT, sitio Sp1, región Inr, sitio CRE (elemento regulador de AMPc), sitio ATF-1/CRE, caja ARBb, caja APBa, caja CArG, caja CCAC o cualquier otro elemento implicado en la transcripción como se conoce en la técnica.

[0237] Los aspectos de la presente divulgación proporcionan un método para clonar una molécula de ácido nucleico descrita en la presente memoria, que comprende insertar una molécula de ácido nucleico susceptible de estructuras secundarias complejas en un vector adecuado e introducir el vector resultante en una cepa anfitriona bacteriana adecuada. Como se conoce en la técnica, las estructuras secundarias complejas (p. ej., regiones palindrómicas largas) de ácidos nucleicos pueden ser inestables y difíciles de clonar en cepas anfitrionas bacterianas. Por ejemplo, las moléculas de ácido nucleico que comprenden una primera ITR y una segunda ITR (p. ej., ITR parvovirales que no son de AAV, p. ej., ITR de B19 o GPV) de la presente divulgación pueden ser difíciles de clonar utilizando metodologías convencionales. Los palíndromos de ADN largos inhiben la replicación del ADN y son inestables en los genomas de *E.*

coli, *Bacillus*, *Streptococcus*, *Streptomyces*, *S. cerevisiae*, ratones y seres humanos. Estos efectos resultan de la formación de estructuras de horquilla o cruciformes mediante el emparejamiento de bases intrahebra. En *E. coli*, la inhibición de la replicación del ADN se puede superar significativamente en mutantes SbcC o SbcD. SbcD es la subunidad nucleasa y SbcC es la subunidad ATPasa del complejo SbcCD. El complejo SbcCD de *E. coli* es un complejo de exonucleasa responsable de prevenir la replicación de palíndromos largos. El complejo SbcCD es un complejo nuclear con actividad exonucleasa de ADN de doble hebra dependiente de ATP y actividad endonucleasa de ADN de hebra sencilla independiente de ATP. SbcCD puede reconocer los palíndromos de ADN y colapsar las horquillas de replicación al atacar las estructuras de horquilla que surgen.

[0238] En ciertas realizaciones, una cepa anfitriona bacteriana adecuada es incapaz de resolver estructuras de ADN cruciforme. En determinadas realizaciones, una cepa anfitriona bacteriana adecuada comprende una alteración en el complejo SbcCD. En algunas realizaciones, la alteración en el complejo SbcCD comprende una alteración genética en el gen SbcC y/o el gen SbcD. En ciertas realizaciones, la alteración en el complejo SbcCD comprende una alteración genética en el gen SbcC. Se conocen en la técnica diversas cepas anfitrionas bacterianas que comprenden una alteración genética en el gen SbcC. Por ejemplo, sin limitación, la cepa anfitriona bacteriana PMC103 comprende el genotipo *sbcC*, *recD*, *mcrA*, $\Delta mcrBCF$; la cepa anfitriona bacteriana PMC107 comprende el genotipo *recBC*, *recJ*, *sbcBC*, *mcrA*, $\Delta mcrBCF$; y la cepa anfitriona bacteriana SURE comprende el genotipo *recB*, *recJ*, *sbcC*, *mcrA*, $\Delta mcrBCF$, *umuC*, *uvrC*. Por consiguiente, en algunas realizaciones, un método de clonación de una

molécula de ácido nucleico descrito en la presente memoria comprende insertar una molécula de ácido nucleico susceptible de estructuras secundarias complejas en un vector adecuado e introducir el vector resultante en la cepa anfitriona PMC103, PMC107 o SURE. En ciertas realizaciones, el método de clonación de una molécula de ácido nucleico descrito en la presente memoria comprende insertar una molécula de ácido nucleico susceptible de estructuras secundarias complejas en un vector adecuado, e introducir el vector resultante en la cepa anfitriona PMC103.

[0239] Los vectores adecuados se conocen en la técnica y se describen en otra parte de la presente memoria. En ciertas realizaciones, un vector adecuado para su uso en una metodología de clonación de la presente divulgación es un vector de bajo número de copias. En ciertas realizaciones, un vector adecuado para su uso en una metodología de clonación de la presente divulgación es pBR322.

[0240] Por consiguiente, la presente divulgación proporciona un método para clonar una molécula de ácido nucleico, que comprende insertar una molécula de ácido nucleico susceptible de estructuras secundarias complejas en un vector adecuado e introducir el vector resultante en una cepa anfitriona bacteriana que comprende una alteración en el complejo SbcCD, en donde la molécula de ácido nucleico comprende una primera repetición terminal invertida (ITR) y una segunda ITR, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

B. Proteínas Terapéuticas

[0241] Ciertos aspectos de la presente divulgación están dirigidos a una molécula de ácido nucleico que comprende una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica. En algunas realizaciones, el casete genético codifica una proteína terapéutica. En algunas realizaciones, el casete genético codifica más de una proteína terapéutica. En algunas realizaciones, el casete genético codifica dos o más copias de la misma proteína terapéutica. En algunas realizaciones, el casete genético codifica dos o más variantes de la misma proteína terapéutica. En algunas realizaciones, el casete genético codifica dos o más proteínas terapéuticas diferentes.

[0242] Ciertas realizaciones de la presente divulgación están dirigidas a una molécula de ácido nucleico que comprende una primera ITR, una segunda ITR y un casete genético que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un factor de coagulación. En algunas

realizaciones, el factor de coagulación se selecciona del grupo que consiste en FI, FII, FIII, FIV, FV, FVI, FVII, FVIII, FIX, FX, FXI, FXII, FXIII, VWF, precalicreína, cininógeno de alto peso molecular, fibronectina, antitrombina III, cofactor II de heparina, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionado con la proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador del plasminógeno tisular (tPA), uroquinasa, inhibidor del activador de plasminógeno 1 (PAI-1), inhibidor 2 del activador de plasminógeno (PAI2), cualquier cimógeno de los mismos, cualquier forma activa de los mismos y cualquier combinación de los mismos. En una realización, el factor de coagulación comprende FVIII o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende FIX o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende FVII o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende VWF o una variante o fragmento del mismo.

1. Factores de Coagulación

[0243] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido de factor VIII. "Factor VIII", abreviado en la presente solicitud como "FVIII", como se emplea en la presente memoria, significa polipéptido funcional de FVIII en su papel normal en la coagulación, a menos que se especifique lo contrario. Por tanto, el término FVIII incluye variantes de polipéptidos que son funcionales. "Una proteína FVIII" se utiliza de manera indistinta con polipéptido (o proteína) de FVIII o FVIII. Los ejemplos de las funciones de FVIII incluyen, pero no se limitan a, la capacidad de activar la coagulación, la capacidad de actuar como cofactor del factor IX, o la capacidad para formar un complejo de tenasa con el factor IX en presencia de Ca^{2+} y fosfolípidos, que a continuación convierte el Factor X en la forma Xa activada. La proteína FVIII puede ser la proteína FVIII humana, porcina, canina, de rata o murina. Además, las comparaciones entre el FVIII de seres humanos y otras especies han identificado restos conservados que probablemente sean necesarios para la función (Cameron et al., *Thromb. Haemost.* 79:317-22 (1998); US 6.251.632). Se conocen las secuencias de los polipéptidos y polinucleótidos de longitud completa, al igual que muchos fragmentos funcionales, mutantes y versiones modificadas. Varias secuencias de aminoácidos y nucleótidos de FVIII se describen, p. ej., en las Publicaciones de EE.UU. Núm. 2015/0158929 A1, 2014/0308280 A1 y 2014/0370035 A1 y la Publicación Internacional Núm. WO 2015/106052 A1. Los polipéptidos de FVIII incluyen, p. ej., FVIII de longitud completa, FVIII de longitud completa menos Met en el extremo N-terminal, FVIII maduro (menos la secuencia señal), FVIII maduro con una Met adicional en el extremo N-terminal y/o FVIII con

una delección total o parcial del dominio B. Las variantes de FVIII incluyen delecciones del dominio B, ya sean delecciones parciales o totales.

a. FVIII y Secuencias de Polinucleótidos que Codifican la Proteína FVIII

[0244] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una

segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido de factor VIII.

"Factor VIII", abreviado en la presente solicitud como "FVIII", como se emplea en la presente memoria, significa un polipéptido funcional de FVIII en su papel normal en la coagulación, a menos que se especifique lo contrario. Por tanto, el término FVIII incluye variantes de polipéptidos que son

funcionales. "Una proteína de FVIII" se utiliza de manera indistinta con polipéptido (o proteína) de FVIII o FVIII. Los ejemplos de las funciones de FVIII incluyen, pero no se limitan a, la capacidad de activar la coagulación, la capacidad de actuar como cofactor del factor IX, o la capacidad de formar un complejo de tenasa con el factor IX en presencia de Ca^{2+} y fosfolípidos, que a continuación convierte el Factor X en la forma Xa activada. La proteína FVIII puede ser la proteína FVIII humana, porcina, canina, de rata o murina. Además, las comparaciones entre FVIII de seres humanos y otras especies han identificado restos conservados que probablemente sean necesarios para la función (Cameron et al., *Thromb. Haemost.* 79:317-22 (1998); US 6.251. 632). Se conocen las secuencias de polipéptidos y polinucleótidos de longitud completa, al igual que muchos fragmentos funcionales, mutantes y versiones modificadas.

Varias secuencias de aminoácidos y nucleótidos de FVIII se describen, por ejemplo, en las Publicaciones de EE.UU. Núm. 2015/0158929 A1, 2014/0308280 A1 y 2014/0370035 A1 y la Publicación Internacional Núm. WO 2015/106052 A1. Los polipéptidos de FVIII incluyen, p. ej., FVIII de longitud completa, FVIII de longitud completa menos Met en el extremo N-terminal, FVIII maduro (menos la secuencia señal), FVIII maduro con una Met adicional en el extremo N-terminal y/o FVIII con una delección total o parcial del dominio B. Las variantes de FVIII incluyen delecciones del dominio B, ya sean delecciones parciales o totales.

[0245] La porción de FVIII en la proteína quimérica utilizada en la presente memoria tiene actividad de FVIII. La actividad de FVIII se puede medir mediante cualquier método conocido en la técnica. Se encuentra disponible una serie de pruebas para evaluar la función del sistema de coagulación: prueba de tiempo de tromboplastina parcial activada (aPTT), ensayo cromogénico, ensayo ROTEM, prueba de tiempo de protrombina (PT) (también utilizada para determinar INR), ensayo de fibrinógeno (a menudo por el Método de Clauss), recuento de plaquetas, pruebas de función plaquetaria (a menudo por PFA-100), TCT, tiempo de sangrado, prueba de mezclado (si una anomalía se corrige si el plasma del paciente se mezcla con plasma normal), ensayos de factor de coagulación, anticuerpos antifosfolípidos, dímero D,

pruebas genéticas (p. ej., factor V Leiden, mutación de protrombina G20210A), tiempo de veneno de víbora de Russell diluido (dRWT), pruebas de función plaquetaria diversas, tromboelastografía (TEG o Sonoclot), tromboelastomería (TEM[®], p. ej., ROTEM[®]), o tiempo de lisis de euglobulina (ELT).

[0246] La prueba aPTT es un indicador de rendimiento que mide la eficacia tanto de las vías tanto "intrínseca" (también denominada vía de activación por contacto) como de coagulación común. Esta prueba se utiliza comúnmente para medir la actividad de coagulación de factores de coagulación recombinantes disponibles comercialmente, p. ej., FVIII. Se utiliza junto con el tiempo de protrombina (TP), que mide la vía extrínseca.

[0247] El análisis ROTEM proporciona información sobre toda la cinética de la hemostasia: tiempo de coagulación, formación de coágulos, estabilidad de coágulos y lisis. Los diferentes parámetros de la tromboelastometría dependen de la actividad del sistema de coagulación plasmática, la función plaquetaria, la fibrinólisis o de muchos factores que influyen en estas interacciones. Este ensayo puede proporcionar una visión completa de la hemostasia secundaria.

[0248] El mecanismo de ensayo cromogénico se basa en los principios de la cascada de coagulación sanguínea, donde el FVIII activado acelera la conversión del Factor X en Factor Xa en presencia de Factor IX activado, fosfolípidos e iones calcio. La actividad del factor Xa se evalúa mediante la hidrólisis de un sustrato de p-nitroanilida (pNA) específico del factor Xa. La velocidad inicial de liberación de p-nitroanilina medida a 405 nM es directamente proporcional a la actividad del factor Xa y, por tanto, a la actividad de FVIII en la muestra.

[0249] El ensayo cromogénico es recomendado por el Subcomité de FVIII y Factor IX del Comité Científico y de Normalización (SSC) de la Sociedad Internacional de Trombosis y Hemostasis (ISTH). Desde 1994, el ensayo cromogénico también ha sido el método de referencia de la Farmacopea Europea para la asignación de la potencia al producto concentrado de FVIII. Por lo tanto, en una realización, el polipéptido quimérico que comprende FVIII tiene actividad de FVIII comparable a un polipéptido quimérico que comprende FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]).

[0250] En otra realización, la proteína quimérica que comprende FVIII de esta descripción tiene una tasa de generación de Factor Xa comparable a una proteína quimérica que comprende FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], DE REFACTO[®], o ELOCTATE[®]).

[0251] Para activar el Factor X a Factor Xa, el Factor IX activado (Factor IXa) hidroliza un enlace de arginina-isoleucina en el Factor X para formar el Factor Xa en presencia de Ca²⁺, fosfolípidos de

membrana y un cofactor de FVIII. Por lo tanto, la interacción de FVIII con el Factor IX es fundamental en la vía de coagulación. En ciertas realizaciones, el polipéptido quimérico que comprende FVIII puede interactuar con el Factor IXa a una velocidad comparable a un polipéptido quimérico que comprende una secuencia de FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]).

5 **[0252]** Además, el FVIII se une al Factor de von Willebrand mientras está inactivo en circulación. El FVIII se degrada rápidamente cuando no está unido al VWF y se libera del VWF por la acción de la trombina. En algunas realizaciones, el polipéptido quimérico que comprende FVIII se une al Factor de von Willebrand a un nivel comparable a un polipéptido quimérico que comprende la secuencia de FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]).

10 **[0253]** El FVIII se puede inactivar por medio de la proteína C activada en presencia de calcio y fosfolípidos. La proteína C activada escinde la cadena pesada del FVIII después de la Arginina 336 en el dominio A1, que altera un sitio de interacción del sustrato del Factor X, y escinde después de la Arginina 562 en el dominio A2, que mejora la disociación del dominio A2 y también altera un sitio de interacción con el Factor IXa. Esta escisión también escinde el dominio A2 (43 kDa) y genera dominios A2-N (18
15 kDa) y A2-C (25 kDa). Por tanto, la proteína C activada puede catalizar múltiples sitios de escisión en la cadena pesada. En una realización, el polipéptido quimérico que comprende FVIII se inactiva por la proteína C activada a un nivel comparable a un polipéptido quimérico que comprende una secuencia de FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]).

20 **[0254]** En otras realizaciones, la proteína quimérica que comprende FVIII tiene actividad de FVIII *in vivo* comparable a la de un polipéptido quimérico que comprende una secuencia de FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]). En una realización particular, el polipéptido quimérico que comprende FVIII es capaz de proteger a un ratón de HemA a un nivel comparable al de un polipéptido quimérico que comprende una secuencia de FVIII maduro o un FVIII BDD (p. ej., ADVATE[®], REFACTO[®], o ELOCTATE[®]) en un modelo de transección de la vena de la
25 cola de ratón HemA.

30 **[0255]** Un "dominio B" de FVIII, como se emplea en la presente memoria, es el mismo que el dominio B conocido en la técnica que se define por la identidad de la secuencia de aminoácidos interna y los sitios de escisión proteolítica por trombina, p. ej., los restos Ser741-Arg1648 de FVIII humano maduro. Los otros dominios de FVIII humano están definidos por los siguientes restos de aminoácidos, con respecto al FVIII humano maduro: A1, restos Ala1-Arg372; A2, restos Ser373-Arg740; A3, restos Ser1690-Ile2032; C1, restos Arg2033-Asn2172; C2, restos Ser2173-Tyr2332 del FVIII maduro. Los números de los restos de la secuencia utilizados en la presente memoria sin hacer referencia a ningún Número de SEQ ID

corresponden a la secuencia de FVIII sin la secuencia del péptido señal (19 aminoácidos) a menos que se indique lo contrario. La secuencia A3-C1-C2, también conocida como cadena pesada de FVIII, incluye los restos Ser1690-Tyr2332. La secuencia restante, los restos Glu1649-Arg1689, se denominan habitualmente péptido de activación de la cadena ligera de FVIII. Las ubicaciones de los límites para todos los dominios, incluidos los dominios B, para el FVIII porcino, de ratón y canino también se conocen en la técnica. En una realización, el dominio B de FVIII se elimina ("FVIII con dominio B suprimido" o "FVIII BDD"). Un ejemplo de un FVIII BDD es REFACTO® (FVIII BDD recombinante). En una realización particular, la variante de FVIII con el dominio B suprimido comprende una delección de los restos de aminoácidos 746 a 1648 del FVIII maduro.

10 **[0256]** Un "FVIII con el dominio B suprimido" puede tener las delecciones totales o parciales descritas en las Patentes de Estados Unidos 6.316.226, 6.346.513, 7.041.635, 5.789.203, 6.060.447, 5.595.886, 6.228.620, 5.972.885, 6.048.720, 5.543.502, 5.610.278, 5.171, 844, 5.112.950, 4.868.112 y 6.458.563 y la Publicación Internacional Núm. WO 2015106052 A1 (PCT/US2015/010738). En algunas realizaciones, una secuencia de FVIII con el dominio B suprimido utilizada en los métodos de la presente divulgación comprende cualquiera de las supresiones descritas en la col. 4, línea 4 hasta la col. 5, línea 28 y los Ejemplos 1-5 de la Patente de Estados Unidos 6.316.226 (también en el documento US 6.346.513). En otra realización, un Factor VIII con el dominio B suprimido es el Factor VIII con el dominio B suprimido S743/Q1638 (FVIII BDD SQ) (p. ej., Factor VIII que tiene una delección del aminoácido 744 al aminoácido 1637, p. ej., Factor VIII que tiene los aminoácidos 1-743 y los aminoácidos 1638-2332 del FVIII maduro). En algunas realizaciones, un FVIII con el dominio B suprimido utilizado en los métodos de la presente divulgación tiene una delección descrita en la col. 2, líneas 26-51 y Ejemplos 5-8 de la Patente de Estados Unidos Núm. 5.789.203 (también los documentos US 6.060.447, US 5.595.886 y US 6.228.620). En algunas realizaciones, un Factor VIII con el dominio B suprimido tiene una delección descrita en la col. 1, líneas 25 a col. 2, línea 40 de la Patente de Estados Unidos Núm. 5.972.885; columna. 6, líneas 1-22 y el Ejemplo 1 de la Patente de Estados Unidos 6.048.720; columna. 2, líneas 17-46 de la Patente de Estados Unidos Núm. 5.543.502; col. 4, línea 22 hasta la col. 5, línea 36 de la Patente de Estados Unidos Núm. 5.171.844; col. 2, líneas 55-68, figura 2 y el Ejemplo 1 de la Patente de Estados Unidos Núm. 5.112.950; col. 2, línea 2 hasta la col. 19, línea 21 y la tabla 2 de la Patente de Estados Unidos Núm. 4.868.112; col. 2, línea 1 hasta col. 3, línea 19, col. 3, línea 40 hasta col. 4, línea 67, col. 7, línea 43 hasta col. 8, línea 26 y col. 11, línea 5 hasta la col. 13, línea 39 de la Patente de Estados Unidos 7.041.635; o col. 4, líneas 25-53, de la Patente de Estados Unidos Núm. 6.458.563. En algunas realizaciones, un FVIII con el dominio B suprimido tiene una delección de la mayor parte del dominio B, pero aún contiene secuencias amino-terminales del dominio B que son esenciales para el procesamiento proteolítico *in vivo* del producto de traducción primario en dos cadenas polipeptídicas, como se describe

en el documento WO 91/09122. En algunas realizaciones, un FVIII con el dominio B suprimido se construye con una delección de los aminoácidos 747-1638, es decir, virtualmente una delección completa del dominio B. Hoeben R. C. et al. *J. Biol. Chem.* 265 (13): 7318 - 7323 (1990). Un Factor VIII con el dominio B suprimido también puede contener una delección de los aminoácidos 771-1666 o los

5 aminoácidos 868-1562 del FVIII. Meulien P. et al. *Protein Eng.* 2(4): 301-6 (1988). Las delecciones adicionales del dominio B que forman parte de la invención incluyen: delección de los aminoácidos 982 a 1562 o 760 a 1639 (Toole et al., *Proc. Natl. Acad. Sci. USA* (1986) 83, 5939-5942)), 797 a 1562 (Eaton, et al. *Biochemistry* (1986) 25:8343-8347)), 741 a 1646 (Kaufman (Solicitud Publicada PCT Núm. WO 87/04187)), 747-1560 (Sarver, et al., *DNA* (1987) 6:553-564)), 741 a 1648 (Pasek (Solicitud PCT Núm. 10 88/00831)), o 816 a 1598 o 741 a 1648 (Lagner (Behring Inst. Mitt. (1988) Núm. 82: 16- 25, documento EP 295597)). En una realización particular, el FVIII con el dominio B suprimido comprende una delección de los restos de aminoácidos 746 a 1648 del FVIII maduro. En otra realización, el FVIII con el dominio B suprimido comprende una delección de los restos de aminoácidos 745 a 1648 del FVIII maduro. En algunas realizaciones, el FVIII BDD comprende FVIII monocatenario que contiene una 15 delección en los aminoácidos 765 a 1652 correspondientes al FVIII de longitud completa maduro (también conocido como rVIII-SingleChain y AFSTYLA®). Véase la Patente de Estados Unidos Núm. 7.041.635.

[0257] En otras realizaciones, el FVIII BDD incluye un polipéptido de FVIII que contiene fragmentos del dominio B que conservan uno o más sitios de glicosilación unidos a N, p. ej., restos 757, 784, 828, 900, 963 u opcionalmente 943, que corresponden a la secuencia de aminoácidos de la secuencia de FVIII 20 de longitud completa. Los ejemplos de los fragmentos del dominio B incluyen 226 aminoácidos o 163 aminoácidos del dominio B como describen Miao, H. Z., et al., en *Blood* 103(a): 3412-3419 (2004), Kasuda, A, et al., *J. Thromb. Haemost.* 6:1352-1359 (2008), y Pipe, S. W., et al., *J. Thromb. Haemost.* 9: 2235-2242 (2011) (es decir, se conservan los primeros 226 aminoácidos o 163 aminoácidos del dominio B). En otras realizaciones más, el FVIII BDD comprende adicionalmente una mutación puntual en el 25 resto 309 (de Phe a Ser) para mejorar la expresión de la proteína FVIII BDD. Véase Miao, H. Z. et al., *Blood* 103(a): 3412-3419 (2004). En otras realizaciones más, el FVIII BDD incluye un polipéptido de FVIII que contiene una porción del dominio B, pero que no contiene uno o más sitios de escisión de furina (p. ej., Arg1313 y Arg 1648). Véase Pipe, S.W, et al., *J. Thromb. Haemost.* 9: 2235-2242 (2011). En algunas realizaciones, el FVIII BDD comprende FVIII monocatenario que contiene una delección en 30 los aminoácidos 765 a 1652 correspondientes al FVIII de longitud completa maduro (también conocido como rVIII-SingleChain y AFSTYLA®). Véase la Patente de Estados Unidos Núm. 7.041.635. Cada una de las delecciones anteriores se puede realizar en cualquier secuencia de FVIII.

[0258] Se conoce una gran cantidad de variantes funcionales de FVIII, como se describe anteriormente y a continuación. Además, se han identificado cientos de mutaciones no funcionales en el FVIII en pacientes con hemofilia, y se ha determinado que el efecto de estas mutaciones en la función del FVIII se debe más a dónde se encuentran dentro de la estructura tridimensional del FVIII que a la naturaleza de la sustitución (Cutler et al., *Hum. Mutat.* 19:274-8 (2002)), incorporada a la presente memoria como referencia en su totalidad. Además, las comparaciones entre el FVIII de seres humanos y otras especies han identificado restos conservados que probablemente sean necesarios para la función (Cameron et al., *Thromb. Haemost.* 79: 317-22 (1998); documento US 6.251.632), incorporado a la presente memoria como referencia en su totalidad.

[0259] En algunas realizaciones, el polipéptido de FVIII comprende una variante de FVIII o fragmento del mismo, en donde la variante de FVIII o fragmento del mismo tiene actividad de FVIII. En algunas realizaciones, el casete genético codifica un polipéptido de FVIII de longitud completa. En otras realizaciones, el casete genético codifica un polipéptido de FVII con el dominio BDD suprimido (BDD), en donde todo o una porción del dominio B de FVIII está suprimido. En una realización particular, el casete genético codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con SEQ ID NO: 106, 107, 109, 110, 111, o 112. En algunas realizaciones, el casete genético codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 17 o un fragmento de la misma. En algunas realizaciones, el casete genético codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 106 o un fragmento de la misma. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 107. En algunas realizaciones, el casete genético codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 109 o un fragmento de la misma. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 16. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 109.

[0260] En algunas realizaciones, el casete genético de la divulgación codifica un polipéptido de FVIII que comprende un péptido señal o un fragmento del mismo. En otras realizaciones, el casete genético codifica un polipéptido de FVIII que carece de un péptido señal. En algunas realizaciones, el péptido señal comprende los aminoácidos 1-19 de SEQ ID NO: 17.

- 5 [0261] En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos tiene un codón optimizado. En ciertas realizaciones, el casete genético comprende una secuencia de nucleótidos que se describe en la Solicitud Internacional Núm. PCT/US2017/015879, que se incorpora como referencia en su totalidad. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que codifica un
- 10 polipéptido de FVIII, en donde la secuencia de nucleótidos tiene un codón optimizado. En ciertas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con una secuencia de nucleótidos seleccionada entre SEQ ID NO: 1-14. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al
- 15 menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 71. En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 19.

20 *i. Secuencia de Nucleótidos con Codones Optimizados que Codifican Polipéptidos de FVIII*

- [0262] En algunas realizaciones, una molécula de ácido nucleico de la presente divulgación comprende una primera ITR, una segunda ITR, y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica, en donde la primera ITR y la segunda ITR se obtienen de un genoma de AAV, y en donde el casete genético comprende una secuencia de nucleótidos con
- 25 codones optimizados que codifica un polipéptido de FVIII. En algunas realizaciones, la secuencia de nucleótidos con codones optimizados codifica un polipéptido de FVIII de longitud completa. En otras realizaciones, la secuencia de nucleótidos con codones optimizados codifica un polipéptido de FVIII con el dominio B suprimido (BDD), en donde todo o una porción del dominio B de FVIII está suprimido. En una realización particular, la secuencia de nucleótidos con codones optimizados codifica un polipéptido
- 30 que comprende una secuencia de aminoácidos que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos

aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con SEQ ID NO: 17 o un fragmento de la misma. En una realización, la secuencia de nucleótidos con codones optimizados codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 17 o un fragmento de la misma.

- 5 **[0263]** En algunas realizaciones, la secuencia de nucleótidos con codones optimizados codifica un polipéptido de FVIII que comprende un péptido señal o un fragmento del mismo. En otras realizaciones, la secuencia con codones optimizados codifica un polipéptido de FVIII que carece de un péptido señal. En algunas realizaciones, el péptido señal comprende los aminoácidos 1-19 de SEQ ID NO: 17.

- 10 **[0264]** En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3 o (ii) los nucleótidos 58-1791 de SEQ ID NO: 4; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En una realización particular, la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 3. En otra realización, la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 4. En otras realizaciones, la primera secuencia de nucleótidos comprende los nucleótidos 58-1791 de SEQ ID NO: 3 o los nucleótidos 58-1791 de SEQ ID NO: 4.

- 25 **[0265]** En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1-1791 de SEQ ID NO: 3 o (ii) los nucleótidos 1-1791 de SEQ ID NO: 4; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En una realización, la primera secuencia de nucleótidos comprende los nucleótidos 1-1791 de SEQ ID NO: 3 o los nucleótidos 1-1791 de SEQ ID NO: 4. En otra realización, la segunda secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos

99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4. En una realización particular, la segunda secuencia de nucleótidos comprende los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4. En otra realización más, la segunda secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 3 o 1792-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En una realización particular, la segunda secuencia de nucleótidos comprende los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 3 o 1792-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o fragmento del dominio B).

[0266] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 5 o (ii) 1792-4374 de SEQ ID NO: 6; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En ciertas realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 5. En otras realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 6. En una realización particular, la segunda secuencia de ácido nucleico comprende los nucleótidos 1792-4374 de SEQ ID NO: 5 o 1792-4374 de SEQ ID NO: 6. En algunas realizaciones, la primera secuencia de ácido nucleico conectada a la segunda secuencia de ácido nucleico enumerada más arriba tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 5 o los nucleótidos 58-1791 de SEQ ID NO: 6. En otras realizaciones, la primera secuencia de ácido nucleico conectada a la segunda secuencia de ácido nucleico enumerada más arriba tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1-1791 de SEQ ID NO: 5 o los nucleótidos 1-1791 de SEQ ID NO: 6.

[0267] En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda

5 secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o (ii) 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o

10 fragmento del dominio B); y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En ciertas realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o fragmento

15 del dominio B). En otras realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En una realización particular, la segunda secuencia de ácido nucleico comprende los nucleótidos 1792-

20 2277 y 2320-4374 de SEQ ID NO: 5 o 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 o 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En algunas realizaciones, la primera secuencia de ácido nucleico conectada a la segunda secuencia de ácido nucleico enumerada más arriba tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al

25 menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 5 o los nucleótidos 58-1791 de SEQ ID NO: 6. En otras realizaciones, la primera secuencia de ácido nucleico conectada a la segunda secuencia de ácido nucleico enumerada más arriba tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1-1791 de SEQ ID NO: 5 o los nucleótidos 1-

30 1791 de SEQ ID NO: 6.

[0268] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera

secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 1, (ii) los nucleótidos 58-1791 de SEQ ID NO: 2, (iii) los nucleótidos 58-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 58-1791 de SEQ ID NO: 71 ; y en donde la porción N-terminal y la porción C-terminal
5 juntas tienen una actividad de polipéptido de FVIII. En otras realizaciones, la primera secuencia de nucleótidos comprende los nucleótidos 58-1791 de SEQ ID NO: 1, los nucleótidos 58-1791 de SEQ ID NO: 2, (iii) los nucleótidos 58-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 58-1791 de SEQ ID NO: 71.

[0269] En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un
10 polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1-1791 de SEQ ID NO: 1, (ii) los
15 nucleótidos 1-1791 de SEQ ID NO: 2, (iii) los nucleótidos 1-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 1-1791 de SEQ ID NO: 71 ; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En una realización, la primera secuencia de nucleótidos comprende los nucleótidos 1-1791 de SEQ ID NO: 1, los nucleótidos 1-1791 de SEQ ID NO: 2, (iii) los nucleótidos 1-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 1-1791 de SEQ ID NO: 71. En otra
20 realización, la segunda secuencia de nucleótidos conectada a la primera secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 1, 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71. En una realización particular, la segunda secuencia de
25 nucleótidos conectada a la primera secuencia de nucleótidos comprende (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71. En otras realizaciones, la segunda secuencia de nucleótidos conectada a la primera secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al
30 menos 99% de identidad de secuencia con (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71. En una realización, la segunda secuencia de nucleótidos comprende (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos

1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71.

[0270] En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un

polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de

ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia

de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda

secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos

98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii)

los nucleótidos 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv)

los nucleótidos 1792-4374 de SEQ ID NO: 71 ; y en donde la porción N-terminal y la porción C-terminal

juntas tienen una actividad de polipéptido de FVIII. En una realización particular, la segunda secuencia

de ácido nucleico comprende (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-

4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-

4374 de SEQ ID NO: 71. En algunas realizaciones, la secuencia con codones optimizados que codifica

un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia

de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda

secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la

segunda secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%,

al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-2277 y 2320-4374

de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos

1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID

NO: 71 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 1, los nucleótidos 1792-4374 de SEQ ID

NO: 2, los nucleótidos 1792-4374 de SEQ ID NO: 70, o los nucleótidos 1792-4374 de SEQ ID NO: 71

sin los nucleótidos que codifican el dominio B o fragmento del dominio B); y en donde la porción N-

terminal y la porción C-terminal juntas tienen una actividad de polipéptido de FVIII. En una realización,

la segunda secuencia de ácido nucleico comprende (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ

ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-2277

y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71 (es

decir, los nucleótidos 1792-4374 de SEQ ID NO: 1, los nucleótidos 1792-4374 de SEQ ID NO: 2, los

nucleótidos 1792-4374 de SEQ ID NO: 70, o los nucleótidos 1792-4374 de SEQ ID NO: 71 sin los

nucleótidos que codifican el dominio B o fragmento del dominio B).

[0271] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un

polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad

de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 1. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 1. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 1. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o nucleótidos 1 a 4374 de SEQ ID NO: 1.

[0272] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 2. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 2. En otras realizaciones, la secuencia de ácido nucleico tiene al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 2. En otras realizaciones, la secuencia de nucleótidos comprende nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 2 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 2 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 2. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 2 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 2 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 2.

[0273] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 70. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91 %, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 70. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 70. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 1-4374 of SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 70.

[0274] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 71. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91 %, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 71. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 71 sin los

nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 71. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 71.

- 5 **[0275]** En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 3. En otras
- 10 realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o
- 15 fragmento del dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 3. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los
- 20 nucleótidos 58 a 4374 de SEQ ID NO: 3. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 3.

- [0276]** En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad
- 25 de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 4. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que
- 30 tiene al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos

96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 4. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 4. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 4.

[0277] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 5. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 89%, al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 5. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 5. En otras realizaciones más, la secuencia de nucleótidos comprende nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 5.

[0278] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 6. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%,

al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o fragmento del dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 89%, al menos 90%, al menos 91 %, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con SEQ ID NO: 6. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 6. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 6.

[0279] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de ácido nucleico que codifica un péptido señal. En ciertas realizaciones, el péptido señal es un péptido señal de FVIII. En algunas realizaciones, la secuencia de ácido nucleico que codifica un péptido señal tiene codones optimizados. En una realización particular, la secuencia de ácido nucleico que codifica un péptido señal tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% de identidad de secuencia con (i) los nucleótidos 1 a 57 de SEQ ID NO: 1; (ii) los nucleótidos 1 a 57 de SEQ ID NO: 2; (iii) los nucleótidos 1 a 57 de SEQ ID NO: 3; (iv) los nucleótidos 1 a 57 de SEQ ID NO: 4; (v) los nucleótidos 1 a 57 de SEQ ID NO: 5; (vi) los nucleótidos 1 a 57 de SEQ ID NO: 6; (vii) los nucleótidos 1 a 57 de SEQ ID NO: 70; (viii) los nucleótidos 1 a 57 de SEQ ID NO: 71 ; o (ix) los nucleótidos 1 a 57 de SEQ ID NO: 68.

[0280] SEQ ID NO: 1-6, 70, y 71 son versiones optimizadas de SEQ ID NO: 16, la secuencia de nucleótidos de FVIII de partida o "parental" o "de tipo salvaje". SEQ ID NO: 16 codifica un FVIII humano con el dominio B suprimido. Si bien SEQ ID NO: 1-6, 70, y 71 se obtienen de una forma de FVIII con el dominio B suprimido específica (SEQ ID NO: 16), se debe entender que la presente divulgación también incluye versiones optimizadas de ácidos nucleicos que codifican otras versiones de FVIII. Por ejemplo, otras versiones de FVIII pueden incluir FVIII de longitud completa, otras delecciones del dominio B de FVIII (descritas en la presente memoria), u otros fragmentos de FVIII que conservan la actividad de FVIII.

[0281] En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos enumerada en las Tablas 2A-2F. En una realización, el casete genético

comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2A. En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2B. En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2C. En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2D. En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2E. En una realización, el casete genético comprende una construcción de FVIII, que incluye una secuencia de polinucleótidos expuesta en la Tabla 2F

- 10 **[0282]** En ciertas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos
- 15 aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 179, 182, 189, o 194. En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al
- 20 menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 179. En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente
- 25 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 182. En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos
- 30 aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 189.

En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 194. En algunas realizaciones, la molécula de ácido nucleico aislada conserva la capacidad de expresar una proteína de FVIII funcional.

Tabla 2A: Ejemplo de construcción de AAV-FVIII (nucleótidos 1-6526; SEQ ID NO: 110)

Descripción	Secuencia
ITR 5' (repetición terminal invertida de AAV2 extremo 5') (SEQ ID NO:111)	1 -- CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACC TTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCC ATCACTAGGGGTTCCT -- 130
Secuencia del Esqueleto del Plásmido (PBS)-1 (SEQ ID NO:112)	131 -- GCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTC TCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCG GCCGCCTCGACGACTAGCGTTTAATTAA -- 272
TTPp (promotor específico de hígado) (SEQ ID NO:113)	273 -- ACGCGTGTCTGTCTGCACATTTCTGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTA GGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAA TCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTGGAAGG AGGGGGTATAAAAGCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCT G -- 501
PBS-2 (SEQ ID NO:114)	502 -- AG -- 503
Intrón Sintético (SEQ ID NO:115)	504 -- GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGC GTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTCTCCACAG --609
PBS-3 (SEQ ID NO:116)	610 -- CTAGCGCCACC -- 620
FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto la versión 6 de FVIII con codones optimizados que contiene XTEN144; la	621 -- ATGCAGATTGAGCTGTCCACTTGTTTCTTCCCTGTGCCTCCTGCGCTTCTGTTTCTCC GCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGC GACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTC CCCTTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTCTGAGTTCACTGACCAC CTGTTCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATT CAAGCTGAAGTGACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCACCCC GTGTCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTAC GACGACCAGACTAGCCAGCGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCG CATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTG

secuencia XTEN
está marcada por
un doble
subrayado (SEQ
ID NO:118))

TGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGG
CTGATTGGTGCACCTTCTCGTGTGCCGCGAAGGTTTCGCTCGCTAAGGAAAAGACCCAG
ACCCTCCATAAGTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCAT
TCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGG
CCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGT
TGTCACAGAAAGTCCGTGTACTGGCACGTTCATCGGCATGGGCACTACGCCTGAAGTG
CACTCCATCTTCTTGGGAAGGGCACACCTTCTCGTGCACAACACCGCCAGGCCTCT
CTGGAAATCTCCCCGATTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGG
CAGTTCTCTTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTAC
GTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAG
GCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCTGCGGTTTCGAT
GACGACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAA
ACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTG
CTGGCACCCGATGACCGGTCTGTACAAGTCCAGTATCTGAACAATGGTCCGCAGCGG
ATTGGCAGAAAGTACAAGAAAGTGCAGTTTCATGGCGTACACTGACGAAACGTTTAAAG
ACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAG
GTGCGCGATACCCGTCTCATCTTCAAAAACAGGCCTCCCGGCCTTACAACATC
TACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGCGCCTGCCGAAGGGC
GTCAAGCACCTGAAAGACTTCCCTATCTGCGGGCGAAATCTTCAAGTATAAGTGG
ACCGTCACCGTGGAGGACGGGCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTAC
TACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTG
CTGATCTGTACAAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAG
CGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAA
AACATCCAGAGGTTCTCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCTTGAGTTT
CAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTTCGACTCGCTGCAACTG
AGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCAGACT
GACTTCTCTGAGCGTGTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAA
GATACCTTGACCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTGATGGAGAAC
CCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACT
GCCCTGTGTAAGGTGTCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCC
TACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACCGCATCGAGCCGCGCAGC
TTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGG
AGCGAGCCAGCCACATCTGGGTTCGAAACGCCAGGCACAAGTGAGTCTGCAACTCCC
GAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCC
GAGAGCGCTACACCAGAAAGCGGACCCGGAACCAAGTACCGAACCTAGCGAGGGCTCT
GCTCCGGGCAGCCAGCCGGCTCTCTACATCCACGGAGGAGGGCACTTCCGAATCC
GCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCA
GGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCT
ACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCACCAGCACTGAAGAAGGTGCC
TCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGACCACTCCAA
TCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAA
GATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAA
ACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCC
AGCCCTCACGTCTTTCGGAACCGGGGCCAGAGCGGATCGGTGCCTCAGTTCAAGAAA
GTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAA
CTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAAC
ATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTG
ATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAG
CCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG
GATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGAT
GTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCTGAACCCA
GCGCATGGACGCCAGGTCACCGTCCAGGAGTTTGTCTGTTCTTACCATTTTTGAC

	GAAACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGC AATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAAC GGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATTCGG TGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGT CACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTAC CCCGGGGTGTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTG GAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTAC TCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAG ATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCCGCTTGCACTAC TCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGAC CTCCTGGCCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAAGTTC TCCTCCCTGTACATCTCGCAATTATCATCATGTACAGCCTGGACGGGAAGAAGTGG CAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGAT TCCTCCGGCATTAAGCACAAACATCTTCAACCCACCGATCATAGCCAGATATATTAGG CTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACATCATGGGGTGC GACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCC CAGATCACCGCGAGCTCCTACTTCACTAACATGTTTCGCCACCTGGAGCCCCCTCCAAG GCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCG AAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACC CAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCTGATTAGCAGC AGCCAGGACGGGCACCAAGTGGACCCTGTTCTTCCAAAACGGAAGGTCAAGGTGTTT CAGGGGAACCAAGGACTCGTTACACCCGTGGTGAACCTCCCTGGACCCCCCACTGCTG ACGCGGTACTTGAGGATTATCCTCAGTCTGGGTCCATCAGATTGCATTGCGAATG GAAGTCTTGGGCTGCGAGGCCAGGACCTGTACTGA -- 5444
PBS-4 (SEQ ID NO:119)	5445 -- ATCAGCCTGAGCTCGCTGA -- 5463
WPRe (elemento regulador post-transcripcional del virus de marmota mutado) (SEQ ID NO:120)	5464 -- TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTAT TGCTTCCCGTATGGCTTTTCAATTTCTCCTTGTATAAATCCTGGTTGCTGTCTCT TTATGAGGAGTTGTGGCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGC TGACGCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGAC TTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCCG CTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTGCGGGGAA ATCATCGTCCTTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGAC GTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCTTCCCGCGGCCCT GCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT CTCCCTTTGGGCGCCTCCCCGCTG -- 6058
PBS-5 (SEQ ID NO:121)	6059 -- ATCAGCCT -- 6066
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	6067 -- CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCT TGACCCTGGAAGGTGCCACTCCCACTGTCTTTCCTAATAAAATGAGGAAATTGCAT CGCATTGTCTGAGTAGGTGTATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCA AGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA --6277
PBS-6 (SEQ ID NO:123)	6278 -- TGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGAT

	ATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGG GCCGC -- 6396
ITR 3' (repetición terminal invertida de AAV2 extremo 3') (SEQ ID NO:124)	6397 -- AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTG AGGCCGGGCGACCAAAGGTCGCCCCGACGCCGGGCTTTGCCCGGGCGGCCTCAGTGA GCGAGCGAGCGCGCAG -- 6526

Tabla 2B: Ejemplo de construcción de B19-FVIII que porta ITR de B19d135 (nucleótidos 1-6762; SEQ ID NO: 179)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 180)	1 - CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTG ATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCAGAGCCAACCTAATTCCGGAAG TCCCGCCCACCGGAAGTGACGTACAGGAAATGACGTACAGGAAATGACGTAATTG TCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTT GGTGTCTTCTTTTAAATTTT -- 248
TTPp (promotor específico de hígado) (SEQ ID NO:113)	391 - ACGCGTGTCTGTCTGCACATTTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTA GGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAA TCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGG AGGGGGTATAAAAGCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCT G -- 619
Intrón Sintético (SEQ ID NO:115)	622 - GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGC GTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTCTCCACAG -- 727
FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto para la versión 6 de FVIII con codones optimizados Que contiene XTEN144; la secuencia XTEN está marcada por un <u>doble</u> <u>subrayado</u> (SEQ ID NO:118))	739 - ATGCAGATTGAGCTGTCCACTTGTCTTCTCCTGTGCCTCCTGCGCTTCTGTTTCTCC GCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGC GACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTC CCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTGCGAGTTCACTGACCAC CTGTTCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATT CAAGCTGAAGTGTACGACACCGTGGTGTACCCCTGAAGAACATGGCGTCCCACCCC GTGTCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTAC GACGACCAGACTAGCCAGCGGGAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCG CATACTTACGTGTGGCAAGTCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTG TGCTTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGG CTGATTGGTGCATTCTCGTGTGCCGCGAAGGTTTCGCTCGCTAAGGAAAAGACCCAG ACCCTCCATAAGTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCAT TCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGG CCTAAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGT TGTCACAGAAAGTCCGTGTACTGGCACGTTCATCGGCATGGGCACTACGCCTGAAGTG CACTCCATCTTCTGGAAGGGCACACCTTCCTCGTGCACCAACACCGCCAGGCCTCT CTGGAAATCTCCCCGATTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGG CAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTAC GTGAAGGTGGACTCATGCCCGGAAGAACCCTCAGTTGCGGATGAAGAACAACGAGGAG GCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGAT GACGACAACAGCCCCAGCTTCATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCAAA ACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTG

	CTGGCACCCGATGACCGGTGCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGG ATTGGCAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGACGAAACGTTTAAAG ACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAG GTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACATC TACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGC GTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGG ACCGTCACCGTGGAGGACGGGCCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTAC TACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTG CTGATCTGTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAG CGAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAA AACATCCAGAGGTTCTTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTT CAGGCCTCGAATATCATGCACTCGATTAAACGGTTACGTGTTGACTCGCTGCAACTG AGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACT GACTTCTTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAA GATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTGATGGAGAAC CCGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACT GCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCC TACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGC TTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGG AGCGAGCCAGCCACATCTGGGTTCGGAAACGCCAGGCACAAGTGAGTCTGCAACTCCC GAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTTC GAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCT GCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCC GCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCA GGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCT ACGTCCACTGAGGAAGGGTCACCAGCGGGGCTCGCCCACCAGCACTGAAGAAGGTGCC TCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAA TCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAA GATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTTCCAAAAGAAA ACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCC AGCCCTCACGTCTTTCGGAACCGGGGCCAGAGCGGATCGGTGCCTCAGTTCAAGAAA GTGGTGTTCAGGAGTTACCGACGGCAGCTTACCCAGCCGCTGTACCGGGGAGAA CTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAAC ATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTG ATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAG CCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG GATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGAT GTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCA GCGCATGGACGCCAGGTACCCGTCCAGGAGTTTGCTCTGTTCTTACCATTCTTTGAC GAAACTAAGTCTTGGTACTTACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGC AATATCCAGATGGAAGATCCGACTTTCAGGAGAATATAGATTCCACGCCATCAAC GGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCCGG TGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGT CACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTAC CCCGGGGTGTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTG GAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCTGGTGTAC TCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAG ATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTAC TCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGAC CTCCTGGCCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTC TCCTCCCTGTACATCTCGCAATTATCATCATGTACAGCCTGGACGGGAAGAAGTGG CAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGAT
--	---

	TCCTCCGGCATTAAGCACAAACATCTTCAACCCACCGATCATAGCCAGATATATTAGG CTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACATCATGGGGTGC GACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCC CAGATCACCGCGAGCTCCTACTTCACTAACATGTTGCCACCTGGAGCCCCCTCCAAG GCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCG AAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACC CAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGC AGCCAGGACGGGCACCAGTGGACCCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTT CAGGGGAACCAGGACTCGTTACACCCGTGGTGAACCTCCCTGGACCCCCCACTGCTG ACGCGGTACTTGAGGATTATCCTCAGTCCCTGGGTCCATCAGATTGCATTGCGAATG GAAGTCTTGGGCTGCGAGGCCAGGACCTGTACTGA -- 5562
WPRe (elemento regulador post-transcripcional del virus de la hepatitis de marmota) (SEQ ID NO:120)	5582 - TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTAT TGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCT TTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGC TGACGCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGAC TTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCTGCCTTGCCCG CTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAA ATCATCGTCCCTTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGAC GTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCT GCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT CTCCCTTTGGGCCGCTTCCCCGCTG -- 6176
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	6185 - CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCT TGACCCTGGAAGGTGCCACTCCCACTGTCCTTTTCTAATAAAATGAGGAAATTGCAT CGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCA AGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA -- 6395
ITR 3' repetición terminal invertida (SEQ ID NO: 181)	6515 - AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTGCCTGCCGGTAGGCGGGAC TTCCGGTACAAGATGGCGGACAATTACGTCATTTCTGTGACGTCATTTCTGTGAC GTCACTTCCGGTGGCGGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGG GGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACC CCAAGCAAGCTGGCCCAGAG -- 6762
Secuencia de Longitud Completa (SEQ ID NO: 179)	
CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGT CAACCCCAAGCGCTGGCCAGAGCCAACCTAATTCGGGAAGTCCCGCCACCGGAAGTGACGTACAGGAA ATGACGTACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGC GGCATCTGATTTGGTGTCTTCTTTTAAATTTTGC GGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGC GATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCG CGGCCGCTCGACGACTAGCGTTTAAATTAACGCGTGTCTGTCTGCACATTTCTGTAGAGCGAGTGTTCGAT ACTCTAATCTCCCTAGGCAAGGTTCAATTTGTGTAGGTTACTTATCTCCTTTTGTGACTAAGTCAATAA TCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGC CCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGGG CCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTC TCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCACTTGTCTTCTCTGTGCCTCCTGCGCTTCTGTTTC TCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAA CTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAG	

AAAACCCCTCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCACCTTGATGGGCCTC
 CTGGGACCGACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCC
 GTGTCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGC
 CAGCGGGAAAAGGAGGACGATAAAGTGTTCGCGGGCGCTCGCATACTTACGTGTGGCAAGTCTGAAGGAA
 AACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGAC
 CTGAACAGCGGGCTGATTGGTGCACCTCTCGTGTGCCGCGAAGGTTTCGCTCGCTAAGGAAAAGACCCAGACC
 CTCCATAAGTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCG
 CTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAAT
 CGGTCACTGCCCGGGCTCATCGGTTGTACAGAAAGTCCGTGTACTGGCACGTATCGGCATGGGCACTACG
 CCTGAAGTGCATCCATCTTCTGGAAGGGCACACCTTCTCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAA
 ATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTCTGCCAC
 ATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTG
 CGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCTGTGCGG
 TTCGATGACGACAACAGCCCCAGCTTCATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCAAAACCTGGGTG
 CACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCGGTTGGTGCTGGCACCCGATGACCGGTCTGTAC
 AAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTAC
 ACTGACGAAACGTTTAAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGA
 GAGGTGCGCGATAACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACATCTACCCTCACGGA
 ATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATC
 CTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGCGATCCTAGG
 TGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTG
 CTGATCTGCTACAAAGAGTCCGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTG
 TTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCTCCCAAACCCCGCA
 GGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTTCGAC
 TCGCTGCAACTGAGCGTGTGCCTCCATGAAGTTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGAC
 TTCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAAGATGGTGTACGAAGATACCCTGACCCTGTTT
 CCTTTCTCCGGCGAAACGGTGTTTCATGTGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGC
 GACTTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGGAGACTACTAC
 GAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGC
 CAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGG
 TCGGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGC
 TCCGAGACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACAGTACCGAACCTAGCGAG
 GGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCG
 GAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCG
 GAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTACCAGCGGGCTCGCCCACC
 AGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTC
 CAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAATGAAGAAGGAAGATTTTCGATATC
 TACGACGAGGACGAAAATCAGTCCCCCTCGCTCATTTCAAAAAGAAAACCTAGACACTACTTTATCGCCGCGGTG
 GAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCTTCGGAACCGGGGCCAGAGCGGATCGGTG
 CCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAA
 CTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTC
 CGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCTCGATCTCATACGAGGAGGACCAGCGCCAAGGC
 GCCGAGCCCCGCAAGAACTTCGTCAAGCCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATG
 GCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTC
 CATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCCTGAACCCAGCGCATGGACGCCAGGTC
 ACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCTTGGTACTTCACCGAGAATATG
 GAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCAC
 GCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCCGTGGTAC
 TTGCTGTCAATGGGATCGAACGAAAACATTCCTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAG
 AAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTTCGAAACTGTGGAGATGCTGCCGTCC
 AAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCTG
 GTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCA

AGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCC
ACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCCTATGATTATCCACGGAATTAAGACCCAG
GGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAG
TGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCCCTCCGGCATT
AAGCACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGC
TCAACTCTTCGGATGGAACCTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAG
GCTATTAGCGACGCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCCGCCACCTGGAGCCCCCTCCAAG
GCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAA
GTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATG
TATGTGAAGGAGTTCTTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCAAAACGGAAAG
GTCAAGGTGTTCCAGGGGAACCAGGACTCGTTACACCCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACG
CGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCTTGGGCTGCGAG
GCCCAGGACCTGTACTGAATCAGCCTGAGCTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAA
GATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATC
ATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGG
AGTTGTGGCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGTGCTGACGCAACCCCCCACTGGTTGGG
GCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCA
TCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTGCGG
GGAAATCATCGTCTTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCT
ACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGC
GTCTTCGCCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGATCAGCCTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCTCTTCTTAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCT
ATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTT
AAATTAGGGATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAAGAAGACACCAAATCAGATGCCGCCGGT
CGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCAATTTCTGTGACGTCAATTTCTGT
TGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGC
CACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCGAGAG

Tabla 2C: Ejemplo de construcción de GPV-FVIII que porta ITR de GPVd162 (nucleótidos 1-6830; SEQ ID NO: 182)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 183)	1 - CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTGAGCCCAAT AGTTAAGCCGGAACACGTACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAA CACGTGACAGGAAGCACGTGACCGGAACCTACGTACCGGATGTGCGTCACCGGAAGCA TGTGACCGGAACCTGCGTCACTTCCCCCTCCCCCTGATTGGCTGGTTTGAACGAACGAA CCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGCAGCGCCAGGAAGTG -- 282
TTPp (promotor específico de hígado) (SEQ ID NO:113)	425 - ACGCGTGTCTGTCTGCACATTTCTGATAGCGAGTGTTCGATACTCTAATCTCCCTAG GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATC AGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTGGAAGGAGG GGGTATAAAAGCCCCCTTACCAGGAGAAGCCGTACACAGATCCACAAGCTCCTG -- 653
Intrón Sintético (SEQ ID NO:115)	656 - GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCG TGCTTGAATTACTGACACTGACATCCACTTTTCTTTTCTCCACAG -- 761

FVIIIco6XTEN
(SEQ ID
NO:117) (marco
de lectura
abierto para la
versión 6 de
FVIII con
codones
optimizados que
contiene
XTEN144; la
secuencia XTEN
está marcada
por un doble
subrayado (SEQ
ID NO:118))

773 -

ATGCAGATTGAGCTGTCCACTTGTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCG
CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGA
CCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCC
TTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGAGTTCACTGACCACCTGT
TCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAGC
TGAAGTGTACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCC
CTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACC
AGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTA
CGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACT
TACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTG
CACTTCTCGTGTGCCGCGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAA
GTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAG
AACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATA
CAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGTTGTACAGAAAGTC
CGTGTACTGGCACGTCACTCGGCATGGGCACTACGCCTGAAGTGCATCCATCTTCCTG
GAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGA
TTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGGCAGTTTCCTTCTTCTTG
CCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGC
CCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACG
ATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGATGACGACAACAGCCCCAGCTT
CATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCG
GCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCTGT
ACAAGTCCCAGTATCTGAACAATGGTCCGCGAGCGGATTGGCAGAAAGTACAAGAAAGT
GCGGTTTCATGGCGTACACTGACGAAACGTTTAAAGACCCGGGAGGCCATTCAACATGAG
AGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCT
TCAAAAACAGGCCTCCCGGCCCTTACAACATCTACCCTCACGGAATCACCGACGTGCG
GCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATC
CTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCCACCA
AGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGA
CCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAA
CGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATG
AAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTTCCTCCCAACCCCGCAGG
AGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGT
TACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACA
TTCTGTCCATCGGCGCCAGACTGACTTCCTGAGCGTGTCTTTTCCGGTTACACCTT
TAAGCACAAGATGGTGTACGAAGATAACCTGACCCTGTTCCCTTTCTCCGGCGAAACG
GTGTTTCATGTGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACT
TTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGG
AGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAAC
GCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCC
CTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTGCGAAACGCCAGGCACAAG
TGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAG
ACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAAGCGGACCCGGAACCAGTACCGAAC
CTAGCGAGGGCTCTGCTCCGGGCGAGCCAGCCGGCTCTCCTACATCCACGGAGGAGGG
CACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGC
AGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTG
CTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCACCCAGCACTGA
AGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACC
ACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGA
AGAAGGAAGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTTCCA
AAAGAAAACCTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATG

	TCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCA AGAAAGTGGTGTTCCAGGAGTTACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGG AGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGAGGAT AACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCCCTCCC TGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAA GCCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG GATGAGTTTACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATG TCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCAGC GCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAA ACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATA TCCAGATGGAAGATCCGACTTTCAGGAGAACTATAGATTCCACGCCATCAACGGGTA CATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTGGTAC TTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGT TCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGT GTTTCGAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTG ATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCTGGTGTACTCGAATAAGT GCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAG CGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATC AACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCCTA TGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACAT CTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGA AACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAGC ACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTA CTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGCGACCTGAACTCCTGCTCC ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACC CGAGCTCCT ACTTCACTAACATGTTTCGCCACCTGGAGCCCCCTCCAAGGCCAGGCTGCACTTGCAAGG ACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGAT TTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGA CCTCGATGTATGTGAAGGAGTTCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGAC CCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTTACA CCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTTCATCTC AGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCCAGGA CCTGTACTGA --5596
WPRE (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)	5616 - TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGAC GCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCG CTTTCCCCCTCCCTATTGCCACGGCGGAACATCATCGCCGCTGCCTTGCCCGCTGCTG GACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCG TCCTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC TCTGCGGCCTCTTCCGCGTCTTCGCTTTCGCCCTCAGACGAGTCGGATCTCCCTTTGG GCCGCTCCCCGCTG -- 6210
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	6219 - CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCCTT GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTAATAAAATGAGGAAATTGCATCG CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA -- 6429

ITR 3' repetición terminal invertida (SEQ ID NO: 184)	6549 -- CACTTCCTGGCGCGCAAAATATCCTCTTGCTCCTTGAGTCTCATTGGAGGGTTTCGTTCCG TTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCGGGTACATGCTTCCG GTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCTGTACGTGTTTCCGGT CACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTTAACATTGGGGCT GACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGAAACACGTCACCG -- 6830
Secuencia de Longitud Completa (SEQ ID NO: 182)	
CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGGGTCAGCCCAATAGTTAAGCCGGAAA CACGTACCCGGAAGTCACATGACCCGGAAGTCACGTGACCCGGAACACGTGACAGGAAGCACGTGACCCGGAAC TACGTACCCGGATGTGCGTCACCCGGAAGCATGTGACCCGGAACCTTGCGTCACTTCCCCCTCCCCTGATTGGCT GGTTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGC GCGCCAGGAAGTGCGGCA ATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACG CGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCGCTCGACGACTAGCGTTTAATTAAACGCGTGT CTGTCTGCACATTTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTTCATATTTGTGTAGG TTACTTATTCTCCTTTTGTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATC AGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGC TCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTG AATTACTGACACTGACATCCACTTTTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCAC TTGTTTCTTCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCT TTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAA GTCCTTCCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTCTGAGTTCACTGACCACCTGTTCAA CATCGCCAAGCCGCGCCACCTTGATGGGCCTCCTGGGACCGACCATTCAGCTGAAGTGTACGACACCGT GGTGATCACCTGAAGAATGCGGTCCACCCCGTGTCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGC CTCCGAAGGAGCTGAGTACGACGACAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCGCGGGCGG CTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTA CTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACCTCTCGTGTGCCG CGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTTCGATGA AGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCTCAGCCCGCGCCTG GCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTACAGAAAAGTC CGTGTACTGGCACGTATCGGCATGGGCACTACGCCTGAAGTGCACCTCCATCTTCTGGAAGGGCACACCTT CCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCAGACTCTGCT CATGGACCTGGGGCAGTTCTTCTCTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGT GAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGA CGATTTGACTGACTCCGAGATGGACGTGCTGCGGTTTCGATGACGACAACAGCCCCAGCTTCATCCAGATTTCG CAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGC CCCGTTGGTGCTGGCACCCGATGACCGGTGCTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGG CAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGACGAAAACGTTTAAGACCCGGGAGGCCATTCAACA TGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTGCGCGATAACCCTGCTCATCATCTTCAAAAACCA GGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCC GAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAAATCTTCAAGTATAAGTGGACCGTCAC CGTGGAGGACGGGCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGA ACGGGACCTGGCATCGGACTCATTGGACCGCTGCTGATCTGCTACAAAAGAGTCGGTGGATCAACGCGGCAA CCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCAC TGAAAACATCCAGAGGTTCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAA TATCATGCACTCGATTAACGGTTACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTA CTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCTGAGCGTGTCTTTTCCGGTTACACCTTTAAGCA CAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTTCATGTGATGGAGAA CCCGGGTCTGTGATTCTGGGATGCCACAACAGCGACTTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGT GTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTC CAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGCGCGCCAACATCAGAGAGCGCCACCCCTGA	

AAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAACGCCAGGCACAAGTGAGTCTGCAACTCCCGA
GTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAGAGCGCTACACCAGA
AAGCGGACCCGGAACCACTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATC
CACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAG
TGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACGTC
CACTGAGGAAGGGTCACCAGCGGGCTCGCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCTGTGTGAA
GAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCAT
CTCGGTGGAATGAAGAAGGAAGATTTGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTTCCA
AAAGAAAAC TAGACACTACTTTATCGCCGCGGTGGAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCA
CGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTCACCGA
CGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCCTGCTCGGTCCCTACATCCG
CGCGGAAGTGGAGGATAACATCATGGTGACCTTCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTC
CCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCCAACGAGAC
TAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGC
CTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACAC
CAACACCCTGAACCCAGCGCATGGACGCCAGGTACCGTCCAGGAGTTTGTCTGTCTTCCACATTTTTGA
CGAACTAAGTCTTGGTACTTACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATATCCAGATGGA
AGATCCGACTTTCAAGGAGAACTATAGATTTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCT
GGTCATGGCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTTCACTCCAT
TCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCC
CGGGGTGTTGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGA
GCACCTCCACGCGGGGATGTCCACCCTCTTCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGC
CTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCCG
CTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCT
GGCCCCATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAAGTTCTCCTCCCTGTACATCTCGCA
ATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCCTGAT
GGTCTTTTTTCGGCAACGTGGATTCCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAGATA
TATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACATCATGGGGTGCACCTGAA
CTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCTACTT
CACTAACATGTTGCCACCTGGAGCCCCCTCAAGGCCAGGCTGCACCTTGACGGGACGGTCAAATGCCTGGCG
GCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCAC
CACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGG
GCACCACTGGACCTGTTCTTCCAAAACGAAAGGTCAAGGTGTTCCAGGGGAACAGGACTCGTTACACCC
CGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCCTGGGTCCATCA
GATTGCATTGCGAATGGAAGTCCCTGGGCTGCGAGGCCAGGACCTGTACTGAATCAGCCTGAGCTCGCTGAT
CATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACG
CTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCC
TTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGC
ACTGTGTTTGCTGACGCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTC
GCTTTCCCCCTCCCTATTGCCACGGCGGAACATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACAGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTTCTTGGCTGCTCGCCTGTGTT
GCCACCTGGATTCTGCGCGGGACGTCCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCC
CGCGGCCCTGCTGCCGGCTCTGCGGCCCTTCCGCGCTCTTCGCTTCGCCCTCAGACGAGTCGGATCTCCCTT
TGGGCCGCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCC
CGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCAATCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGA
CAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTT
CTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCACT
TCCTGGCGCGCAAAATATCCTCTTGTCTTGTAGTCTCATTGGAGGGTTTCGTTTCGTTTCAACCAGCCAATCAG
GGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCA
CGTGCTTCTGTACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTT
AACTATTGGGCTGACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAAACACGTACCG

Tabla 2D: Ejemplo de construcción de B19-FVIII que porta ITR de B19 de longitud completa
(nucleótidos 1-7032; SEQ ID NO: 189)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 185)	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGG ACAATTACGTCATTTCTGTGACGTCATTTCTGTGACGTCACCTCCGGTGGGCGGGA CTTCCGGAATTAGGGTTGGCTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGA CAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCAGAGCC AACCCTAATTCCGGAAGTCCCGCCACCGGAAGTGACGTACAGGAAATGACGTCACA GGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACC GGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
TTPp (promotor específico de hígado) (SEQ ID NO:113)	ACGCGTGTCTGTCTGCACATTTCTGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAG GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATC AGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTGGAAGGAGG GGGTATAAAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTG
Intrón Sintético (SEQ ID NO:115)	GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCG TGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCACAG
FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto para la versión 6 de FVIII con codones optimizados que contiene XTEN144; la secuencia XTEN está marcada por un <u>doble subrayado</u> (SEQ ID NO:118))	ATGCAGATTGAGCTGTCCACTTGTTTTCTTCTGTGCCTCCTGCGCTTCTGTTTTCTCCG CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGA CCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCC TTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGAGTTCACTGACCACCTGT TCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAGC TGAAGTGACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTC CTGCATGCGGTGCGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACC AGACTAGCCAGCGGGAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTA CGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACT TACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTG CACTTCTCGTGTGCCGCGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAA GTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAAGTCATGGCATTCCGAAACTAAG AACTCGCTGATGCAGGACCGGGATGCCGCTCAGCCGCGCCTGGCCTAAAATGCATA CAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGTTGTACAGAAAGTC CGTGTACTGGCACGTATCGGCATGGGCACTACGCTGAAGTGCATCCATCTTCCTG GAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGA TTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGGCAGTTCTCTCTCTCTG CCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGC CCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACG ATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGATGACGACAACAGCCCCAGCTT CATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCG GCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGT ACAAGTCCCAGTATCTGAACAATGGTCCGACGCGGATTGGCAGAAAGTACAAGAAAGT GCGGTTTCATGGCGTACACTGACGAAACGTTTAAAGACCCGGGAGGCCATTCAACATGAG AGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTTCGGCGATACCCTGCTCATCATCT TCAAAAACCAGGCCTCCCGGCCCTTACAACATCTACCCTCACGGAATCACCGACGTGCG GCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATC CTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCCACCA AGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGA CCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAA CGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATG AAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCCCAAACCCCGCAGG AGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAAACGGT

	TACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACA TTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTT TAAGCACAAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACG GTGTTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACT TTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGG AGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAAC GCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCC CTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAG TGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAG ACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAAC CTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGG CACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGC AGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTG CTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGA AGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACC ACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGA AGAAGGAAGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTTCA AAAGAAAAC TAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATG TCATCCAGCCCTCACGTCCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCA AGAAAGTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGG AGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGAT AACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCTCCC TGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAA GCCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG GATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATG TCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCAGC GCATGGACGCCAGGTACCGTCCAGGAGTTTGCTCTGTTCTTACCATTTTTTGACGAA ACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATA TCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTA CATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTTCGGTGGTAC TTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGT TCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGT GTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTG ATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGT GCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAG CGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATC AACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCCTA TGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACAT CTCGCAATTTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGA AACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAGC ACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTA CTCAATCCGCTCAACTCTTCGGATGGAACATCATGGGGTGCGACCTGAACTCCTGCTCC ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCT ACTTCACTAACATGTTTCGCCACCTGGAGCCCCCTCCAAGGCCAGGCTGCACTTGCAAGG ACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGAT TTCCAAAAGACCATGAAAGTGACCGGAGTCAACACCCAGGGAGTGAAGTCCCTTCTGA CCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGAC CCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTTACA CCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTC AGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGA CCTGTACTGA
--	--

WPRe (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)+	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGAC GCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCG CTTTCCCCCTCCCTATTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTG GACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCG TCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCTCTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC TCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGG CCGCGCTCCCCGCTG
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTT GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTAATAAAATGAGGAAATTGCATCG CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
ITR 3' repetición terminal invertida (SEQ ID NO: 186)	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTTCGCCGCCGGTAGGCGGGACT TCCGGTACAAGATGGCGGACAATTACGTCATTTTCTGTGACGTCATTTTCTGTGACGT CACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGT TGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAA GCAAGCTGGCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTC ACAGGAAATGACGTACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCC CGCCTACCGGCGGCGACCGGCGGCATCTGATTG
Secuencia de Longitud Completa (SEQ ID NO: 189)	
CCAAATCAGATGCCGCCGGTTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATT TCCTGTGACGTCATTTCTGTGACGTCATTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCC AGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAA GCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTACAGGAAATGACGTCA CAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGA TTTGGTGTCTTCTTTTAAATTTTGCGGCAATTACGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAAT ACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTGCGGGCCGCT CGACGACTAGCGTTTAAATTAACGCGTGTCTGTCTGCACATTTCTGTAGAGCGAGTGTTCGATACTCTAATC TCCCTAGGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATCAGAATCA GCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCCTTCACC AGGAGAAGCCGTACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTC TTTACGGGTTATGGCCCTTGCCTGCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCACAGCT AGCGCCACCATGCAGATTGAGCTGTCCACTTGTCTTCTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACT CGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTG GATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCTTTAACACCTCCGTGGTGTACAAGAAAACCTC TTTGTGCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCG ACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCCCTG CATGCGGTTCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGCCAGCGGGAA AAGGAGGACGATAAAGTGTTCGCGGGCGGCTCGCATACTTACGTGTGGCAAGTCTGAAGGAAAACGGACCT ATGGCATCCGATCCTCTGTGCTGACTTACTCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGC GGGCTGATTGGTGCATTTCTCGTGTGCCGCAAGGTTGCTCGCTAAGGAAAAGACCCAGACCTCCATAAG TTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAACTAAGAACTCGCTGATGCAG GACCGGGATGCCGCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTG CCCGGGCTCATCGGTTGTACAGAAAGTCCGTGTACTGGCACGTATCGGCATGGGCACTACGCCTGAAGTG CACTCCATCTTCTGGAAGGGCACACCTTCTCGTGCACAACACCGCCAGGCCTCTTGGAAATCTCCCCG	

ATTACCTTTCTGACCGCCAGACTCTGCTCATGGACCTGGGGCAGTTCTTCTCTTCTGCCACATCTCCAGC
 CATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCCGAAGAACCTCAGTTGCGGATGAAG
 AACACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGATGAC
 GACAACAGCCCCAGCTTCATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATC
 GCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCAG
 TATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGACGAA
 ACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGC
 GATACCTTGCTCATCATCTTCAAAAACAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGAC
 GTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGC
 GAAATCTTCAAGTATAAGTGGACCGTCAACGTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACT
 CGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGC
 TACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTG
 TTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCCCAAACCCCGCAGGAGTGCAA
 CTGGAGGACCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTGCACTCGCTGCAA
 CTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCAGACTGACTTCCTGAGC
 GTGTTCTTTTCCGGTTACACCTTTAAGCACAAAGATGGTGTACGAAGATAACCTGACCCTGTTCCCTTTCTCC
 GGCGAAACGGTGTTCATGTGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGG
 AACCGCGGAATGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCC
 TACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGC
 GCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAAACG
 CCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGGCTCCGAGACT
 CCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAAGTACCGAACCTAGCGAGGGCTCTGCT
 CCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAGGG
 CCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGG
 CCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCCACCAGCACTGAA
 GAAGGTGCCTCGAGCCCGCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGAT
 CAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAATGAAGAAGGAAGATTTTCGATATCTACGACGAG
 GACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAAGTACGACTACTTTATCGCCGCGGTGGAAAGACTG
 TGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGGCCAGAGCGGATCGGTGCCTCAGTTC
 AAGAAAGTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAA
 CACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAA
 GCATCCAGACCTTACTCCTTCTATTCTCCTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCC
 CGCAAGAATTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACC
 AAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGC
 CTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTACCGTCCAG
 GAGTTTGCTCTGTTCTTACCATTTTTGACGAACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAAAC
 TGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAAC
 GGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCA
 ATGGGATCGAACGAAAACATTCCTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAG
 TACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGC
 ATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCTGGTGTACTCG
 AATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAA
 TACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAA
 CCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCATGATTATCCACGGAATTAAGACCCAGGGCGCCAGG
 CAGAAGTTCTCCTCCCTGTACATCTCGCAATTATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACT
 TACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAGCACAA
 ATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTT
 CGGATGGAATCATGGGTGCGACCTGAACTCCTGCTCCATGCCGTGGGGATGGAATCAAAGGCTATTAGC
 GACGCCCAGATACCGCGAGCTCCTACTTCACTAACATGTTGCGCACCTGGAGCCCCCTCAAAGGCCAGGCTG
 CACTTGCAGGGACGGTCAAATGCCTGGCGGCCGAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTC
 CAAAAGACCATGAAAGTGACCGGAGTCAACACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAG
 GAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCTGTCTTCCAAAACGGAAGGTCAAGGTG

TTCCAGGGGAACCAGGACTCGTTACACCCGTGGTGAACCTCCCTGGACCCCCACTGCTGACGCGGTACTTG
 AGGATTTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCAGGAC
 CTGTACTGAATCAGCCTGAGCTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTG
 GTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG
 CTTCCCGTATGGCTTTTCATTTTCTCCTCCTTGATATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGC
 CCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGGCTGACGCAACCCCCACTGGTTGGGGCATTGCCA
 CCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACATCATGCCGCCT
 GCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCAT
 CGTCCTTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTT
 CGGCCCTCAATCCAGCGGACCTTCCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCC
 TTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGT
 TGCCAGCCATCTGTTGTTTGGCCCTCCCCGTCCTTCCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCCTT
 TCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGTGGGGTGGGG
 CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT
 GAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGG
 ATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGG
 TAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCTGTGACGTCATTTCTGTGACGTCAC
 TTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGAT
 CAAGCGGCGCGCCGCTTGTCTTAGTGTCAGGCAACCCCAAGCAAGCTGGCCCAGAGCCAACCTAATTCGG
 GAAGTCCCGCCACCGGAAGTGACGTACAGGAAATGACGTACAGGAAATGACGTAATTGTCCGCCATCTT
 GTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG

Tabla 2E: Ejemplo de una construcción de AAV-FVIII (nucleótidos 1-6824; SEQ ID NO: 190)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 111)	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCT TTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCAT CACTAGGGGTTCCT
CAGp (promotor ubicuo) (SEQ ID NO:191)	TCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCC AATTTTGTATTTATTTATTTTATTTTAAATATTTTGTGACGCGATGGGGGCGGGGGGGGGG GGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGA GAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCTTTTATGGCGAG GCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCG
Intrón Sintético (SEQ ID NO:192)	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTAATGAC GGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTT GTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGACGGCTGCCTTCGGGGGGGACG GGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAAC CATGTTTCATGCCTTCTTCTTTTCTTCTACAG
FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto la versión 6 de FVIII con codones optimizados que contiene XTEN144; la secuencia XTEN está marcada por un <u>doble subrayado</u> (SEQ ID NO:118))	ATGCAGATTGAGCTGTCCACTTGTTCCTTCTTCTGTCCTCCTGCGCTTCTGTTTCTCCG CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGA CCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCC TTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGAGTTCACTGACCACCTGT TCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAGC TGAAGTGACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTC CTGCATGCGGTTCGGAGTGCTTACTGGAAGGCTCCGAAGGAGCTGAGTACGACGACC AGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTA CGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACT TACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTG CACTTCTCGTGTGCCGCGAAGGTTCTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAA GTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAAGTCATGGCATTCCGAAACTAAG AACTCGCTGATGCAGGACCGGGATGCCGCTCAGCCGCGCCTGGCCTAAAATGCATA CAGTCAACGGATACGTGAATCGGTCACTGCCCCGGCTCATCGGTTGTACAGAAAAGTC CGTGTACTGGCACGTATCGGCATGGGCACTACGCTGAAGTGCACTCCATCTTCTG

	GAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGA TTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTTCCTTCTTCTTG CCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGC CCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACG ATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGATGACGACAACAGCCCCAGCTT CATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCCAAAACCTGGGTGCACTACATCGCG GCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGT ACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGT GCGGTTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAG AGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCT TCAAAAACCAGGCCTCCCGGCCCTTACAACATCTACCCTCACGGAATCACCGACGTGCG GCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATC CTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCCACCA AGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTTCGTGAACATGGAACGGGA CCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAA CGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATG AAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTTCCTCCCAAACCCCGCAGG AGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAAACGGT TACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACA TTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTT TAAGCACAAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACG GTGTTTCATGTTCGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACT TTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGG AGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAAC GCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCC CTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAAACGCCAGGCACAAG TGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAG ACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAAC CTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGG CACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGC AGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTG CTGGATCTCCTACGTCCACTGAGGAAGGGTACCAGCGGGCTCGCCCCACCAGCACTGA AGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACC ACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGA AGAAGGAAGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTTCA AAAGAAAACCTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATG TCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCA AGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGG AGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGAT AACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTCC TGATCTCATAACGAGGAGGACCAGCGCCAAGGCGCGGAGCCCCGCAAGAACTTCGTCAA GCCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG GATGAGTTTACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATG TCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCAGC GCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTACCATTTTTGACGAA ACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATA TCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTA CATCATGGATACTCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATTCGGTGGTAC TTGCTGTCAATGGGATCGAACGAAAACATTCCTCCATTCACTTCTCCGGTCACGTGT TCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGT GTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGCATCTGGAGAGTGGAGTGCCTG ATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTCCTGGTGTACTCGAATAAGT
--	---

	GCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAG CGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATC AACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTA TGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACAT CTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGA AACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCGGCATTAAAGC ACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTA CTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGGCAGCTGAACTCCTGCTCC ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCT ACTTCACTAACATGTTTCGCCACCTGGAGCCCCCTCCAAGGCCAGGCTGCACTTGCAAGG ACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGAT TTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGA CCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGAC CCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTTACA CCCGTGGTGAATCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTC AGTCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCTGGGCTGCGAGGCCAGGA CCTGTACTGA
WPRe (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGAC GCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTTCG CTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCTGCCTTGCCCCGCTGCTG GACAGGGGCTCGGCTGTTGGGCACTGACAATTCCTGTTGTTGTCGGGGAAATCATCG TCCTTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCTTCT GCTACGTCCCTTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC TCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTTG GCCGCTCCCCGCTG
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTT GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCG CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGG GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
ITR 3' repetición terminal invertida (SEQ ID NO: 193)	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGA GGCCGGGCGACCAAAGGTCGCCCAGCGCCCGGCTTTGCCGGGCGGCCTCAGTGAGC GAGCGAGCGCGCAG
Secuencia de Longitud Completa (SEQ ID NO: 190)	
CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTGCCTCGCCGGCC TCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTTCTGCGGCAATTCAGTC GATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATT GAGATCTGGATCCGGTACCGAATTCGCGGCCGCTCGACGACTAGCGTTTAATTAAATCGAGGTGAGCCCCA CGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTGTATTTATTTATTTTAAATTAT TTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGG GCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCTTTTATGGCGA GGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGCCTTCGCC CCGTGCCCGCTCCGCCGCCGCTCGCGCCCGCCCGCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGA GCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTAATGACGGCTTGTCTTCTTCTGT GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGCTGTCCGCGG	

GGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGA
 GCCTCTGCTAACCATGTTTCATGCCTTCTTCTTTTTCTACAGGCTAGCGCCACCATGCAGATTGAGCTGTCC
 ACTTGTTTCTTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGAGCCGTGGAG
 CTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCCGGTGCCA
 AAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGAGTTCACTGACCACCTGTTT
 AACATCGCCAAGCCGCGCCACCTTGATGGGCTCCTGGGACCGACCATTCAGCTGAAGTGTACGACACC
 GTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGGTCGGAGTGTCTCTACTGGAAG
 GCCTCCGAAGGAGCTGAGTACGACGACCACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCGCGGC
 GGCTCGCATACTTACGTGTGGCAAGTCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACT
 TACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACCTTCTCGTGTGC
 CGCGAAGGTTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTTCGAT
 GAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCTCAGCCCGCGCC
 TGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTACAGAAAAG
 TCCGTGTACTGGCACGTATCGGCATGGGCACTACGCCTGAAGTGCATCCATCTTCTGGAAGGGCACACC
 TTCTCGTGCACAACACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCAGACTCTG
 CTCATGGACCTGGGGCAGTTCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTAC
 GTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGAC
 GACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTTCGATGACGACAACAGCCCCAGCTTCATCCAGATT
 CGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTAC
 GCCCCGTGGTGTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATT
 GGCAGAAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAAACGTTTAAAGACCCGGGAGGCCATTCAA
 CATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTTCGGCGATACCCTGCTCATCATCTTCAAAAAC
 CAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTG
 CCGAAGGGCGTCAAGCACCTGAAAAGACTTCCCTATCTGCGGGCGAAAATCTTCAAGTATAAGTGGACCGTC
 ACCGTGGAGGACGGGCCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATG
 GAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTTCGGTGGATCAACGCGGC
 AACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTC
 ACTGAAAACATCCAGAGGTTCTTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCG
 AATATCATGCACTCGATTAACGGTTACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTTCGT
 TACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCCTGAGCGTGTCTTTTCCGGTTACACCTTTAAG
 CACAAGATGGTGTACGAAGATAACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTGATGGAG
 AACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTTCGGAACCGCGGAATGACTGCCCTGCTGAAG
 GTGTCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTG
 TCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCT
 GAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAACGCCAGGCACAAGTGAGTCTGCAACTCCC
 GAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAGAGCGCTACACCA
 GAAAGCGGACCCGGAACAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACA
 TCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGC
 AGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACG
 TCCACTGAGGAAGGGTACCAGCGGGCTCGCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTG
 AAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACC
 ATCTCGGTGGAAATGAAGAAGGAAGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTC
 CAAAAGAAAACCTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCT
 CACGTCCTTCGGAACCGGGGCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTCACC
 GACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATC
 CGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCC
 TCCCTGATCTCATAAGGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACGAG
 ACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGG
 GCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTAC
 ACCAACACCCTGAACCCAGCGCATGGACGCCAGGTACCGTCCAGGAGTTTGCTCTGTTCTTCAACATTTTT
 GACGAAAATAAGTCTGTTACTTACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATATCCAGATG
 GAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGG

CTGGTCATGGCCAGGATCAGAGGATTCTGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCC
ATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTAC
CCCGGGGTGTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGCATCTGGAGAGTGGAGTGCCTGATCGGA
GAGCACCTCCACGCGGGGATGTCCACCTCTTCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATG
GCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCC
CGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTC
CTGGCCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAAGTTCTCCTCCCTGTACATCTCG
CAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTG
ATGGTCTTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAAGCACAACATCTTCAACCCACCGATCATAGCCAGA
TATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGGCAGCTG
AACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCTAC
TTACTAACATGTTTCGCCACCTGGAGCCCCCTCAAGGCCAGGCTGCACCTGCAGGGACGGTCAAATGCCTGG
CGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTC
ACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGAC
GGGCACCACTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTTACA
CCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTATCCTCAGTCTCTGGGTCCAT
CAGATTGCATTGCGAATGGAAGTCTGGGCTGCGAGGCCAGGACCTGTACTGAATCAGCCTGAGCTCGCTG
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTA
CGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCT
CCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTTCAGGCAACGTGGCGTGGTGT
GCACTGTGTTTGCTGACGCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTT
TCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC
GGCTGTTGGGCACTGACAATTCCGTGGTGTGTGCGGGGAAATCATCGTCTTTTCTTGGCTGCTCGCCTGTG
TTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTT
CCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCTTCGCCCTCAGACGAGTCGGATCTCCC
TTTGGGCCGCTCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCC
CCCGTGCCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCTTTTCTTAATAAAATGAGGAAATTCATCG
CATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGC
TTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGAG
GAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGGCCGGGCGACCAAAG
GTCGCCCCGACGCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG

Tabla 2F: Ejemplo de construcción de GPV-FVIII que porta las ITR de GPV de longitud completa (nucleótidos 1-7154; SEQ ID NO: 194)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 187)	CTCATTGGAGGGTTTCGTTTCGTTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAA GTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCT TCCTGTACAGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTT TCCGGCTGTTAGGTTGACCACGCGCATGCCGCGGGTCAGCCCAATAGTTAAGCCGGA AACACGTACCCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGACAGGA AGCACGTGACCGGAACTACGTACCCGGATGTGCGTCACCGGAAGCATGTGACCGGAAC TTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAG ACTCAAGGACAAGAGGATATTTTGC GCGCCAGGAAGTG
TTPp (promotor específico de hígado) (SEQ ID NO:113)	ACGCGTGTCTGTCTGCACATTTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAG GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATC AGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCTGGGTTGGAAGGAGG GGGTATAAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTG

Intrón Sintético (SEQ ID NO:115)	GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCG TGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCACAG
FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto para la versión 6 de FVIII con codones optimizados que contiene XTEN144; la secuencia XTEN está marcada por un <u>doble subrayado</u> (SEQ ID NO:118))	ATGCAGATTGAGCTGTCCACTTGTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCG CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGA CCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCC TTTAACACCTCCGTGGTGTACAAGAAAACCTCTTTGTGCGAGTTCACTGACCACCTGT TCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAGC TGAAGTGACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCC CTGCATGCGGTCCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACC AGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTA CGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACT TACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTG CACTTCTCGTGTGCCGCGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCTCCATAA GTTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCGGAAACTAAG AACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATA CAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGTTGTACAGAAAAGTC CGTGTACTGGCACGTCAATCGGCATGGGCACTACGCCTGAAGTGCATCCATCTTCCTG GAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGA TTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTTCTTCTTCTTG CCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGC CCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACG ATTTGACTGACTCCGAGATGGACGTGCTGCGGTTTCGATGACGACAACAGCCCCAGCTT CATCCAGATTTCGACGCGTGGCCAAGAAGCACCCCCAAAACCTGGGTGCACTACATCGCG GCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCTG ACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGT GCGGTTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAG AGCGGCATTCTGGGACCCTGCTGTACGGAGAGGTGCGCGATACCCTGCTCATCATCT TCAAAAACCAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTGCG GCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATC CTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCCACCA AGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCTGTAACATGGAACGGGA CCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAA CGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATG AAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTTCTCCAAACCCCGCAGG AGTGCAACTGGAGGACCCTGAGTTTTCAGGCCTCGAATATCATGCACTCGATTAACGGT TACGTGTTTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGTTACTGGTACA TTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTCTTTTCCGGTTACACCTT TAAGCACAAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACG GTGTTTCATGTGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACT TTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGG AGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAAAC GCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCC <u>CTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAACGCCAGGCACAAG</u> <u>TGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAG</u> <u>ACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAAGCGGACCCGGAACCAGTACCGAAC</u> <u>CTAGCGAGGGCTCTGCTCCGGGCGAGCCAGCCGGCTCTCCTACATCCACGGAGGAGGG</u> <u>CACTTCCGAATCCGCCACCCCGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGC</u> <u>AGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCCTG</u> <u>CTGGATCTCCTACGTCCACTGAGGAAGGGTCAACAGCGGGCTCGCCACACGACTGA</u> <u>AGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACC</u> <u>ACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGA</u>

	AGAAGGAAGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCA AAAGAAAACCTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATG TCATCCAGCCCTCACGTCCTTCGGAACCGGGCCAGAGCGGATCGGTGCCTCAGTTCA AGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGG AGAAGTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGAT AACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCTCC TGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAA GCCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAG GATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATG TCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCTGAACCCAGC GCATGGACGCCAGGTACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAA ACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATA TCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTA CATCATGGATACTCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATTCGGTGGTAC TTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGT TCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGT GTTTCGAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTG ATCGGAGAGCACCTCCACGCGGGGATGTCCACCTCTTCTGGTGTACTCGAATAAGT GCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAG CGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATC AACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCCTA TGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACAT CTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGA AACTCCACCGGCACCCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAGC ACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTA CTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGCGACCTGAACTCCTGCTCC ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCT ACTTCACTAACATGTTTCGCCACCTGGAGCCCCCTCCAAGGCCAGGCTGCACTTGCAAGG ACGGTCAAATGCCTGGCGGCCGAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGAT TTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGA CCTCGATGTATGTGAAGGAGTTCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGAC CCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTTACA CCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTTCATCCTC AGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCCAGGA CCTGTACTGA
WPRE (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGAC GCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCG CTTTCCCCCTCCCTATTGCCACGGCGGAACCTCATCGCCGCTGCCTTGCCCGCTGCTG GACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCG TCCTTTCTTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC TCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGG GCCGCTCCCCGCTG
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTT GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTAATAAAATGAGGAAATTGCATCG CATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA

ITR 3' repetición terminal invertida (SEQ ID NO: 188)	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTTCGTTCCG TTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCG GTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCGTGTCACGTGTTTCCGGT CACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTTAACATTATGGGCT GACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAAACACGTACCCGGAAGTCAC ATGACCGGAAGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACACG TCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACCTTGGCTCACTTCCCCCTCCC CTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAG
Secuencia de Longitud Completa (SEQ ID NO: 194)	
CTCATTGGAGGGTTTCGTTTCGTTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATG CTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCGTGTCACGTGTTTCCGGTCACGTGAC TTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTACAGCC CAATAGTTAAGCCGGAAACACGTACCCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGACAG GAAGCACGTGACCGGAACACGTACCCGGATGTGCGTCACCGGAAGCATGTGACCGGAACCTTGGCTCACTTC CCCCCTCCCCGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGGC CGCCAGGAAGTGGCGGCAATTACGTGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCT TAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCTCGACGACTAGC GTTTAATTAAACGCGTGTCTGTCTGCACATTTTCGTAGAGCGAGTGTTCGGATACTCTAATCTCCCTAGGCAA GGTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGGA GTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCCTTACCAGGAGAAGCCG TCACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTA TGGCCCTTGGCTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTCTCCACAGCTAGCGCCACCAT GCAGATTGAGCTGTCCACTTGTTTTCTTCCGTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTA CCTTGGAGCCGTGGAGCTTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATT CCCCCCCCGCGTGCCAAAGTCCTTCCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCCTCTTTGTGCGAGTT CACTGACCACCTGTTCAACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTTCAAGC TGAAGTGTACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGGTTCGG AGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGA TAAAGTGTTCGGGGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGA TCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGG TGCATTCTCGTGTGCCGCGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTTCATCCTTTT GTTGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGC CGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCCGGGCTCAT CGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCACCTCCATCTT CCTGGAAGGGCACACCTTCTCGTGCACAACCACCGCCAGGCCTCTCTGGAATCTCCCCGATTACCTTTCT GACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTTCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGA CGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCCGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGA GGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTGCTGCGGTTTCGATGACGACAACAGCCC CAGCTTCATCCAGATTTCGAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGA AGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTGCTACAAGTCCAGTATCTGAACAA TGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGACGAAACGTTTAAGAC CCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTGGCGGATACCTGCT CATCATCTTCAAAAACAGGCCTCCCGGCCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACT CTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCGGGGCGAAATCTTCAA GTATAAGTGGACCGTCACCGTGGAGGACGGGCCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTC CAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTC GGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGTATGAAAA CAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTTCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCC TGAGTTTTCAGGCCTCGAATATCATGCACTCGATTAAACGGTTACGTGTTTCGACTCGCTGCAACTGAGCGTGTG CCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCAGACTGACTTCTGAGCGTGTCTTTTTC CGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCTGTTCCCTTTCTCCGGCGAAACGGT GTTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAAT	

GACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATAT
 CTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGGCCAACATC
 AGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTTCGGAACGCCAGGCACAAG
 TGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTTC
 CGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCC
 AGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGA
 ACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGCCC
 TGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCACCAGCACTGAAGAAGGTGCCTC
 GAGCCCGCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACTTCCAATCGGATCAGGAGGAAAT
 CGACTACGACGACACCATCTCGGTGGAATGAAGAAGGAAGATTTGATATCTACGACGAGGACGAAAATCA
 GTCCCTCGCTCATTTCCAAAAGAAAAGTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGG
 AATGTCATCCAGCCCTCACGTCTTTCGGAACCGGGCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGT
 GTTCCAGGAGTTACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCCT
 GCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACC
 TTAATCCTTCTATTCTCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTT
 CGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTT
 TGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCC
 GCTGCTCGTGTGTACACCAACACCTGAACCCAGCGCATGGACGCCAGGTACCCGTCCAGGAGTTTGCTCT
 GTTCTTACCATTTTTTGACGAAACTAAGTCCCTGGTACTTACCCGAGAATATGGAGCGAAACTGTAGAGCGCC
 CTGCAATATCCAGATGGAAGATCCGACTTTCAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCAT
 GGATACTCTGCCGGGGCTGGTCATGGCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAA
 CGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGC
 GCTGTACAATCTGTACCCCGGGGTGTTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGT
 GGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCCTGGTGTACTCGAATAAGTGCCA
 GACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCAATG
 GGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTG
 GATTAAGGTGGACCTCCTGGCCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAAGTTCTC
 CTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAA
 CTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCTCCGGCATTAAGCACAAACATCTTCAACCC
 ACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCCGGATGGAAC
 CATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGAT
 CACCGCGAGCTCCTACTTCACTAACATGTTGCCACCTGGAGCCCCCTCAAGGCCAGGCTGCACTTGCAGGG
 ACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCAT
 GAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCTTGAT
 TAGCAGCAGCCAGGACGGGCACCAGTGGACCTGTTCTTCCAAAACGGAAGGTCAAGGTGTTCCAGGGGAA
 CCAGGACTCGTTACACCCGTGGTGAACCTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTATCC
 TCAGTCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCTTGGGCTGCGAGGCCAGGACCTGTACTGAAT
 CAGCCTGAGCTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
 TATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
 GCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGG
 CAACGTGGCGTGGTGTGCACTGTGTTGCTGACGCAACCCCCCACTGGTTGGGGCATTGCCACCACCTGTCAG
 CTCCTTTCCGGGACTTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACATCATCGCCGCTGCCTTGCCCGC
 TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGTGTGTCGGGGAAATCATCGTCTCTTCCCT
 TGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCGGCCCTCAAT
 CCAGCGGACCTTCTTCCCGCGGCCTGTGCGCGCTCTGCGGCCTCTTCCGCGTCTTTCGCTTCGCCCTCAG
 ACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATC
 TGTTGTTTGCCCCCTCCCCCGTGCCTTCTTGACCTGGAAGGTGCCACTCCCACTGTCCTTTCTTAATAAAAA
 TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
 GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAG
 AACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTA
 ATGGCGCGGGCGCCCACTTCTGGCGCGCAAAATATCCTCTTGTCTTGTGCTGAGTCTCATTTGGAGGGTTCGTTTCG
 TTGCAACCAGCCAATCAGGGGAGGGGGAAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCG

```

GTGACGTAGTTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCG
GTGACGTGTTTCCGGCTTAACCTATTGGGCTGACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAAACA
CGTCACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAACACGTGACAGGAAGCACGTGACCGGAACCTA
CGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACCTGCGTCACTTCCCCCTCCCCTGATTGGCTGG
TTCGAACGAACGAACCCCTCCAATGAG

```

[0283] En una realización, el casete genético comprende una construcción de fenilalanina hidroxilasa (PAH), que incluye una secuencia de polinucleótidos como se enumera en las Tablas 10A y 10B. En una realización, el casete genético comprende una construcción de PAH, que incluye una secuencia de polinucleótidos expuesta en la Tabla 10A. En una realización, el casete genético comprende una construcción de PAH, que incluye una secuencia de polinucleótidos expuesta en la Tabla 10B.

[0284] En ciertas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 197 o 198. En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 197. En algunas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 198. En algunas realizaciones, la molécula de ácido nucleico aislada conserva la capacidad de expresar una fenilalanina hidroxilasa funcional.

A. Índice de adaptación de codones

[0285] En una realización, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde el índice de adaptación de codones humanos de la secuencia de nucleótidos con codones optimizados aumenta con respecto a SEQ ID NO: 16. Por ejemplo, la secuencia de nucleótidos con codones optimizados puede tener un índice de adaptación de codones humanos que es al menos aproximadamente 0,75 (75%), al menos aproximadamente 0,76 (76%), al menos aproximadamente 0,77 (77%), al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%), al menos aproximadamente 0,80 (80%), al menos aproximadamente 0,81 (81%), al menos aproximadamente 0,82 (82%), al menos aproximadamente 0,83 (83%), al menos aproximadamente 0,84 (84%), al menos aproximadamente 0,85 (85%), al menos aproximadamente 0,86 (86%), al menos aproximadamente 0,87 (87%), al menos aproximadamente 0,88 (88%), al menos aproximadamente 0,89 (89%), al menos aproximadamente 0,90 (90%), al menos aproximadamente 0,91 (91%), al menos aproximadamente 0,92 (92%), al menos aproximadamente 0,93 (93%), al menos aproximadamente 0,94 (94%), al menos aproximadamente 0,95 (95%), al menos aproximadamente 0,96 (96%), al menos aproximadamente 0,97 (97%), al menos aproximadamente 0,98 (98%), o al menos aproximadamente 0,99 (99%). En algunas realizaciones, la secuencia de nucleótidos con codones optimizados tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,88 (88%). En otras realizaciones, la secuencia de nucleótidos con codones optimizados tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,91 (91%). En otras realizaciones, la secuencia de nucleótidos con codones optimizados tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,91 (97%).

[0286] En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii) nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la

porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde el índice de adaptación de codones humanos de la secuencia de nucleótidos aumenta con respecto a SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,75 (75%), al menos aproximadamente 0,76 (76%), al menos

aproximadamente 0,77 (77%), al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%), al menos aproximadamente 0,80 (80%), al menos aproximadamente 0,81 (81%), al menos aproximadamente 0,82 (82%), al menos aproximadamente 0,83 (83%), al menos aproximadamente 0,84 (84%), al menos aproximadamente 0,85 (85%), al menos aproximadamente 0,86 (86%), al menos aproximadamente 0,87 (87 %), al menos aproximadamente 0,88 (88%), al menos aproximadamente 0,89 (89%), al menos aproximadamente 0,90 (90%) o al menos aproximadamente 0,91 (91%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,88 (88%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,91 (91%).

[0287] En otra realización, la secuencia de nucleótidos con codones optimizados que codifica el polipéptido de FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 o (ii) 1792-2277 y 2320-4374 de SEQ ID NO: 6; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde el índice de adaptación de codones humanos de la secuencia de nucleótidos aumenta con respecto a SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,75 (75%), al menos aproximadamente 0,76 (76%), al menos aproximadamente 0,77 (77%), al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%), al menos aproximadamente 0,80 (80%), al menos aproximadamente 0,81 (81%), al menos aproximadamente 0,82 (82%), al menos aproximadamente 0,83 (83%), al menos aproximadamente 0,84 (84%), al menos aproximadamente 0,85 (85%), al menos aproximadamente 0,86 (86%), al menos aproximadamente 0,87 (87 %), o al menos aproximadamente 0,88 (88%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación de

codones humanos que es al menos aproximadamente 0,83 (83%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,88 (88%).

- 5 **[0288]** En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al
- 10 menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 (es decir, nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde el índice de adaptación de codones humanos de la secuencia
- 15 de nucleótidos aumenta con respecto a SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,75 (75%), al menos aproximadamente 0,76 (76%), al menos aproximadamente 0,77 (77%), al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%), al menos aproximadamente 0,80 (80%), al menos aproximadamente 0,81 (81%), al menos aproximadamente 0,82 (82%), al menos
- 20 aproximadamente 0,83 (83%), al menos aproximadamente 0,84 (84%), al menos aproximadamente 0,85 (85%), al menos aproximadamente 0,86 (86%), al menos aproximadamente 0,87 (87%), o al menos aproximadamente 0,88 (88%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,75 (75%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos
- 25 aproximadamente 0,83 (83%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,88 (88%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,91 (91%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,97 (97%).
- 30 **[0289]** En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII de la presente divulgación tiene una frecuencia aumentada de codones óptimos (FOP) con respecto a SEQ ID NO: 16. En ciertas realizaciones, la FOP de la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII es al menos aproximadamente 40, al

menos aproximadamente 45, al menos aproximadamente 50, al menos aproximadamente 55, al menos aproximadamente 60, al menos aproximadamente 64, al menos aproximadamente 65, al menos aproximadamente 70, al menos aproximadamente 75, al menos aproximadamente 79, al menos aproximadamente 80, al menos aproximadamente 85, o al menos aproximadamente 90.

- 5 **[0290]** En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII de la presente divulgación tiene un aumento de uso de codón sinónimo relativo (RCSU) con respecto a SEQ ID NO: 16. En algunas realizaciones, el RCSU de la molécula de ácido nucleico aislada es mayor que 1,5. En otras realizaciones, el RCSU de la molécula de ácido nucleico aislada es mayor que 2,0. En ciertas realizaciones, el RCSU de la molécula de ácido nucleico aislada es al
- 10 menos aproximadamente 1,5, al menos aproximadamente 1,6, al menos aproximadamente 1,7, al menos aproximadamente 1,8, al menos aproximadamente 1,9, al menos aproximadamente 2,0, al menos aproximadamente 2,1, al menos aproximadamente 2,2, al menos aproximadamente 2,3, al menos aproximadamente 2,4, al menos aproximadamente 2,5, al menos aproximadamente 2,6 o al menos aproximadamente 2,7.
- 15 **[0291]** En otras realizaciones más, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII de la presente divulgación tiene una disminución del número efectivo de codones con respecto a SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos optimizada de codones que codifica un polipéptido de FVIII tiene un número efectivo de codones de menos de
- 20 aproximadamente 50, menos de aproximadamente 45, menos de aproximadamente 40, menos de aproximadamente 35, menos de aproximadamente 30, o menos de aproximadamente 25. En una realización particular, la molécula de ácido nucleico aislada tiene un número efectivo de codones de
- aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25, o aproximadamente 20.

B. Optimización del Contenido de G/C

- 25 **[0292]** En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados contiene un mayor porcentaje de nucleótidos de G/C en comparación con el porcentaje de nucleótidos de G/C en SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al
- 30 menos aproximadamente 45%, al menos aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos aproximadamente 51%, al menos aproximadamente 52%, al menos aproximadamente 53%, al

menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56%, al menos aproximadamente 57%, al menos aproximadamente 58%, al menos aproximadamente 59% o al menos aproximadamente 60%.

[0293] En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica

5 un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos
10 aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii) nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-
15 1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende tiene un contenido de G/C que es al menos aproximadamente 45%, al menos
20 aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos aproximadamente 51%, al menos aproximadamente 52%, al menos aproximadamente 53%, al menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56%, al menos aproximadamente 57% o al menos aproximadamente 58%. En una realización particular, la secuencia de nucleótidos que codifica un
25 polipéptido con actividad de FVIII tiene un contenido de G/C que es al menos aproximadamente 58%.

[0294] En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un

polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos
30 aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos

aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B), o (iv) 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene un porcentaje más alto de nucleótidos de G/C en comparación con el porcentaje de nucleótidos de G/C en SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 45%, al menos aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos aproximadamente 51%, al menos aproximadamente 52%, al menos aproximadamente 53%, al menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56% o al menos aproximadamente 57%. En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 52%. En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 55%. En otra realización, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 57%.

[0295] En otras realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-4374 o (ii) nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 (es decir, nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII

tiene un contenido de G/C que es al menos aproximadamente 45%. En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 52%. En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 55%. En otra realización, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 57%. En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 58%. En otra realización más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII tiene un contenido de G/C que es al menos aproximadamente 60%.

[0296] "Contenido de G/C" (o contenido de guanina-citosina), o "porcentaje de nucleótidos de G/C", se refiere al porcentaje de bases nitrogenadas en una molécula de ADN que son guanina o citosina. El contenido de G/C se puede calcular utilizando la siguiente fórmula:

$$\frac{G + C}{A + T + G + C} \times 100$$

[0297] Los genes humanos son muy heterogéneos en su contenido de G/C, con algunos genes que tienen un contenido de G/C tan bajo como 20%, y otros genes tienen un contenido de G/C tan alto como 95%. En general, los genes ricos en G/C se expresan en mayor medida. De hecho, se ha demostrado que aumentar el contenido de G/C de un gen puede conducir a una mayor expresión del gen, debido principalmente a un aumento en la transcripción y niveles más altos de ARNm en estado estacionario. Véase Kudla et al., PLoS Biol., 4 (6): e180 (2006).

C. *Secuencias Similares a Regiones de Anclaje a la Matriz*

[0298] En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados contiene menos secuencias de MARS/ARS con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3 o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene una secuencia MARS/ARS.

[0299] En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii) nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene menos secuencias MARS/ARS con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene una secuencia MARS/ARS.

[0300] En otra realización, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); o (iv) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad

polipeptídica de FVIII; y en donde la secuencia de nucleótidos contiene menos secuencias MARS/ARS con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada de codones que codifica un polipéptido de FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene una secuencia MARS/ARS.

[0301] En otras realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones

optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados contiene menos secuencias MARS/ARS con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene una secuencia MARS/ARS.

[0302] Se han identificado elementos ricos en AT en la secuencia de nucleótidos de FVIII humano que comparten similitud de secuencia con secuencias de replicación autónoma (ARS) de *Saccharomyces cerevisiae* y regiones de unión a la matriz nuclear (MAR). (Fallux *et al.*, *Mol. Cell. Biol.* 16:4264-4272 (1996). Se ha demostrado que uno de estos elementos se une a factores nucleares *in vitro* y reprime la expresión de un gen indicador de cloranfenicol acetiltransferasa (CAT). *Id.* Se ha planteado la hipótesis de que estas secuencias pueden contribuir a la represión transcripcional del gen del FVIII humano. Por tanto, en una realización, todas las secuencias de MAR/ARS están abolidas en la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII de la presente descripción. Existen cuatro secuencias MAR/ARS ATATTT (SEQ ID NO: 21) y tres secuencias MAR/ARS

AAATAT (SEQ ID NO: 22) en la secuencia de FVIII parental (SEQ ID NO: 16). Todos estos sitios se mutaron para destruir las secuencias MAR/ARS en las secuencias optimizadas de FVIII (SEQ ID NO: 1-6). La ubicación de cada uno de estos elementos y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3, a continuación.

5 **Tabla 3:** Resumen de cambios en elementos represivos

Localización de Elemento	Secuencia de FVIII BDD de Partida (SEQ ID NO: 16)	Secuencia de FVIII BDD Optimizada							
		SEQ ID NO: 1	SEQ ID NO: 2	SEQ ID NO: 3	NO: 4	NO: 5	NO: 6	NO: 70	NO: 71
Secuencias Desestabilizadoras									
639	ATTTA	GTTTA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA
1338	ATTTA	GTTTA	GTTCA	CTTCA	GTTCA	GTTCA	GTTCA	CTTCA	GTTCA
1449	ATTTA	CTTTA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA	CTTCA
1590	TAAAT	TAAAT	CAAGT	CAAGT	TAAGT	CAAGT	CAAGT	CAAGT	TAAGT
1623	TAAAT	CAAAA	GAAGA	CTAAG	CAAGA	CAAGA	CAAGA	TAAGT	CAAGA
2410	ATTTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA	ATCTA
2586	ATTTA	GTTTA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA	GTTCA
2630	TAAAT	TGAAT	TGAAC	TGAAC	TGAAC	TCAAT	TGAAC	TCAAT	TGAAC
3884	ATTTA	ATCTG	ACCTG	ACCTG	ACCTG	ATCTG	ACCTG	ATCTG	ACCTG
3887	TAAAT	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC	TGAAC
Sitios Potenciales de Unión al Promotor									
641	TTATA	TTATC	TCATC	TCATT	TCATC	TCATC	TCATC	TCATT	TCATC
1275	TATAA	CTATA	TTACA	CTACA	GTACA	CTACA	CTACA	CTACA	GTACA
1276	TTATA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
1445	TTATA	TCATC	TCATC	TTATC	TCATC	TCATC	TCATC	TTATC	TCATC
1474	TATAA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
1588	TATAA	TATAA	TACAA	TACAA	TATAA	TACAA	TACAA	TACAA	TATAA
2614	TTATA	CTGTA	CTGTA	CTGTA	CTGTA	TTGTA	CTGTA	TTGTA	CTGTA
2661	TATAA	CATCA	CATCA	CATCA	CATCA	CATCA	CATCC	CATCA	CATCC

3286	TATAA	TATAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA	TACAA
3840	TTATA	TTATA	TTACT	CTACA	CTACA	CTACA	CTACT	CTACA	CTACT
Secuencias de Tipo Anclaje a la Matriz (MARS/ARS)									
1287	ATATTT	GTATCT	GTACCT	GTACCT	GTATCT	GTACCT	GTACCT	GTACCT	GTATCT
1447	ATATTT	ATCTTT	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC	ATCTTC
1577	AAATAT	AAATCT	AGATCT	AAATCT	AAATCT	AGATCT	AGATCT	AAATCT	AAATCT
1585	AAATAT	AAGTAT	AAGTAC	AAGTAC	AAGTAT	AAGTAC	AAGTAC	AAGTAC	AAGTAT
2231	ATATTT	ACATCA	ATATCA	ACATCA	ACATCA	ACATCT	ATATCT	ACATCT	ATATCT
3054	AAATAT	AAACAT	GAATAT	GAACAT	GAACAT	GAACAT	GAATAT	GAACAT	GAATAT
3788	ATATTT	ATATCT	ATATCT	ACATCT	ACATCT	ACATCT	ACATCT	ACATCT	ACATCT
Elementos de Secuencia Ricos en AU (ARE)									
2468	ATTTTA TT	ACTTCAT C	ACTTCAT C	ACTTCAT T	ACTTCAT T	ACTTTAT T	ACTTTAT C	ACTTTAT T	ACTTTAT C
3790	ATTTTT AA	ATCTTTA A	ATCTTCA A	ATCTTCA A	ATCTTCA A	ATCTTCA A	ATCTTCA A	ATCTTCA A	ATCTTCA A
Secuencias Poli A/Poli T									
3273	AAAAAA A	GAAAAAA	GAAGAAG	GAAGAAG	GAAGAAG	GAAGAAG	CAAGAAG	GAAGAAG	CAAGAAG
4195	TTTTTT	TTCTTT	TTCTTC	TTCTTC	TTCTTC	TTCTTC	TTCTTC	TTCTTC	TTCTTC
Sitios de Corte y Empalme									
2203	GGTGAT	GGGGAC	GGCGAC	GGGGAC	GGGGAC	GGAGAC	GGAGAC	GGAGAC	GGAGAC

D. Secuencias Desestabilizadoras

[0303] En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados contiene menos elementos desestabilizadores con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como máximo 1 elementos desestabilizadores. En otras realizaciones más, la secuencia de

nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un elemento desestabilizador.

[0304] En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii) nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene menos elementos desestabilizadores con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como máximo 1 elementos desestabilizadores. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un elemento desestabilizador.

[0305] En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es

decir, nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); o (iv) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene menos elementos desestabilizadores con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como máximo 1 elementos desestabilizadores. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un elemento desestabilizador.

[0306] En otras realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados contiene menos elementos desestabilizadores con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elementos desestabilizadores. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un elemento desestabilizador.

[0307] Existen diez elementos desestabilizadores en la secuencia de FVIII parental (SEQ ID NO: 16): seis secuencias ATTTA (SEQ ID NO: 23) y cuatro secuencias TAAAT (SEQ ID NO: 24). En una

realización, las secuencias de estos sitios se mutaron para destruir los elementos desestabilizadores en SEQ ID NO: 1-6, 70 y 71 de FVIII optimizado. La ubicación de cada uno de estos elementos y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

E. Sitios Potenciales de Unión al Promotor

- 5 **[0308]** En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos contiene menos sitios de unión al promotor potenciales con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 sitios potenciales de
- 10 unión al promotor. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como máximo 1 sitios potenciales de unión al promotor. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un sitio potencial de unión al promotor.
- 15 **[0309]** En una realización particular, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos
- 20 aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii)
- 25 nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene menos sitios de unión al promotor potenciales con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII
- 30 contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 sitios potenciales de unión al promotor. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como máximo 1 sitios potenciales de unión al promotor. En otras realizaciones más, la

secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un sitio potencial de unión al promotor.

[0310] En otra realización, la secuencia de nucleótidos con codones optimizados que codifica un

polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-

terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos

aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos

aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos

aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos

aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos

aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos

aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-4374 de SEQ ID NO: 5; (ii)

nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es

decir, nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el

fragmento del dominio B); o (iv) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir,

nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento

del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad

polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados contiene menos

sitios de unión al promotor potenciales con respecto a SEQ ID NO: 16. En otras realizaciones, la

secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como

máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 sitios potenciales de

unión al promotor. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que

codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2 o como

máximo 1 sitios potenciales de unión al promotor. En otras realizaciones más, la secuencia de

nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un sitio potencial

de unión al promotor.

[0311] En otras realizaciones, el casete genético comprende una secuencia de nucleótidos con codones

optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos comprende una

secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%,

al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al

menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al

menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al

menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los

nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados contiene menos sitios potenciales de unión al promotor con respecto a SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 sitios potenciales de unión al promotor. En otras realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 sitios potenciales de unión al promotor. En otras realizaciones más, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII no contiene un sitio potencial de unión al promotor.

[0312] Las cajas TATA son secuencias reguladoras que se encuentran a menudo en las regiones promotoras de eucariotas. Sirven como el sitio de unión de la proteína de unión TATA (TBP), un factor de transcripción general. Las cajas TATA normalmente comprenden la secuencia TATAA (SEQ ID NO: 28) o una variante cercana. Sin embargo, las cajas TATA dentro de una secuencia codificante pueden inhibir la traducción de la proteína de longitud completa. Existen diez secuencias potenciales de unión al promotor en la secuencia de FVIII BDD de tipo salvaje (SEQ ID NO: 16): cinco secuencias TATAA (SEQ ID NO: 28) y cinco secuencias TTATA (SEQ ID NO: 29). En algunas realizaciones, al menos 1, al menos 2, al menos 3, o al menos 4 de los sitios de unión al promotor están abolidos en los genes de FVIII de la presente divulgación. En algunas realizaciones, al menos 5 de los sitios de unión al promotor están abolidos en los genes del FVIII de la presente divulgación. En otras realizaciones, al menos 6, al menos 7 o al menos 8 de los sitios de unión al promotor están abolidos en los genes del FVIII de la presente divulgación. En una realización, al menos 9 de los sitios de unión al promotor están abolidos en los genes de FVIII de la presente divulgación. En una realización particular, todos los sitios de unión al promotor se eliminan en los genes de FVIII de la presente divulgación. La ubicación de cada sitio de unión del promotor potencial y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

F. Otros Elementos Reguladores Negativos que Actúan en Cis

[0313] Además de las secuencias MAR/ARS, los elementos desestabilizadores y los sitios promotores potenciales descritos anteriormente, se pueden identificar varias secuencias potencialmente inhibidoras adicionales en la secuencia de FVIII BDD de tipo salvaje (SEQ ID NO: 16). Se pueden identificar dos elementos de secuencia ricos en AU (ARE) (ATTTTATT (SEQ ID NO: 30); y ATTTTTAA (SEQ ID

NO: 31), junto con un sitio poli-A (AAAAAAA; SEQ ID NO: 26), un sitio poli-T (TTTTTT; SEQ ID NO: 25) y un sitio de empalme (GGTGAT; SEQ ID NO: 27) en la secuencia de FVIII BDD no optimizada. Uno o más de estos elementos se pueden eliminar de las secuencias optimizadas de FVIII. La ubicación de cada uno de estos sitios y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

[0314] En ciertas realizaciones, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 58-1791 de SEQ ID NO: 3; (ii) nucleótidos 1-1791 de SEQ ID NO: 3; (iii) nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene uno o más elementos reguladores que actúan en cis, por ejemplo, un sitio de corte y empalme, una secuencia de poli-T, una secuencia de poli-A, una secuencia ARE, o cualquier combinación de las mismas.

[0315] En otra realización, la secuencia de nucleótidos optimizada por codones que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el

fragmento del dominio B); o (iv) nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene uno o más elementos reguladores negativos que actúan en cis, por ejemplo, un sitio de corte y empalme, una secuencia de poli-T, una secuencia de poli-A, una secuencia ARE o cualquier combinación de las mismas.

[0316] En otras realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados no contiene uno o más elementos reguladores negativos que actúan en cis, por ejemplo, un sitio de corte y empalme, una secuencia de poli-T, una secuencia de poli-A, una secuencia ARE o cualquier combinación de las mismas.

[0317] En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas

tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 27). En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene una secuencia de poli-T (SEQ ID NO: 25). En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene una secuencia de poli-A (SEQ ID NO: 26). En algunas realizaciones, la secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido de FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido de FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos

aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad polipeptídica de FVIII; y en donde la secuencia de nucleótidos con codones optimizados no contiene un elemento ARE (SEQ ID NO: 30 o SEQ ID NO: 31).

[0318] En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% % de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 (es decir, nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 27). En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98 %, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia

de nucleótidos con codones optimizados no contiene una secuencia de poli-T (SEQ ID NO: 25). En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98 %, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados no contiene una secuencia de poli-A (SEQ ID NO: 26). En algunas realizaciones, el casete genético comprende una secuencia de nucleótidos con codones optimizados que codifica un polipéptido de FVIII, en donde la secuencia de nucleótidos con codones optimizados comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58 -4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada entre SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos con codones optimizados no contiene un elemento ARE (SEQ ID NO: 30 o SEQ ID NO: 31).

[0319] En otras realizaciones, una secuencia optimizada de FVIII de la divulgación no comprende uno o más motivos antivirales, estructuras de tallo-bucle y secuencias repetidas.

[0320] En otras realizaciones más, los nucleótidos que rodean el sitio de inicio de la transcripción se cambian a una secuencia consenso de kozak (GCCGCCACCATGC (SEQ ID NO: 32), en donde los nucleótidos subrayados son el codón de inicio). En otras realizaciones, se pueden añadir o eliminar sitios de restricción para facilitar el proceso de clonación.

b. Secuencias de Polinucleótidos y FIX que Codifican la Proteína FIX

[0321] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido de FIX. En algunas realizaciones, el polipéptido de FIX comprende FIX o una variante o fragmento del mismo, en donde el FIX o la variante o fragmento del mismo tienen actividad de FIX.

[0322] El FIX humano es una serina proteasa que es un componente importante de la ruta intrínseca de la cascada de coagulación sanguínea. "Factor IX" o "FIX", como se emplean en la presente memoria, se refieren a una proteína del factor de coagulación y especies y variantes de secuencia de la misma, e incluyen, pero no se limitan a, la secuencia de 461 aminoácidos de cadena sencilla del polipéptido precursor de FIX humano ("prepro"), la secuencia de 415 aminoácidos de cadena sencilla de FIX humano maduro (SEQ ID NO: 125), y la variante R338L de FIX (Padua) (SEQ ID NO: 126). FIX incluye cualquier forma de molécula de FIX con las características típicas de la coagulación sanguínea FIX. Como se emplea en la presente memoria, se pretende que "Factor IX" y "FIX" abarquen polipéptidos que comprenden los dominios Gla (región que contiene restos de ácido γ -carboxiglutámico), EGF1 y EGF2 (regiones que contienen secuencias homólogas del factor de crecimiento epidérmico humano), péptido de activación ("AP", formado por los residuos R136-R180 del FIX maduro) y el dominio de proteasa C-terminal ("Pro"), o sinónimos de estos dominios conocidos en la técnica, o pueden ser un fragmento truncado o una variante de secuencia que retiene al menos una parte de la actividad biológica de la proteína nativa. Se han clonado FIX o variantes de secuencia, como se describe en las Patentes de Estados Unidos Núm. 4.770.999 y 7.700.734, y se ha aislado, caracterizado y clonado ADNc que codifica FIX humano en vectores de expresión (véanse, por ejemplo, Choo et al., Nature 299:178-180 (1982); Fair et al., Blood 64:194-204 (1984); y Kurachi et al., Proc. Natl. Acad. Sci., USA 79:6461-6464 (1982)). Una variante particular de FIX, la variante R338L FIX (Padua) (SEQ ID NO: 2), caracterizada por Simioni et al., 2009, comprende una mutación de ganancia de función, que se correlaciona con un aumento de casi 8 veces en la actividad de la variante Padua con respecto al FIX nativo (Tabla 4). Las variantes de FIX también pueden incluir cualquier polipéptido de FIX que tenga una o más sustituciones conservativas de aminoácidos, que no afecten a la actividad FIX del polipéptido de FIX. En algunas realizaciones, la variante FIX comprende rFIX-albúmina fusionada mediante un conector escindible, por ejemplo, IDELVION®. Véase el documento US 7.939.632, incorporado en la presente memoria como referencia en su totalidad.

Tabla 4: Ejemplos de Secuencias FIX

<u>SEQ ID NO: 125 (polipéptido de FIX maduro)</u>					
1:	YNSGKLEEFV	QGNLERECME	EKCSFEEARE	VFENTERTE	FWKQYVDGDQ
61:	SCKDDINSYE	CWCPFGFEGK	NCELDVTCNI	KNGRCEQFCK	NSADNKVVCS
121:	QKSCEPAVPF	PCGRVSVSQT	SKLTRAETVF	PDVDYVNSTE	AETILDNITQ
181:	VVGGEDAKPG	QFPWQVVLNG	KVDAFCGGS	VNEKWIPTAA	HCVETGVKIT
241:	TEHTEQKRN	IRIIPHHN	YNAINKYNHDI	ALLELDEPLV	LNSYVTPICI
301:	KFGSGYVSGW	GRVFKHGRSA	LVLQYLRVPL	VDRATCLRST	KFTIYNNMFC
361:	CQGDSSGPHV	TEVEGTSFLT	GIISWGEECA	MKGKYGITYK	VSRYVNWIKE
					KTKLT
<u>SEQ ID NO: 126 (Polipéptido de FIX Padua (R338L) maduro)</u>					
1:	YNSGKLEEFV	QGNLERECME	EKCSFEEARE	VFENTERTE	FWKQYVDGDQ
61:	SCKDDINSYE	CWCPFGFEGK	NCELDVTCNI	KNGRCEQFCK	NSADNKVVCS
121:	QKSCEPAVPF	PCGRVSVSQT	SKLTRAETVF	PDVDYVNSTE	AETILDNITQ
181:	VVGGEDAKPG	QFPWQVVLNG	KVDAFCGGS	VNEKWIPTAA	HCVETGVKIT
241:	TEHTEQKRN	IRIIPHHN	YNAINKYNHDI	ALLELDEPLV	LNSYVTPICI
301:	KFGSGYVSGW	GRVFKHGRSA	LVLQYLRVPL	VDRATCLRST	KFTIYNNMFC
361:	CQGDSSGPHV	TEVEGTSFLT	GIISWGEECA	MKGKYGITYK	VSRYVNWIKE
					KTKLT
<u>SEQ ID NO: 127 (Polipéptido Señal de FIX y Propéptido)</u>					
1:	MQRVNMIMAE	SPGLITICLL	GYLLSAECTV	FLDHENANKI	LNRPKR
<u>SEQ ID NO: 160 (FIX-Conector-Albúmina)</u>					
YNSGKLEEFV	QGNLERECME	EKCSFEEARE	VFENTERTE	FWKQYVDGDQ	50
CESNPCLNGG	SCKDDINSYE	CWCPFGFEGK	NCELDVTCNI	KNGRCEQFCK	100
NSADNKVVCS	CTEGYRLAEN	QKSCEPAVPF	PCGRVSVSQT	SKLTRAETVF	150
PDVDYVNSTE	AETILDNITQ	STQSFNDFTR	VVGGEDAKPG	QFPWQVVLNG	200
KVDAFCGGS	VNEKWIPTAA	HCVETGVKIT	VVAGEHNIEE	TEHTEQKRN	250
IRIIPHHN	YNAINKYNHDI	ALLELDEPLV	LNSYVTPICI	ADKEYTNIFL	300
KFGSGYVSGW	GRVFKHGRSA	LVLQYLRVPL	VDRATCLRST	KFTIYNNMFC	350
AGFHEGGRDS	CQGDSSGPHV	TEVEGTSFLT	GIISWGEECA	MKGKYGITYK	400
VSRYVNWIKE	KTKLTPVSQT	SKLTRAETVF	PDVDAHKSEV	AHRFKDLGEE	450
NFKALVLIAF	AQYLQQCPFE	DHVKLNVET	EFATKCVADE	SAENCDKSLH	500
TLFGDKLCTV	ATLRETYGEM	ADCCAQEQPE	RNECFLOHKD	DNPNLPRLVR	550
PEVDVMCTAF	HDNEETFLKK	YLYEIARRHP	YFYAPELLFF	AKRYKAAATE	600
CCQAADKAAC	LLPKLDELRL	EGKASSAKQR	LKCAQLQKFG	ERAFKAWAVA	650
RLSQRFPKAE	FAEVSKLVTD	LTKVHTECCH	GDLLECADDR	ADLAKYICEN	700
QDSISSKLKE	CCEKPLLEKS	HCIAEVENDE	MPADLPSLAA	DFVESKDVCK	750
NYAEAKDVFL	GMFLYEYARR	HPDYSVVL	LLRLAKTYETTL	EKCCAAADPH	800
ECYAKVFDEF	KPLVEEPQNL	IKQNCLEFEQ	LGEYKFQNAL	LVRYTKKVPQ	850
VSTPTLVEVS	RNLGKVGSKC	CKHPEAKRMP	CAEDYLSVVL	NQLCVLHEKT	900
PVSDRVTKCC	TESLVNRRPC	FSALEVDETY	VPKEFNAETF	TFHADICTLS	950
EKERQIKKQT	ALVELVKHKP	KATKEQLKAV	MDDFAAFVEK	CCKADDKETC	1000
FAEEGKKLVA	ASQAALGL				1018

SEQ ID NO: 161 (FIX)

YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNPCLNGGSKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAEETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIIVTAAHCVETGVKITVVAGEHNIETEHETEQKRNVIIRIIPHHYNAAINKYNHDIALLELDEPLVLNSYVTPICIAADKEYTNIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLRSTKFTIYNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKGIYTKVSRVYVNWIKETKLTTPVSQTSKLT

SEQ ID NO: 162 (Conector)

RAETVFPDV

SEQ ID NO: 163 (Albúmina)

DAHKSEVAHRFKDLGEENFKALVLIIFAQYLQQCPFEDHVKLVNEVTEFAKTCVADESAENCDSLHTLFGDKLCTVATLRETYGEMADCCAQEPERNECFHQHDDNPNLRLVRPEVDVMCTAFHDNEETFLKKYLYEIAARRHPYFYAPELLFFAKRYKAAFTCECQAADKAAACLLPKLDELDEGKASSAKQRLKCASLQKFGERAFAKAWAVARLSQRFPAEFAEVSKLVTDLTGVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYFYARRHPDYSVVLRLAKTYETTLKCCAAADPHECYAKVDFEFKPLVEEPQNLIKQNCLEFQGLGEYKFNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCKADDKETCFAEEGKKLVAASQAALGL

SEQ ID NO: 164 (FIX(XTEN)-Fc)*

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNPCLNGGSKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAEETILDGSPSGSPTSTEEGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGASSNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIIVTAAHCVETGVKITVVAGEHNIETEHETEQKRNVIIRIIPHHYNAAINKYNHDIALLELDEPLVLNSYVTPICIAADKEYTNIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLLSTKFTIYNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKGIYTKVSRVYVNWIKETKLTDKTHTCPPCPAPELLGGPSVFLFPKPKDTLMSRTPPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTIKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCVMHEALHNHYTQKSLSLSPG

SEQ ID NO: 165 (FIX-FX1a-AE288)*

YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNPCLNGGSKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAEETILDNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIIVTAAHCVETGVKITVVAGEHNIETEHETEQKRNVIIRIIPHHYNAAINKYNHDIALLELDEPLVLNSYVTPICIAADKEYTNIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLRSTKFTIYNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKGIYTKVSRVYVNWIKETKLTGKLTRAETGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP

aminoácido que contiene aproximadamente 17% de carbohidratos en peso (Schmidt, AE, et al. (2003) Trends Cardiovasc Med, 13: 39).

[0324] El FIX maduro se compone de varios dominios que en una configuración N- a C-terminal son: un dominio GLA, un dominio EGF1, un dominio EGF2, un dominio de péptido de activación (AP) y un dominio proteasa (o catalítico). Un conector corto conecta el dominio EGF2 con el dominio AP. El FIX contiene dos péptidos de activación formados por R145-A146 y R180-V181, respectivamente. Después de la activación, el FIX monocatenario se convierte en una molécula de 2 cadenas, en la que las dos cadenas están conectadas por un enlace disulfuro. Los factores de coagulación se pueden diseñar reemplazando sus péptidos de activación dando como resultado una especificidad de activación alterada. En los mamíferos, el FIX maduro debe ser activado por el Factor XI activado para producir el Factor IXa. El dominio de proteasa proporciona, tras la activación de FIX a FIXa, la actividad catalítica de FIX.

[0325] En ciertas realizaciones, un polipéptido de FIX comprende una forma alélica Thr148 de FIX derivado de plasma y tiene características estructurales y funcionales similares al FIX endógeno.

[0326] Se conocen en la técnica muchas variantes funcionales de FIX. La Publicación Internacional Núm. WO 02/040544 A3 describe mutantes que exhiben una mayor resistencia a la inhibición por heparina en la página 4, líneas 9-30 y en la página 15, líneas 6-31. La Publicación Internacional Núm. WO 03/020764 A2 describe mutantes de FIX con menor inmunogenicidad de células T en las Tablas 2 y 3 (en las páginas 14-24), y en la página 12, líneas 1-27. La Publicación Internacional Núm. WO 2007/149406 A2 describe moléculas de FIX mutantes funcionales que exhiben una mayor estabilidad proteica, una mayor semivida in vivo e in vitro y una mayor resistencia a las proteasas en la página 4, línea 1 a la página 19, línea 11. El documento WO 2007/149406 A2 también describe moléculas de FIX quiméricas y otras variantes de FIX en la página 19, línea 12 a página 20, línea 9. La Publicación Internacional Núm. WO 08/118507 A2 describe mutantes de FIX que exhiben una mayor actividad de coagulación en la página 5, línea 14 a página 6, línea 5. La Publicación Internacional Núm. WO 09/051717 A2 describe mutantes de FIX que tienen mayor número de sitios de glicosilación conectados a N y conectados a O, que dan como resultado una mayor semivida y/o recuperación en la página 9, línea 11 a página 20, línea 2. La Publicación Internacional Núm. WO 09/137254 A2 también describe mutantes del Factor IX con mayor número de sitios de glicosilación en la página 2, párrafo [006] a página 5, párrafo [011] y página 16, párrafo [044] a página 24, párrafo [057]. La Publicación Internacional Núm. WO 09/130198 A2 describe moléculas de FIX mutantes funcionales que tienen un mayor número de sitios de glicosilación, que dan como resultado una mayor semivida, en la página 4, línea 26 a página 12, línea 6. La Publicación Internacional Núm. WO 09/140015 A2 describe mutantes de FIX funcionales que tienen un mayor número de restos de Cys, que se pueden utilizar para la

conjugación de polímeros (por ejemplo, PEG), en la página 11, párrafo a la página 13, párrafo [0053]. Los polipéptidos de FIX descritos en la Solicitud Internacional Núm. PCT/US2011/043569 presentada el 11 de julio de 2011 y publicada como documento WO 2012/006624 el 12 de enero de 2012 también se incorporan en la presente memoria como referencia en su totalidad. En algunas realizaciones, el polipéptido de FIX comprende un polipéptido de FIX fusionado a una albúmina, por ejemplo, FIX-albúmina. En ciertas realizaciones, el polipéptido de FIX es IDELVION® o rIX-FP.

[0327] Además, se han identificado cientos de mutaciones no funcionales en FIX en sujetos con hemofilia, muchas de las cuales se describen en la Tabla 6, en las páginas 11-14 de la Publicación Internacional Núm. WO 09/137254 A2. Tales mutaciones no funcionales no se incluyen en la invención, pero proporcionan una guía adicional para las mutaciones que es más o menos probable que den como resultado un polipéptido de FIX funcional.

[0328] En una realización, el polipéptido de FIX (o la porción de Factor IX de un polipéptido de fusión) comprende una secuencia de aminoácidos al menos 70%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% idéntica a la secuencia expuesta en SEQ ID NO: 1 o 2 (aminoácidos 1 a 415 de SEQ ID NO: 125 o 126), o alternativamente, con una secuencia de propéptido, o con una secuencia de propéptido y señal (FIX de longitud completa). En otra realización, el polipéptido de FIX comprende una secuencia de aminoácidos al menos 70%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98 %, al menos 99% o 100% idéntica a la secuencia expuesta en SEQ ID NO: 2.

[0329] La actividad de coagulación de FIX se expresa como unidades internacionales (UI). Una UI de actividad de FIX corresponde aproximadamente a la cantidad de FIX en un mililitro de plasma humano normal. Existen varios ensayos disponibles para medir la actividad de FIX, incluido el ensayo de coagulación de una fase (tiempo de tromboplastina parcial activada; aPTT), tiempo de generación de trombina (TGA) y tromboelastometría rotacional (ROTEM®). La invención contempla secuencias que tienen homología con secuencias FIX, fragmentos de secuencia que son naturales, tales como de seres humanos, primates no humanos, mamíferos (incluidos animales domésticos) y variantes de secuencia no naturales que retienen al menos una parte de la actividad biológica o función biológica de FIX y/o que son útiles para prevenir, tratar, mediar o mejorar una enfermedad, deficiencia, trastorno o afección relacionados con el factor de coagulación (p. ej., episodios hemorrágicos relacionados con un traumatismo, cirugía o deficiencia de un factor de coagulación). Las secuencias con homología con FIX humano se pueden encontrar mediante técnicas de búsqueda de homología convencionales, tales como NCBI BLAST.

[0330] En ciertas realizaciones, la secuencia de FIX tiene codones optimizados. Los ejemplos de secuencias FIX con codones optimizados incluyen, pero no se limitan a, SEQ ID NO: 1 y 54-58 de la Publicación Internacional Núm. WO 2016/0041 13 A1, que se incorpora en la presente memoria como referencia en su totalidad.

5 **c. FVII y Secuencias de Polinucleótidos que Codifican la Proteína FVII**

[0331] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido del Factor VII. En algunas realizaciones, el polipéptido de FVII comprende FVII o una variante o fragmento del mismo, en donde la variante o fragmento del mismo tiene actividad de FVII.

10

[0332] El "Factor VII" ("FVII" o "F7;" también denominado Factor 7, factor de coagulación VII, factor VII sérico, acelerador de conversión de protrombina sérica, SPCA, proconvertina y eptacog alfa) es una serina proteasa que es parte de la cascada de coagulación. En una realización, el factor de coagulación en el ácido nucleico descrito en la presente memoria es FVII. El Factor VII activado recombinante ("FVII") se ha vuelto ampliamente utilizado para el tratamiento de hemorragias graves, tales como las que se presentan en pacientes con hemofilia A o B, deficiencia del factor de coagulación XI, FVII, función plaquetaria defectuosa, trombocitopenia o enfermedad de von Willebrand.

15

[0333] El FVII activado recombinante (rFVIIa; NOVOSEVEN®) se utiliza para tratar episodios de sangrado en pacientes con hemofilia (i) con anticuerpos neutralizantes contra FVIII o FIX (inhibidores), (ii) pacientes con deficiencia de FVII, o (iii) pacientes con hemofilia A o B con inhibidores sometidos a procedimientos quirúrgicos. Sin embargo, NOVOSEVEN® muestra poca eficacia. A menudo se requieren dosis repetidas de FVIIa a alta concentración para controlar un sangrado, debido a su baja afinidad por las plaquetas activadas, su corta semivida y su escasa actividad enzimática en ausencia de factor tisular. En consecuencia, existe una necesidad médica insatisfecha de mejores opciones de tratamiento y prevención para pacientes con hemofilia con inhibidores de FVIII y FIX y/o con deficiencia de FVII.

20

25

[0334] En una realización, el casete genético codifica una forma madura de FVII o una variante del mismo. El FVII incluye un dominio Gla, dos dominios EGF (EGF-1 y EGF-2) y un dominio serina proteasa (o dominio de peptidasa S1) que está altamente conservado entre todos los miembros de la familia peptidasa S1 de serina proteasas, tales como por ejemplo con quimotripsina. El FVII se presenta como un zimógeno de cadena sencilla (es decir, FVII activable) y una forma de dos cadenas completamente activada.

30

C. Factores de Crecimiento

[0335] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, donde la secuencia diana codifica una proteína terapéutica y donde la proteína terapéutica comprende un factor de crecimiento. El factor de crecimiento se puede seleccionar entre cualquier factor de crecimiento conocido en la técnica. En algunas realizaciones, el factor de crecimiento es una hormona. En otras realizaciones, el factor de crecimiento es una citocina. En algunas realizaciones, el factor de crecimiento es una quimiocina.

[0336] En algunas realizaciones, el factor de crecimiento es adrenomedulina (AM). En algunas realizaciones, el factor de crecimiento es angiopoyetina (Ang). En algunas realizaciones, el factor de crecimiento es un factor de motilidad autocrina. En algunas realizaciones, el factor de crecimiento es una proteína morfogenética ósea (BMP). En algunas realizaciones, la BMP se selecciona entre BMP2, BMP4, BMP5 y BMP7. En algunas realizaciones, el factor de crecimiento es un miembro de la familia del factor neurotrófico ciliar. En algunas realizaciones, el miembro de la familia del factor neurotrófico ciliar se selecciona entre factor neurotrófico ciliar (CNTF), factor inhibidor de leucemia (LIF), interleucina-6 (IL-6). En algunas realizaciones, el factor de crecimiento es un factor estimulante de colonias. En algunas realizaciones, el factor estimulante de colonias se selecciona entre factor estimulante de colonias de macrófagos (m-CSF), factor estimulante de colonias de granulocitos (G-CSF), y factor estimulante de colonias de granulocitos y macrófagos (GM-CSF). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento epidérmico (EGF). En algunas realizaciones, el factor de crecimiento es una efrina. En algunas realizaciones, la efrina se selecciona entre efrina A1, efrina A2, efrina A3, efrina A4, efrina A5, efrina B1, efrina B2 y efrina B3. En algunas realizaciones, el factor de crecimiento es eritropoyetina (EPO). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento de fibroblastos (FGF). En algunas realizaciones, el FGF se selecciona entre FGF1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF20, FGF21, FGF22 y FGF23. En algunas realizaciones, el factor de crecimiento es somatotropina bovina fetal (FBS). En algunas realizaciones, el factor de crecimiento es un miembro de la familia GDNF. En algunas realizaciones, el miembro de la familia GDNF se selecciona entre factor neurotrófico derivado de la línea celular glial (GDNF), neurturina, persefina y artemina. En algunas realizaciones, el factor de crecimiento es el factor de diferenciación del crecimiento-9 (GDF9). En algunas realizaciones, el factor de crecimiento es factor de crecimiento de hepatocitos (HGF). En algunas realizaciones, el factor de crecimiento es factor de crecimiento derivado de hepatoma (HDGF). En algunas realizaciones, el factor de crecimiento es insulina. En algunas realizaciones, el factor de crecimiento es un factor de crecimiento de tipo insulínico. En algunas realizaciones, el factor de

crecimiento de tipo insulínico es el factor de crecimiento de tipo insulínico 1 (IGF-1) o IGF-2. En algunas realizaciones, el factor de crecimiento es una interleucina (IL). En algunas realizaciones, la IL se selecciona entre IL-1, IL-2, IL-3, IL-4, IL-5, IL-6 e IL-7. En algunas realizaciones, el factor de crecimiento es el factor de crecimiento de queratinocitos (KGF). En algunas realizaciones, el factor de crecimiento es el factor estimulante de la migración (MSF). En algunas realizaciones, el factor de crecimiento es una proteína estimulante de macrófagos (MSP o proteína de tipo factor de crecimiento de hepatocitos (HGFLP)). En algunas realizaciones, el factor de crecimiento es miostatina (GDF-8). En algunas realizaciones, el factor de crecimiento es una neurregulina. En algunas realizaciones, la neurregulina se selecciona entre neurregulina 1 (NRG1), NRG2, NRG3 y NRG4. En algunas realizaciones, el factor de crecimiento es una neurotrofina. En algunas realizaciones, el factor de crecimiento es el factor neurotrófico derivado del cerebro (BDNF). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento nervioso (NGF). En algunas realizaciones, el NGF es neurotrofina-3 (NT-3) o NT-4. En algunas realizaciones, el factor de crecimiento es el factor de crecimiento placentario (PGF). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento derivado de plaquetas (PDGF). En algunas realizaciones, el factor de crecimiento es la renalasa (RNLS). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento de células T (TCGF). En algunas realizaciones, el factor de crecimiento es trombopoyetina (TPO). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento transformador. En algunas realizaciones, el factor de crecimiento transformante es el factor de crecimiento transformante alfa (TGF-a) o TGF-b. En algunas realizaciones, el factor de crecimiento es el factor de necrosis tumoral alfa (TNF-a). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento endotelial vascular (VEGF).

D. Micro ARN (miARN)

[0337] Los microARN (miARN) son pequeñas moléculas de ARN no codificantes (aproximadamente 18-22 nucleótidos) que regulan negativamente la expresión génica inhibiendo la traducción o induciendo la degradación del ARN mensajero (ARNm). Desde su descubrimiento, los miARN han estado implicados en varios procesos celulares, incluida la apoptosis, la diferenciación y la proliferación celular, y se ha demostrado que desempeñan un papel clave en la carcinogénesis. La capacidad de los miARN para regular la expresión génica hace que la expresión de miARN in vivo sea una herramienta valiosa en la terapia génica.

[0338] Ciertos aspectos de la presente divulgación están dirigidos a moléculas de ácido nucleico de tipo plásmido que comprenden una primera ITR, una segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica un miARN, y donde la primera ITR y/o la segunda

ITR son una ITR de un virus no adenoasociado (p. ej., la primera ITR y/o la segunda ITR son de un no AAV). El miARN puede ser cualquier miARN conocido en la técnica. En algunas realizaciones, el miARN regula negativamente la expresión de un gen diana. En ciertas realizaciones, el gen diana se selecciona entre SOD1, HTT, RHO o cualquier combinación de los mismos.

5 [0339] En algunas realizaciones, el casete genético codifica un miARN. En algunas realizaciones, el casete genético codifica más de un miARN. En algunas realizaciones, el casete genético codifica dos o más miARN diferentes. En algunas realizaciones, el casete genético codifica dos o más copias del mismo miARN. En algunas realizaciones, el casete genético codifica dos o más variantes de la misma proteína terapéutica. En ciertas realizaciones, el casete genético codifica uno o más miARN y una o más proteínas
10 terapéuticas.

[0340] En algunas realizaciones, el miARN es un miARN de origen natural. En algunas realizaciones, el miARN es un miARN modificado por ingeniería genética. En algunas realizaciones, el miARN es un miARN artificial. En ciertas realizaciones, el miARN comprende el miARN modificado por ingeniería genética miHTT descrito por Evers et al., *Molecular Therapy* 26 (9):1-15 (epub antes de la impresión de
15 junio de 2018). En ciertas realizaciones, el miARN comprende el miARN artificial miR SOD1 divulgado por Dirren et al., *Annals of Clinical and Translational Neurology* 2(2):167-84 (Febrero de 2015). En ciertas realizaciones, el miARN comprende miR-708, que se dirige a RHO (véase Behrman et al., *JCB* 192(6):919-27 (2011).

[0341] En algunas realizaciones, el miARN regula por incremento la expresión de un gen regulando por
20 disminución la expresión de un inhibidor del gen. En algunas realizaciones, el inhibidor es un inhibidor natural, p. ej., de tipo salvaje. En algunas realizaciones, el inhibidor resulta de un gen mutado, heterólogo y/o mal expresado.

E. Radicales Heterólogos

[0342] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una
25 segunda ITR y un casete genético que codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica y en donde la proteína terapéutica comprende al menos un radical heterólogo. En algunas realizaciones, el radical heterólogo se fusiona con el extremo N-terminal o C-terminal de la proteína terapéutica. En otras realizaciones, el radical heterólogo se inserta entre dos aminoácidos dentro de la proteína terapéutica.

30 [0343] En algunas realizaciones, la proteína terapéutica comprende un polipéptido de FVIII y un radical heterólogo, que se inserta entre dos aminoácidos dentro del polipéptido de FVIII. En algunas

realizaciones, el radical heterólogo se inserta dentro del polipéptido de FVIII en uno o más sitios de inserción seleccionados de la Tabla 5. En algunas realizaciones, la secuencia de aminoácidos heteróloga se puede insertar dentro del polipéptido del factor de coagulación codificado por la molécula de ácido nucleico de la divulgación en cualquier sitio descrito en la Publicación Internacional Núm. WO

2013/123457 A1, WO 2015/106052 A1 o la Publicación de EE.UU. Núm. 2015/0158929 A1, que se incorporan en la presente memoria como referencia en su totalidad. En una realización particular, la proteína terapéutica comprende un FVIII y un radical heterólogo, en donde el radical heterólogo se inserta dentro del FVIII inmediatamente aguas abajo del aminoácido 745 con respecto al FVIII maduro. En una realización particular, la proteína terapéutica comprende un FVIII y un XTEN en donde el XTEN se inserta dentro del FVIII inmediatamente aguas abajo del aminoácido 745 con respecto al FVIII maduro. En una realización particular, el FVIII comprende una delección de los aminoácidos 746-1646, correspondientes al FVIII humano maduro (SEQ ID NO: 15).

Tabla 5: Sitios de Inserción de Radicales Heterólogos de FVIII

Sitio de Inserción	Dominio	Sitio de Inserción	Dominio	Sitio de Inserción	Dominio
3	A1	375	A2	1749	A3
18	A1	378	A2	1796	A3
22	A1	399	A2	1802	A3
26	A1	403	A2	1827	A3
40	A1	409	A2	1861	A3
60	A1	416	A2	1896	A3
65	A1	442	A2	1900	A3
81	A1	487	A2	1904	A3
116	A1	490	A2	1905	A3
119	A1	494	A2	1910	A3
130	A1	500	A2	1937	A3
188	A1	518	A2	2019	A3
211	A1	599	A2	2068	C1
216	A1	603	A2	2111	C1
220	A1	713	A2	2120	C1
224	A1	745	B	2171	C2
230	A1	1656	región a3	2188	C2
333	A1	1711	A3	2227	C2
336	A1	1720	A3	2332	CT
339	A1	1725	A3		

[0344] En algunas realizaciones, la proteína terapéutica comprende un polipéptido de FIX y un radical heterólogo, que se inserta entre dos aminoácidos dentro del polipéptido de FIX. En algunas realizaciones, el radical heterólogo se inserta dentro del polipéptido de FIX en uno o más sitios de inserción seleccionados de la Tabla 5. En algunas realizaciones, la secuencia de aminoácidos heteróloga se puede insertar dentro del polipéptido del factor de coagulación codificado por la molécula de ácido nucleico de la divulgación en cualquier sitio descrito en la Solicitud Internacional Núm. PCT/US2017/015879, que se incorpora en la presente memoria como referencia en su totalidad. En una realización particular, la proteína terapéutica comprende un polipéptido de FIX y un radical heterólogo, en donde el radical heterólogo se inserta dentro del polipéptido de FIX inmediatamente aguas abajo del aminoácido 166 con respecto al FIX maduro. En una realización particular, la proteína terapéutica comprende un polipéptido de FIX y un XTEN, en donde el XTEN se inserta dentro del FIX inmediatamente aguas abajo del aminoácido 166 con respecto al FVIII maduro.

Tabla 6: Sitios de Inserción de Radicales Heterólogos de FIX

Sitio de Inserción	Dominio	Sitio de Inserción	Dominio	Sitio de Inserción	Dominio
52	EGF1	149	AP	257	Catalítico
59	EGF1	162	AP	265	Catalítico
66	EGF1	166	AP	277	Catalítico
80	EGF1	174	AP	283	Catalítico
85	EGF2	188	Catalítico	292	Catalítico
89	EGF2	202	Catalítico	316	Catalítico
103	EGF2	224	Catalítico	341	Catalítico
105	EGF2	226	Catalítico	354	Catalítico
113	EGF2	228	Catalítico	392	Catalítico
129	Conector	230	Catalítico	403	Catalítico
142	Conector	240	Catalítico	413	Catalítico

[0345] En otras realizaciones, las proteínas terapéuticas de la divulgación comprenden adicionalmente dos, tres, cuatro, cinco, seis, siete u ocho secuencias de nucleótidos heterólogas. En algunas realizaciones, todos los radicales heterólogos son idénticos. En algunas realizaciones, al menos un radical heterólogo es diferente de los otros radicales heterólogos. En algunas realizaciones, la divulgación puede comprender dos, tres, cuatro, cinco, seis o más de siete radicales heterólogos en tándem.

[0346] En algunas realizaciones, el radical heterólogo aumenta la semivida (es un "prolongador de la semivida") de la proteína terapéutica.

[0347] En algunas realizaciones, el radical heterólogo es un péptido o un polipéptido con características estructuradas o no estructuradas que están asociadas con la prolongación de la semivida in vivo cuando se incorpora en una proteína de la divulgación. Los ejemplos no limitantes incluyen albúmina, fragmentos de albúmina, fragmentos Fc de inmunoglobulinas, el péptido C-terminal (CTP) de la

subunidad b de la gonadotropina coriónica humana, una secuencia HAP, una secuencia XTEN, una transferrina o un fragmento de la misma, un polipéptido PAS, conectores de poliglicina, conectores de poliserina, radicales de unión a albúmina, o cualquier fragmento, derivado, variante o combinación de estos polipéptidos. En una realización particular, la secuencia de aminoácidos heteróloga es una región constante de inmunoglobulina o una porción de la misma, transferrina, albúmina o una secuencia PAS.

En algunos aspectos, un radical heterólogo incluye factor de von Willebrand o un fragmento del mismo. En otros aspectos relacionados, un radical heterólogo puede incluir un sitio de unión (p. ej., Un aminoácido cisteína) para un radical no polipeptídico como polietilenglicol (PEG), hidroxietil almidón (HES), ácido polisiálico o cualquier derivado, variante o combinaciones de estos elementos. En algunos aspectos, un radical heterólogo comprende un aminoácido cisteína que funciona como un sitio de unión para un radical no polipeptídico tal como polietilenglicol (PEG), hidroxietil almidón (HES), ácido polisiálico o cualquier derivado, variante o combinación de estos elementos.

[0348] En una realización específica, un primer radical heterólogo es una molécula que prolonga la semivida que se conoce en la técnica, y un segundo radical heterólogo es una molécula que prolonga la semivida que se conoce en la técnica. En ciertas realizaciones, el primer radical heterólogo (p. ej., un primer radical Fc) y el segundo radical heterólogo (p. ej., un segundo radical Fc) se asocian entre sí para formar un dímero. En una realización, el segundo radical heterólogo es un segundo radical Fc, en donde el segundo radical Fc está unido o asociado con el primer radical heterólogo, por ejemplo, el primer radical Fc. Por ejemplo, el segundo radical heterólogo (p. ej., el segundo radical Fc) se puede conectar al primer radical heterólogo (p. ej., el primer radical Fc) mediante un conector o asociado con el primer radical heterólogo mediante un enlace covalente o no covalente.

[0349] En algunas realizaciones, el radical heterólogo es un polipéptido que comprende, consiste esencialmente en, o consiste en al menos aproximadamente 10, al menos aproximadamente 100, al menos aproximadamente 200, al menos aproximadamente 300, al menos aproximadamente 400, al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1000, al menos aproximadamente 1100, al menos aproximadamente 1200, al menos aproximadamente 1300, al menos aproximadamente 1400, al menos aproximadamente 1500, al menos aproximadamente 1600, al menos aproximadamente 1700, al menos aproximadamente 1800, al menos aproximadamente 1900, al

menos aproximadamente 2000, al menos aproximadamente 2500, al menos aproximadamente 3000, o al menos aproximadamente 4000 aminoácidos. En otras realizaciones, el radical heterólogo es un polipéptido que comprende, que consiste esencialmente en, o consiste en aproximadamente 100 a aproximadamente 200 aminoácidos, aproximadamente 200 a aproximadamente 300 aminoácidos, aproximadamente 300 a aproximadamente 400 aminoácidos, aproximadamente 400 a aproximadamente 500 aminoácidos, aproximadamente 500 a aproximadamente 600 aminoácidos, aproximadamente 600 a aproximadamente 700 aminoácidos, aproximadamente 700 a aproximadamente 800 aminoácidos, aproximadamente 800 a aproximadamente 900 aminoácidos, o aproximadamente 900 a aproximadamente 1000 aminoácidos.

- 10 **[0350]** En ciertas realizaciones, un radical heterólogo mejora una o más propiedades farmacocinéticas de la proteína terapéutica sin afectar significativamente a su actividad o función biológica.

- [0351]** En ciertas realizaciones, un radical heterólogo aumenta la semivida *in vivo* y/o *in vitro* de la proteína terapéutica de la divulgación. En otras realizaciones, un radical heterólogo facilita la visualización o localización de la proteína terapéutica de la divulgación o un fragmento de la misma (p. ej., un fragmento que comprende un radical heterólogo después de la escisión proteolítica de la proteína FVIII). La visualización y/o localización de la proteína terapéutica de la divulgación o un fragmento de la misma puede ser *in vivo*, *in vitro*, *ex vivo* o combinaciones de las mismas.
- 15

- [0352]** En otras realizaciones, un radical heterólogo aumenta la estabilidad de la proteína terapéutica de la divulgación o un fragmento de la misma (p. ej., un fragmento que comprende un radical heterólogo después de la escisión proteolítica de la proteína terapéutica, por ejemplo, un factor de coagulación). Como se emplea en la presente memoria, el término "estabilidad" se refiere a una medida reconocida en la técnica del mantenimiento de una o más propiedades físicas de la proteína terapéutica en respuesta a una condición ambiental (p. ej., una temperatura elevada o reducida). En ciertos aspectos, la propiedad física puede ser el mantenimiento de la estructura covalente de la proteína terapéutica (p. ej., la ausencia de escisión proteolítica, oxidación o desamidación no deseadas). En otros aspectos, la propiedad física también puede ser la presencia de la proteína terapéutica en un estado correctamente plegado (p. ej., la ausencia de agregados o precipitados solubles o insolubles). En un aspecto, la estabilidad de la proteína terapéutica se mide sometiendo a ensayo una propiedad biofísica de la proteína terapéutica, por ejemplo, estabilidad térmica, perfil de desplegamiento con el pH, eliminación estable de glicosilación, solubilidad, función bioquímica (p. ej., capacidad para unirse a una proteína, receptor o ligando), etc., y/o combinaciones de los mismos. En otro aspecto, la función bioquímica se demuestra por la afinidad de unión de la interacción. En un aspecto, una medida de la estabilidad de la proteína es la estabilidad térmica, es decir, la resistencia al reto térmico. La estabilidad se puede medir utilizando métodos
- 20
- 25
- 30

conocidos en la técnica, tales como HPLC (cromatografía líquida de alta resolución), SEC (cromatografía de exclusión de tamaño), DLS (dispersión dinámica de luz), etc. Los métodos para medir la estabilidad térmica incluyen, pero no se limitan a calorimetría diferencial de barrido (DSC), fluorimetría diferencial de barrido (DSF), dicroísmo circular (CD), y ensayo de reto térmico.

5 **[0353]** En ciertos aspectos, una proteína terapéutica codificada por la molécula de ácido nucleico de la divulgación comprende al menos un prolongador de la semivida, es decir, un radical heterólogo que aumenta la semivida *in vivo* de la proteína terapéutica con respecto a la semivida *in vivo* de la proteína terapéutica correspondiente que carece de tal radical heterólogo. La semivida *in vivo* de una proteína terapéutica se puede determinar mediante cualquier método conocido por los expertos en la técnica, p.

10 ej., ensayos de actividad (p. ej., ensayo cromogénico o ensayo de aPTT de coagulación en una etapa en el que la proteína terapéutica comprende un polipéptido de FVIII), ELISA, ROTEM[®], etc.

[0354] En algunas realizaciones, la presencia de uno o más prolongadores de la semivida da como resultado que la semivida de la proteína terapéutica aumente en comparación con la semivida de la proteína correspondiente que carece de uno o más prolongadores de la semivida. La semivida de la

15 proteína terapéutica que comprende un prolongador de la semivida es al menos aproximadamente 1,5 veces, al menos aproximadamente 2 veces, al menos aproximadamente 2,5 veces, al menos aproximadamente 3 veces, al menos aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos aproximadamente 10 veces, al menos

20 aproximadamente 11 veces, o al menos aproximadamente 12 veces más que la semivida *in vivo* de la proteína terapéutica correspondiente que carece de tal prolongador de la semivida.

[0355] En una realización, la semivida de la proteína terapéutica que comprende un prolongador de la semivida es de aproximadamente 1,5 veces a aproximadamente 20 veces, de aproximadamente 1,5 veces a aproximadamente 15 veces, o de aproximadamente 1,5 veces a aproximadamente 10 veces más larga

25 que la semivida *in vivo* de la proteína correspondiente que carece de tal prolongador de la semivida. En otra realización, la semivida de la proteína terapéutica que comprende un prolongador de la semivida se prolonga de aproximadamente 2 veces a aproximadamente 10 veces, de aproximadamente 2 veces a aproximadamente 9 veces, de aproximadamente 2 veces a aproximadamente 8 veces, de aproximadamente 2 veces a aproximadamente 7 veces, de aproximadamente 2 veces a aproximadamente 6 veces, de aproximadamente 2 veces a aproximadamente 5 veces, de aproximadamente 2 veces a

30 aproximadamente 4 veces, de aproximadamente 2 veces a aproximadamente 3 veces, de aproximadamente 2,5 a aproximadamente 10 veces, de aproximadamente 2,5 a aproximadamente 9 veces, de aproximadamente 2,5 a aproximadamente 8 veces, de aproximadamente 2,5 a

aproximadamente 7 veces, de aproximadamente 2,5 a aproximadamente 6 veces, de aproximadamente 2,5 veces a aproximadamente 5 veces, de aproximadamente 2,5 veces a aproximadamente 4 veces, de aproximadamente 2,5 veces a aproximadamente 3 veces, de aproximadamente 3 veces a aproximadamente 10 veces, de aproximadamente 3 veces a aproximadamente 9 veces, de aproximadamente 3 veces a aproximadamente 8 veces, de aproximadamente 3 veces a aproximadamente 7 veces, de aproximadamente 3 veces a aproximadamente 6 veces, de aproximadamente 3 veces a aproximadamente 5 veces, de aproximadamente 3 veces a aproximadamente 4 veces, de aproximadamente 4 veces a aproximadamente 6 veces, de aproximadamente 5 veces a aproximadamente 7 veces, o de aproximadamente 6 veces a aproximadamente 8 veces en comparación con la semivida *in vivo* de la proteína correspondiente que carece de tal prolongador de la semivida.

[0356] En otras realizaciones, la semivida de la proteína terapéutica que comprende un prolongador de la semivida es de al menos aproximadamente 17 horas, al menos aproximadamente 18 horas, al menos aproximadamente 19 horas, al menos aproximadamente 20 horas, al menos aproximadamente 21 horas, al menos aproximadamente 22 horas, al menos aproximadamente 23 horas, al menos aproximadamente 24 horas, al menos aproximadamente 25 horas, al menos aproximadamente 26 horas, al menos aproximadamente 27 horas, al menos aproximadamente 28 horas, al menos aproximadamente 29 horas, al menos aproximadamente 30 horas, al menos aproximadamente 31 horas, al menos aproximadamente 32 horas, al menos aproximadamente 33 horas, al menos aproximadamente 34 horas, al menos aproximadamente 35 horas, al menos aproximadamente 36 horas, al menos aproximadamente 48 horas, al menos aproximadamente 60 horas, al menos aproximadamente 72 horas, al menos aproximadamente 84 horas, al menos aproximadamente 96 horas o al menos aproximadamente 108 horas.

[0357] En otras realizaciones más, la semivida de la proteína terapéutica que comprende un prolongador de la semivida es de aproximadamente 15 horas a aproximadamente dos semanas, de aproximadamente 16 horas a aproximadamente una semana, de aproximadamente 17 horas a aproximadamente una semana, de aproximadamente 18 horas a aproximadamente una semana, de aproximadamente 19 horas a aproximadamente una semana, de aproximadamente 20 horas a aproximadamente una semana, de aproximadamente 21 horas a aproximadamente una semana, de aproximadamente 22 horas a aproximadamente una semana, de aproximadamente 23 horas a aproximadamente una semana, de aproximadamente 24 horas a aproximadamente una semana, de aproximadamente 36 horas a aproximadamente una semana, de aproximadamente 48 horas a aproximadamente una semana, de aproximadamente 60 horas a aproximadamente una semana, de aproximadamente 24 horas a aproximadamente seis días, de aproximadamente 24 horas a aproximadamente cinco días, de

aproximadamente 24 horas a aproximadamente cuatro días, de aproximadamente 24 horas a aproximadamente tres días, o de aproximadamente 24 horas a aproximadamente dos días.

- [0358]** En algunas realizaciones, la semivida promedio por sujeto de la proteína terapéutica que comprende un prolongador de la semivida es de aproximadamente 15 horas, aproximadamente 16 horas, 5 aproximadamente 17 horas, aproximadamente 18 horas, aproximadamente 19 horas, aproximadamente 20 horas, aproximadamente 21 horas, aproximadamente 22 horas, aproximadamente 23 horas, aproximadamente 24 horas (1 día), aproximadamente 25 horas, aproximadamente 26 horas, aproximadamente 27 horas, aproximadamente 28 horas, aproximadamente 29 horas, aproximadamente 30 horas, aproximadamente 31 horas, aproximadamente 32 horas, aproximadamente 33 horas, 10 aproximadamente 34 horas, aproximadamente 35 horas, aproximadamente 36 horas, aproximadamente 40 horas, aproximadamente 44 horas, aproximadamente 48 horas (2 días), aproximadamente 54 horas, aproximadamente 60 horas, aproximadamente 72 horas (3 días), aproximadamente 84 horas, aproximadamente 96 horas (4 días), aproximadamente 108 horas, aproximadamente 120 horas (5 días), aproximadamente seis días, aproximadamente siete días (una semana), aproximadamente ocho días, 15 aproximadamente nueve días, aproximadamente 10 días, aproximadamente 11 días, aproximadamente 12 días, aproximadamente 13 días o aproximadamente 14 días.

[0359] Se pueden fusionar uno o más prolongadores de la semivida al extremo C-terminal o N-terminal de la proteína terapéutica o insertar dentro de la proteína terapéutica.

1. Una región Constante de Inmunoglobulina o una Porción de la Misma

- 20 **[0360]** En otro aspecto, un radical heterólogo comprende una o más regiones constantes de inmunoglobulina o porciones de las mismas (p. ej., una región Fc). En una realización, una molécula de ácido nucleico aislada de la divulgación comprende adicionalmente una secuencia de ácido nucleico heteróloga que codifica una región constante de inmunoglobulina o una porción de la misma. En algunas realizaciones, la región constante de inmunoglobulina o una porción de la misma son regiones Fc.
- 25 **[0361]** Una región constante de inmunoglobulina se compone de dominios denominados dominios CH (pesados constantes) (CH1, CH2, etc.). Dependiendo del isotipo (es decir, IgG, IgM, IgA IgD o IgE), la región constante puede estar compuesta por tres o cuatro dominios CH. Algunas regiones constantes de isotipos (p. ej., IgG) también contienen una región bisagra. Véase Janeway *et al.* 2001, *Immunobiology*, Garland Publishing, N.Y., N.Y.
- 30 **[0362]** Se puede obtener una región constante de inmunoglobulina o una porción de la misma de la presente divulgación a partir de varias fuentes diferentes. En una realización, una región constante de

inmunoglobulina o una porción de la misma se obtiene de una inmunoglobulina humana. Se entiende, sin embargo, que la región constante de inmunoglobulina o una porción de la misma se pueden obtener de una inmunoglobulina de otra especie de mamífero, incluyendo, por ejemplo, una especie de roedor (p. ej., un ratón, rata, conejo, cobaya) o de primate no humano (p. ej., chimpancé, macaco). Además, la región constante de inmunoglobulina o una porción de la misma se pueden obtener de cualquier clase de inmunoglobulina, incluidas IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de inmunoglobulina, incluidas IgG 1, IgG2, IgG3 e IgG4. En una realización, se utiliza el isotipo IgG1 humano.

[0363] Una variedad de secuencias de genes de la región constante de inmunoglobulina (p. ej., secuencias de genes de la región constante humana) están disponibles en forma de depósitos de acceso público. Se puede seleccionar que la secuencia de dominios de la región constante tenga una función efectora particular (o que carezca de una función efectora concreta) o con una modificación concreta para reducir la inmunogenicidad. Se han publicado muchas secuencias de anticuerpos y genes que codifican anticuerpos y se pueden obtener secuencias de la región constante de Ig adecuadas (p. ej., secuencias bisagra, CH2 y/o CH3, o porciones de las mismas) a partir de estas secuencias utilizando mecanismos reconocidos en la técnica. El material genético obtenido utilizando cualquiera de los métodos anteriores se puede modificar o sintetizar para obtener polipéptidos de la presente divulgación. Se apreciará adicionalmente que el alcance de esta divulgación abarca alelos, variantes y mutaciones de secuencias de ADN de la región constante.

[0364] Las secuencias de la región constante de inmunoglobulina o una porción de la misma se pueden clonar, p. ej., utilizando la reacción en cadena de la polimerasa y cebadores que se seleccionan para amplificar el dominio de interés. Para clonar una secuencia de la región constante de inmunoglobulina o una porción de la misma de un anticuerpo, se puede aislar ARNm de hibridoma, bazo o células linfáticas, transcribir de forma inversa a ADN y amplificar los genes de anticuerpos mediante PCR. Los métodos de amplificación por PCR se describen en detalle en la Patente de Estados Unidos Núm. 4.683.195; 4.683.202; 4.800.159; 4.965.188; y, por ejemplo, en "PCR Protocols: A Guide to Methods and Applications" Innis et al. eds., Academic Press, San Diego, CA (1990); Ho et al. 1989. *Gene* 77:51; Horton et al. 1993. *Methods Enzymol.* 217:270). La PCR se puede iniciar mediante cebadores de la región constante consenso o mediante cebadores más específicos basados en las secuencias de aminoácidos y ADN de cadena pesada y ligera publicadas. La PCR también se puede utilizar para aislar clones de ADN que codifican las cadenas ligeras y pesadas del anticuerpo. En este caso, las bibliotecas se pueden escrutar mediante cebadores consenso o sondas homólogas más grandes, tales como sondas de la región constante de ratón. Se conocen en la técnica numerosos conjuntos de cebadores adecuados para la amplificación de genes de anticuerpos (p. ej., cebadores 5' basados en la secuencia N-terminal de

anticuerpos purificados (Benhar y Pastan. 1994. *Protein Engineering* 7: 1509); amplificación rápida de extremos de ADNc (Ruberti, F. et al. 1994. *J. Immunol. Methods* 173: 33); secuencias líder de anticuerpos (Larrick et al. 1989 *Biochem. Biophys. Res. Commun.* 160:1250). La clonación de secuencias de anticuerpos es descrita adicionalmente por Newman et al., en la Patente de Estados Unidos Núm.

5 5.658.570, presentada el 25 de enero de 1995, que se incorpora en la presente memoria como referencia.

[0365] Una región constante de inmunoglobulina utilizada en la presente memoria puede incluir todos los dominios y la región bisagra o porciones de los mismos. En una realización, la región constante de inmunoglobulina o una porción de la misma comprenden el dominio CH2, el dominio CH3 y una región bisagra, es decir, una región Fc o un compañero de unión a FcRn.

10 **[0366]** Como se emplea en la presente memoria, el término "región Fc" se define como la porción de un polipéptido que corresponde a la región Fc de la Ig nativa, es decir, formada por la asociación dimérica de los respectivos dominios Fc de sus dos cadenas pesadas. Una región Fc nativa forma un homodímero con otra región Fc. En contraste, el término "región Fc fusionada genéticamente" o "región Fc monocatenaria" (región scFc), como se emplea en la presente memoria, se refiere a una región Fc
15 dimérica sintética compuesta por dominios Fc genéticamente unidos dentro de una única cadena polipeptídica (es decir, codificados en una sola secuencia genética contigua). Véase la Publicación Internacional Núm. WO 2012/006635, incorporada en la presente memoria como referencia en su totalidad.

[0367] En una realización, la "región Fc" se refiere a la porción de una única cadena pesada de Ig que
20 comienza en la región bisagra justo antes del sitio de escisión de papaína (es decir, el residuo 216 en IgG, tomando el primer residuo de la región constante de cadena pesada 114) y termina en el extremo C-terminal del anticuerpo. Por consiguiente, una región Fc completa comprende al menos un dominio bisagra, un dominio CH2 y un dominio CH3.

[0368] Una región constante de inmunoglobulina o una porción de la misma puede ser un compañero de
25 unión de FcRn. El FcRn es activo en tejidos epiteliales adultos y se expresa en el lumen de los intestinos, las vías respiratorias pulmonares, las superficies nasales, las superficies vaginales, las superficies del colon y del recto (Patente de Estados Unidos Núm. 6.485.726). Un compañero de unión de FcRn es una porción de una inmunoglobulina que se une a FcRn.

[0369] El receptor FcRn se ha aislado de varias especies de mamíferos, incluidos los seres humanos. Se
30 conocen las secuencias del FcRn humano, FcRn de mono, FcRn de rata y FcRn de ratón (Story et al. 1994, *J. Exp. Med.* 180: 2377). El receptor FcRn se une a IgG (pero no a otras clases de inmunoglobulinas tales como IgA, IgM, IgD e IgE) a un pH relativamente bajo, transporta activamente la

IgG transcelularmente en una dirección luminal a serosa y a continuación libera la IgG a un pH relativamente más alto que se encuentra en los fluidos intersticiales. Se expresa en tejido epitelial adulto (Patentes de Estados Unidos Núm. 6.485.726, 6.030.613, 6.086.875; WO 03/077834; US2003-0235536A1) incluyendo epitelio pulmonar e intestinal (Israel et al. 1997, Immunology 92:69) epitelio tubular proximal renal (Kobayashi et al.2002, Am. J. Physiol. Renal Physiol. 282:F358), así como epitelio nasal, superficies vaginales, y superficies del árbol biliar.

[0370] Los compañeros de unión de FcRn útiles en la presente divulgación abarcan moléculas que pueden unirse específicamente por el receptor de FcRn que incluye IgG completa, el fragmento Fc de IgG y otros fragmentos que incluyen la región de unión completa del receptor de FcRn. La región de la porción Fc de IgG que se une al receptor FcRn se ha descrito basándose en cristalografía de rayos X (Burmeister et al. 1994, Nature 372:379). La zona de contacto principal del Fc con el FcRn está cerca de la unión de los dominios CH2 y CH3. Los contactos Fc-FcRn están todos dentro de una sola cadena pesada de Ig. Los compañeros de unión de FcRn incluyen IgG completa, el fragmento Fc de IgG y otros fragmentos de IgG que incluyen la región de unión completa de FcRn. Los principales sitios de contacto incluyen los restos de aminoácido 248, 250-257, 272, 285, 288, 290-291, 308-311, y 314 del dominio CH2 y los restos de aminoácido 385-387, 428 y 433-436 del dominio CH3. Las referencias hechas a la numeración de aminoácidos de inmunoglobulinas o fragmentos o regiones de inmunoglobulinas se basan todas en Kabat et al. 1991, Sequences of Proteins of Immunological Interest, Departamento de Salud Pública de EE. UU., Bethesda, Md.

[0371] Las regiones Fc o los compañeros de unión de FcRn unidos a FcRn se pueden transportar eficazmente a través de barreras epiteliales mediante FcRn, proporcionando así un medio no invasivo para administrar sistémicamente una molécula terapéutica deseada. Además, las proteínas de fusión que comprenden una región Fc o un compañero de unión de FcRn son sometidas a endocitosis por células que expresan el FcRn. Pero en lugar de marcarse para su degradación, estas proteínas de fusión se reciclan de nuevo a la circulación, aumentando así la semivida in vivo de estas proteínas. En ciertas realizaciones, las porciones de las regiones constantes de inmunoglobulina son una región Fc o un compañero de unión de FcRn que típicamente se asocian, a través de enlaces disulfuro u otras interacciones no específicas, con otra región Fc u otro compañero de unión para formar dímeros y multímeros de orden superior.

[0372] Dos receptores de FcRn pueden unirse a una única molécula de Fc. Los datos cristalográficos sugieren que cada molécula de FcRn se une a un único polipéptido del homodímero de Fc. En una realización, la unión del compañero de unión de FcRn, por ejemplo, un fragmento Fc de una IgG, a una molécula biológicamente activa proporciona un medio para administrar la molécula biológicamente

activa por vía oral, bucal, sublingual, rectal, vaginal, como un aerosol administrado por vía nasal o vía pulmonar o vía ocular. En otra realización, la proteína del factor de coagulación se puede administrar de forma invasiva, p. ej., subcutáneamente, intravenosamente.

[0373] Una región compañera de unión a FcRn es una molécula o porción de la misma que puede unirse específicamente al receptor FcRn con el consiguiente transporte activo por el receptor FcRn de la región Fc. Unida específicamente se refiere a dos moléculas que forman un complejo que es relativamente estable en condiciones fisiológicas. La unión específica se caracteriza por una alta afinidad y una capacidad de baja a moderada, a diferencia de la unión no específica, que normalmente tiene una baja afinidad con una capacidad de moderada a alta. Normalmente, la unión se considera específica cuando la constante de afinidad KA es superior a 10^6 M^{-1} o superior a 10^8 M^{-1} . Si fuera necesario, la unión no específica se puede reducir sin afectar sustancialmente a la unión específica variando las condiciones de unión. Las condiciones de unión apropiadas tales como la concentración de las moléculas, la fuerza iónica de la solución, la temperatura, el tiempo permitido para la unión, la concentración de un agente de bloqueo (p. ej., albúmina de suero, caseína de leche), etc., pueden ser optimizadas por un experto en la técnica utilizando técnicas de rutina.

[0374] En ciertas realizaciones, una proteína terapéutica codificada por la molécula de ácido nucleico de la divulgación comprende una o más regiones Fc truncadas que, no obstante, son suficientes para conferir propiedades de unión al receptor de Fc (FcR) a la región Fc. Por ejemplo, la porción de una región Fc que se une a FcRn (es decir, la porción de unión a FcRn) comprende aproximadamente los aminoácidos 282-438 de IgG1, numeración EU (siendo los sitios de contacto primarios los aminoácidos 248, 250-257, 272, 285, 288, 290-291, 308-311 y 314 del dominio CH2 y restos de aminoácido 385-387, 428 y 433-436 del dominio CH3). Por tanto, una región Fc de la divulgación puede comprender o consistir en una porción de unión a FcRn. Las porciones de unión de FcRn se pueden obtener de cadenas pesadas de cualquier isotipo, incluidas IgG1, IgG2, IgG3 e IgG4. En una realización, se utiliza una porción de unión a FcRn de un anticuerpo del isotipo IgG1 humano. En otra realización, se utiliza una porción de unión a FcRn de un anticuerpo del isotipo IgG4 humano.

[0375] La región Fc se puede obtener de varias fuentes diferentes. En una realización, una región Fc del polipéptido se obtiene de una inmunoglobulina humana. Se entiende, sin embargo, que un radical Fc se puede obtener de una inmunoglobulina de otra especie de mamífero, incluyendo, por ejemplo, una especie de roedor (p. ej., un ratón, rata, conejo, cobaya) o de primate no humano (p. ej., chimpancé, macaco). Además, el polipéptido de los dominios Fc o porciones del mismo se pueden obtener de cualquier clase de inmunoglobulina, incluidas IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de

inmunoglobulina, incluidas IgG1, IgG2, IgG3 e IgG4. En otra realización, se utiliza el isotipo humano IgG1.

[0376] En ciertas realizaciones, la variante Fc confiere un cambio en al menos una función efectora impartida por un radical Fc que comprende dicho dominio Fc de tipo salvaje (p. ej., una mejora o reducción en la capacidad de la región Fc para unirse a receptores de Fc (p. ej., FcγRI, FcγRII o FcγRIII) o proteínas del complemento (p. ej., C1q) o para desencadenar citotoxicidad dependiente de anticuerpos (ADCC), fagocitosis o citotoxicidad dependiente del complemento (CDCC)). En otras realizaciones, la variante de Fc proporciona un residuo de cisteína modificado por ingeniería genética.

[0377] La región Fc de la divulgación puede emplear variantes de Fc reconocidas en la técnica que se sabe que imparten un cambio (p. ej., una mejora o reducción) en la función efectora y/o la unión de FcR o FcRn. Específicamente, una región Fc de la divulgación puede incluir, por ejemplo, un cambio (p. ej., una sustitución) en una o más de las posiciones de aminoácidos descritas en las publicaciones internacionales PCT WO88/07089A1, WO96/14339A1, WO98/05787A1, WO98/23289A1, WO99/51642A1, WO99/58572A1, WO00/09560A2, WO00/32767A1, WO00/42072A2, WO02/44215A2, WO02/060919A2, WO03/074569A2, WO04/016750A2, WO04/029207A2, WO04/035752A2, WO04/063351A2, WO04/074455A2, WO04/099249A2, WO05/040217A2, WO04/044859, WO05/070963A1, WO05/077981A2, WO05/092925A2, WO05/123780A2, WO06/019447A1, WO06/047350A2, y WO06/085967A2; Publicaciones de Patente de Estados Unidos Núm. US2007/0231329, US2007/0231329, US2007/0237765, US2007/0237766, US2007/0237767, US2007/0243188, US20070248603, US20070286859, US20080057056; o Patentes de Estados Unidos 5.648.260; 5.739.277; 5.834.250; 5.869.046; 6.096.871; 6.121.022; 6.194.551; 6.242.195; 6.277.375; 6.528.624; 6.538.124; 6.737.056; 6.821.505; 6.998.253; 7.083.784; 7.404.956 y 7.317.091, cada una de los cuales se incorpora en la presente memoria como referencia. En una realización, se puede realizar el cambio específico (p. ej., la sustitución específica de uno o más aminoácidos descritos en la técnica) en una o más de las posiciones de aminoácidos descritas. En otra realización, se puede realizar un cambio diferente en una o más de las posiciones de aminoácidos descritas (p. ej., la sustitución diferente de una o más posiciones de aminoácidos descritas en la técnica).

[0378] La región Fc o el compañero de unión a FcRn de IgG se puede modificar de acuerdo con procedimientos bien reconocidos tales como mutagénesis dirigida al sitio y similares para producir fragmentos de IgG o Fc modificados o porciones de los mismos que se unirán por FcRn. Tales modificaciones incluyen modificaciones remotas de los sitios de contacto de FcRn, así como modificaciones dentro de los sitios de contacto que conservan o incluso mejoran la unión a FcRn. Por ejemplo, los siguientes restos de aminoácido individuales en Fc de IgG1 (Fc γ1) humana se pueden

sustituir sin una pérdida significativa de la afinidad de unión de Fc por FcRn: P238A, S239A, K246A, K248A, D249A, M252A, T256A, E258A, T260A, D265A, S267A, H268A, E269A, D270A, E272A, L274A, N276A, Y278A, D280A, V282A, E283A, H285A, N286A, T289A, K290A, R292A, E293A, E294A, Q295A, Y296F, N297A, S298A, Y300F, R301A, V303A, V305A, T307A, L309A, Q311A, D312A, N315A, K317A, E318A, K320A, K322A, S324A, K326A, A327Q, P329A, A330Q, P331A, E333A, K334A, T335A, S337A, K338A, K340A, Q342A, R344A, E345A, Q347A, R355A, E356A, M358A, T359A, K360A, N361A, Q362A, Y373A, S375A, D376A, A378Q, E380A, E382A, S383A, N384A, Q386A, E388A, N389A, N390A, Y391F, K392A, L398A, S400A, D401A, D413A, K414A, R416A, Q418A, Q419A, N421A, V422A, S424A, E430A, N434A, T437A, Q438A, K439A, S440A, S444A, y K447A, donde por ejemplo P238A representa prolina de tipo salvaje sustituida por alanina en la posición 238. Como ejemplo, una realización específica incorpora la mutación N297A, eliminando un sitio de N-glicosilación altamente conservado. Además de la alanina, se pueden sustituir otros aminoácidos por los aminoácidos de tipo salvaje en las posiciones especificadas anteriormente. Las mutaciones pueden introducirse individualmente en Fc dando lugar a más de cien regiones Fc distintas de la Fc nativa. Además, se pueden introducir juntas combinaciones de dos, tres o más de estas mutaciones individuales, dando lugar a cientos de regiones Fc más.

[0379] Algunas de las mutaciones anteriores pueden conferir nueva funcionalidad a la región Fc o compañero de unión a FcRn. Por ejemplo, una realización incorpora N297A, eliminando un sitio de N-glicosilación altamente conservado. El efecto de esta mutación es reducir la inmunogenicidad, mejorando así la semivida circulante de la región Fc y hacer que la región Fc sea incapaz de unirse a FcγRI, FcγRIIA, FcγRIIB y FcγRIIIA, sin comprometer la afinidad por FcRn (Routledge et al. 1995, Transplantation 60:847; Friend et al. 1999, Transplantation 68:1632; Shields et al. 1995, J. Biol. Chem. 276: 6591). Como ejemplo adicional de la nueva funcionalidad que surge de las mutaciones descritas anteriormente, la afinidad por FcRn se puede incrementar más allá de la del tipo salvaje en algunos casos. Esta mayor afinidad puede reflejar una mayor tasa de "asociación", una menor tasa de "disociación" o tanto una mayor tasa de "asociación" como una menor tasa de "disociación".

[0380] Además, al menos tres receptores gamma de Fc humanos parecen reconocer un sitio de unión en IgG dentro de la región de bisagra inferior, generalmente los aminoácidos 234-237. Por lo tanto, puede surgir otro ejemplo de nueva funcionalidad y posible disminución de la inmunogenicidad a partir de mutaciones de esta región, como, por ejemplo, al reemplazar los aminoácidos 233-236 de IgG1 "ELLG" humano (SEQ ID NO: 45) por la secuencia correspondiente de IgG2 "PVA" (con una delección de un aminoácido). Se ha demostrado que FcγRI, FcγRII y FcγRIII, que median diversas funciones efectoras,

no se unirán a IgG 1 cuando se hayan introducido tales mutaciones. Ward y Ghetie 1995, Therapeutic Immunology 2:77 y Armor et al. 1999, Eur. J. Immunol. 29: 2613.

[0381] En otra realización, la región constante de inmunoglobulina o una porción de la misma comprende una secuencia de aminoácidos en la región bisagra o una porción de la misma que forma uno o más enlaces disulfuro con una segunda región constante de inmunoglobulina o una porción de la misma. La segunda región constante de inmunoglobulina o una porción de la misma pueden unirse a un segundo polipéptido, uniendo la proteína terapéutica y el segundo polipéptido. En algunas realizaciones, el segundo polipéptido es un radical potenciador. Como se emplea en la presente memoria, el término "radical potenciador" se refiere a una molécula, fragmento de la misma o un componente de un polipéptido que son capaces de potenciar la actividad de la proteína terapéutica. El radical potenciador puede ser un cofactor, por ejemplo, en donde la proteína terapéutica es un factor de coagulación, un factor tisular soluble (sTF) o un péptido procoagulante. Por tanto, después de la activación del factor de coagulación, el radical potenciador es capaz de potenciar la actividad del factor de coagulación.

[0382] En ciertas realizaciones, una proteína terapéutica codificada por una molécula de ácido nucleico de la divulgación comprende una sustitución de aminoácidos por una región constante de inmunoglobulina o una porción de la misma (p. ej., variantes de Fc), que alteran las funciones efectoras independientes del antígeno de la Región constante de Ig, en particular la semivida circulante de la proteína.

2. Regiones scFc

[0383] En otro aspecto, un radical heterólogo comprende una región scFc (Fc de cadena sencilla). En una realización, una molécula de ácido nucleico aislada de la divulgación comprende adicionalmente una secuencia de ácido nucleico heteróloga que codifica una región scFc. La región scFc comprende al menos dos regiones constantes de inmunoglobulina o porciones de las mismas (p. ej., radicales o dominios Fc (p. ej., 2, 3, 4, 5, 6 o más radicales o dominios Fc)) dentro de la misma cadena polipeptídica lineal que son capaces de plegarse (p. ej., plegamiento intramolecular o intermolecular) para formar una región scFc funcional que está conectada mediante un conector peptídico de Fc. Por ejemplo, en una realización, un polipéptido de la divulgación es capaz de unirse, a través de su región scFc, a al menos un receptor de Fc (p. ej., un FcRn, un receptor FcγR (p. ej., FcγRIII) o una proteína del complemento (p. ej., C1q)) para mejorar la semivida o desencadenar una función efectora inmunitaria (p. ej., citotoxicidad dependiente de anticuerpos (ADCC), fagocitosis o citotoxicidad dependiente del complemento (CDCC) y/o mejorar la capacidad de fabricación).

3. CTP

[0384] En otro aspecto, un radical heterólogo comprende un péptido C-terminal (CTP) de la subunidad b de la gonadotropina coriónica humana o un fragmento, variante o derivado de la misma. Se sabe que uno o más péptidos CTP insertados en una proteína recombinante aumentan la semivida *in vivo* de esa proteína. Véase, p. ej., la Patente de Estados Unidos Núm. 5.712.122, incorporada en la presente memoria como referencia en su totalidad.

[0385] Los péptidos CTP ilustrativos incluyen DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPIL (SEQ ID NO: 33) o SSSSKAPPPSLPSPSRLPGPSDTPILPQ (SEQ ID NO: 34). Véase, por ejemplo, la publicación de solicitud de patente de EE.UU. Núm. US 2009/0087411 A1, incorporada como referencia.

4. Secuencia XTEN

[0386] En algunas realizaciones, un radical heterólogo comprende una o más secuencias, fragmentos, variantes o derivados de XTEN de las mismas. Como se emplea aquí, "secuencia XTEN" se refiere a polipéptidos de longitud extendida con secuencias no naturales, sustancialmente no repetitivas que están compuestas principalmente de pequeños aminoácidos hidrófilos, teniendo la secuencia un grado bajo o ninguna estructura secundaria o terciaria en condiciones fisiológicas. Como radical heterólogo, los XTEN pueden servir como radical de prolongación de la semivida. Además, XTEN puede proporcionar propiedades deseables que incluyen, pero no se limitan a, parámetros farmacocinéticos mejorados y características de solubilidad.

[0387] La incorporación de un radical heterólogo que comprende una secuencia XTEN a una proteína de la divulgación puede conferir a la proteína una o más de las siguientes propiedades ventajosas: flexibilidad conformacional, solubilidad en agua mejorada, alto grado de resistencia a proteasas, baja inmunogenicidad, baja unión a receptores de mamíferos, o aumento de los radios hidrodinámicos (o de Stokes).

[0388] En ciertos aspectos, una secuencia XTEN puede aumentar las propiedades farmacocinéticas, tales como una semivida *in vivo* más prolongada o una mayor área bajo la curva (AUC), de modo que una proteína de la divulgación permanece *in vivo* y tiene actividad procoagulante durante un período mayor de tiempo en comparación con una proteína con el mismo, pero sin el radical heterólogo XTEN.

[0389] En algunas realizaciones, la secuencia XTEN útil para la divulgación es un péptido o un polipéptido que tiene más de aproximadamente 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1200, 1400, 1600, 1800 o 2000 restos de aminoácido. En ciertas realizaciones, XTEN es un péptido o polipéptido que tiene más de

aproximadamente 20 a aproximadamente 3000 restos de aminoácido, más de 30 a aproximadamente 2500 residuos, más de 40 a aproximadamente 2000 residuos, más de 50 a aproximadamente 1500 residuos, más de 60 a aproximadamente 1000 residuos, más de 70 a aproximadamente 900 residuos, más de 80 a aproximadamente 800 residuos, más de 90 a aproximadamente 700 residuos, más de 100 a aproximadamente 600 residuos, más de 110 a aproximadamente 500 residuos o más de 120 a aproximadamente 400 residuos.

[0390] La secuencia XTEN de la divulgación puede comprender uno o más motivos de secuencia de 5 a 14 (p. ej., 9 a 14) restos de aminoácido o una secuencia de aminoácidos de al menos 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% idéntico al motivo de secuencia, en donde el motivo comprende, consiste esencialmente en, o consiste en 4 a 6 tipos de aminoácidos (p. ej., 5 aminoácidos) seleccionados del grupo que consiste en glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P). Véase el documento US 2010-0239554 A1.

[0391] En algunas realizaciones, el XTEN comprende motivos de secuencia no solapantes en los que aproximadamente 80%, o al menos aproximadamente 85%, o al menos aproximadamente 90%, o aproximadamente 91%, o aproximadamente 92% o aproximadamente 93%, o aproximadamente 94%, o aproximadamente 95%, o aproximadamente 96%, o aproximadamente 97%, o aproximadamente 98%, o aproximadamente 99% o aproximadamente 100% de la secuencia consiste en múltiples unidades de secuencias que no se solapan seleccionadas a partir de una sola familia de motivos seleccionada de la Tabla 7, que da como resultado una secuencia de familia. Como se emplea en la presente memoria, "familia" significa que el XTEN tiene motivos seleccionados sólo de una única categoría de motivos de la Tabla 7; es decir, XTEN AD, AE, AF, AG, AM, AQ, BC o BD, y que cualquier otro aminoácido en el XTEN que no sea de un motivo de familia se selecciona para lograr una propiedad necesaria, por ejemplo, para permitir la incorporación de un sitio de restricción por los nucleótidos codificantes, incorporación de una secuencia de escisión, o para lograr un mejor enlace con la proteína terapéutica. En algunas realizaciones de familias de XTEN, una secuencia XTEN comprende múltiples unidades de motivos secuenciales no solapantes de la familia de motivos AD, o de la familia de motivos AE, o de la familia de motivos AF, o de la familia de motivos AG, o de la familia de motivos AM, o de la familia de motivos AQ, o de la familia BC, o de la familia BD, exhibiendo el XTEN resultante el intervalo de homología descrito anteriormente. En otras realizaciones, el XTEN comprende múltiples unidades de secuencias de motivos de dos o más de las familias de motivos de la Tabla 7. Estas secuencias se pueden seleccionar para lograr las características físico-químicas deseadas, incluidas propiedades tales como carga neta, carácter hidrófilo, carencia de estructura secundaria, o carencia de repetitividad que son conferidas por la composición de aminoácidos de los motivos, descrito con más detalle a continuación.

En las realizaciones anteriormente descritas en este párrafo, los motivos incorporados al XTEN pueden seleccionarse y ensamblarse utilizando los métodos descritos en la presente memoria para lograr un XTEN de aproximadamente 36 a aproximadamente 3000 restos de aminoácido.

Tabla 7. Motivos de Secuencia XTEN de 12 Aminoácidos y Familias de Motivos

Familia de Motivos*	SECUENCIA MOTIVO	SEQ ID NO:
AD	GESPGGSSGSES	73
AD	GSEGSSGPGESS	74
AD	GSSESGSSEGGP	75
AD	GSGGEPSESGSS	76
AE, AM	GSPAGSPTSTEE	77
AE, AM, AQ	GSEPATSGSETP	78
AE, AM, AQ	GTSESATPESGP	79
AE, AM, AQ	GTSTEPSEGSAP	80
AF, AM	GSTSESPSGTAP	81
AF, AM	GTSTPESGSASP	82
AF, AM	GTSPSGESSTAP	83
AF, AM	GSTSSTAESP GP	84
AG, AM	GTPGSGTASSSP	85
AG, AM	GSSTPSGATGSP	86
AG, AM	GSSPSASTGTGP	87
AG, AM	GASPGTSSTGSP	88
AQ	GEPAGSPTSTSE	89
AQ	GTGEPSTPASE	90
AQ	GSGPSTESAPTE	91
AQ	GSETPSGPSETA	92
AQ	GPSETSTSEPGA	93
AQ	GSPSEPTEGTSA	94
BC	GSGASEPTSTEP	95
BC	GSEPATSGTEPS	96
BC	GTSEPSTSEPGA	97
BC	GTSTEPSEPGSA	98
BD	GSTAGSETSTEA	99
BD	GSETATSGSETA	100

Familia de Motivos*	SECUENCIA MOTIVO	SEQ ID NO:
BD	GTSESATSESGA	101
BD	GTSTEASEGSAS	102

* Denota secuencias de motivos individuales que, cuando se utilizan juntas en varias permutaciones, dan como resultado una "secuencia de familia"

[0392] Se describen ejemplos de secuencias XTEN que se pueden utilizar como radicales heterólogos en las proteínas terapéuticas de la divulgación, por ejemplo, en las publicaciones de patente de EE. UU. Núm. 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1, o 2011/0172146 A1, o en la Publicaciones de Patente Internacional Núm. WO 2010091122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1 o WO 2011028344 A2, cada una de las cuales se incorpora en la presente memoria como referencia en su totalidad.

[0393] XTEN puede tener diferentes longitudes para su inserción o unión a una proteína terapéutica. En una realización, la longitud de la secuencia o secuencias XTEN se elige basándose en la propiedad o función que se vaya a lograr en la proteína de fusión. Dependiendo de la propiedad o función deseada, XTEN puede ser una secuencia de longitud corta o intermedia o una secuencia más larga que pueden servir como portadores. En determinadas realizaciones, el XTEN incluye segmentos cortos de aproximadamente 6 a aproximadamente 99 restos de aminoácidos, longitudes intermedias de aproximadamente 100 a aproximadamente 399 restos de aminoácidos y longitudes más largas de aproximadamente 400 a aproximadamente 1000 y hasta aproximadamente 3000 restos de aminoácidos. Por lo tanto, el XTEN insertado o unido a una proteína terapéutica puede tener longitudes de aproximadamente 6, aproximadamente 12, aproximadamente 36, aproximadamente 40, aproximadamente 42, aproximadamente 72, aproximadamente 96, aproximadamente 144, aproximadamente 288, aproximadamente 400, aproximadamente 500, aproximadamente 576, aproximadamente 600, aproximadamente 700, aproximadamente 800, aproximadamente 864, aproximadamente 900, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500 o hasta aproximadamente 3000 restos de aminoácidos de longitud. En otras realizaciones, las secuencias XTEN tienen de aproximadamente 6 a aproximadamente 50, de aproximadamente 50 a aproximadamente 100, de aproximadamente 100 a 150, de aproximadamente 150 a 250, de aproximadamente 250 a 400, de aproximadamente 400 a aproximadamente 500, de aproximadamente 500 a aproximadamente 900, aproximadamente 900 a 1500, aproximadamente 1500 a 2000, o aproximadamente 2000 a aproximadamente 3000 restos de aminoácidos de longitud. La longitud precisa de un XTEN insertado o

conectado a una proteína terapéutica puede variar sin afectar negativamente a la actividad de la proteína terapéutica. En una realización, uno o más de los XTEN utilizados en la presente memoria tienen 42 aminoácidos, 72 aminoácidos, 144 aminoácidos, 288 aminoácidos, 576 aminoácidos o 864 aminoácidos de longitud y se pueden seleccionar entre uno o más de las secuencias de familias XTEN; es decir, AD, AE, AF, AG, AM, AQ, BC o BD.

[0394] En algunas realizaciones, la proteína terapéutica comprende un polipéptido de FVIII y un XTEN, en donde el XTEN comprende 288 aminoácidos. En una realización, la proteína terapéutica comprende un polipéptido de FVIII y un XTEN, en donde el XTEN comprende 288 aminoácidos y el XTEN se inserta dentro del dominio B del polipéptido de FVIII. En una realización particular, la proteína terapéutica comprende un polipéptido de FVIII y un XTEN que comprende SEQ ID NO: 109, y el XTEN se inserta dentro del dominio B del polipéptido de FVIII. En una realización particular, la proteína terapéutica comprende un polipéptido de FVIII y un XTEN que comprende SEQ ID NO: 109, y el XTEN se inserta dentro del polipéptido de FVIII inmediatamente aguas abajo del aminoácido 745 del FVIII maduro.

[0395] En algunas realizaciones, la proteína terapéutica comprende un polipéptido de FIX y un XTEN, en donde el XTEN comprende 72 aminoácidos. En una realización, la proteína terapéutica comprende un polipéptido de FIX y un XTEN, en donde el XTEN comprende 72 aminoácidos, y el XTEN se inserta dentro del polipéptido de FIX inmediatamente aguas abajo del aminoácido 166 del FIX maduro.

[0396] En algunas realizaciones, la secuencia XTEN utilizada en la divulgación es al menos el 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% idéntica a una secuencia seleccionada del grupo que consiste en AE42, AG42, AE48, AM48, AE72, AG72, AE108, AG108, AE144, AF144, AG144, AE180, AG180, AE216, AG216, AE252, AG252, AE288, AG288, AE324, AG324, AE360, AG360, AE396, AG396, AE432, AG432, AE468, AG468, AE504, AG504, AF504, AE540, AG540, AF540, AD576, AE576, AF576, AG576, AE612, AG612, AE624, AE648, AG648, AG684, AE720, AG720, AE756, AG756, AE792, AG792, AE828, AG828, AD836, AE864, AF864, AG864, AM875, AE912, AM923, AM1318, BC864, BD864, AE948, AE1044, AE1140, AE1236, AE1332, AE1428, AE1524, AE1620, AE1716, AE1812, AE1908, AE2004A, AG948, AG1044, AG1140, AG1236, AG1332, AG1428, AG1524, AG1620, AG1716, AG1812, AG1908, AG2004, y cualquier combinación de las mismas. Véase el documento US 2010-0239554 A1. En una realización particular, el XTEN comprende AE42, AE72, AE144, AE288, AE576, AE864, AG 42, AG72, AG144, AG288, AG576, AG864, o cualquier combinación de las mismas.

[0397] Las secuencias de XTEN ilustrativas que se pueden utilizar como radicales heterólogos en la proteína terapéutica de la divulgación incluyen XTEN AE42-4 (SEQ ID NO: 46, codificado por SEQ ID NO: 47), XTEN AE144-2A (SEQ ID NO: 48, codificado por SEQ ID NO: 49), XTEN AE144-3B (SEQ ID NO: 50, codificado por SEQ ID NO: 51), XTEN AE144-4A (SEQ ID NO: 52, codificado por SEQ ID NO: 53), XTEN AE144-5A (SEQ ID NO: 54, codificado por SEQ ID NO: 55), XTEN AE144-6B (SEQ ID NO: 56, codificado por SEQ ID NO: 57), XTEN AG144-1 (SEQ ID NO: 58, codificado por SEQ ID NO: 59), XTEN AG144-A (SEQ ID NO: 60, codificado por SEQ ID NO: 61), XTEN AG144-B (SEQ ID NO: 62, codificado por SEQ ID NO: 63), XTEN AG144-C (SEQ ID NO: 64, codificado por SEQ ID NO: 65), y XTEN AG144-F (SEQ ID NO: 66, codificado por SEQ ID NO: 67). En una realización particular, el XTEN está codificado por SEQ ID NO: 18.

[0398] En otra realización, la secuencia XTEN se selecciona del grupo que consiste en AE36 (SEQ ID NO: 130), AE42 (SEQ ID NO: 131), AE72 (SEQ ID NO: 132), AE78 (SEQ ID NO: 133), AE144 (SEQ ID NO: 134), AE144_2A (SEQ ID NO: 48), AE144_3B (SEQ ID NO: 50), AE144_4A (SEQ ID NO: 52), AE144_5A (SEQ ID NO: 54), AE144_6B (SEQ ID NO: 135), AG144 (SEQ ID NO: 136), AG144_A (SEQ ID NO: 137), AG144_B (SEQ ID NO: 62), AG144_C (SEQ ID NO: 64), AG144_F (SEQ ID NO: 66), AE288 (SEQ ID NO: 138), AE288_2 (SEQ ID NO: 139), AG288 (SEQ ID NO: 140), AE576 (SEQ ID NO: 141), AG576 (SEQ ID NO: 142), AE864 (SEQ ID NO: 143), AG864 (SEQ ID NO: 144), XTEN_AE72_2A_1 (SEQ ID NO: 145), XTEN_AE72_2A_2 (SEQ ID NO: 146), XTEN_AE72_3B_1 (SEQ ID NO: 147), XTEN_AE72_3B_2 (SEQ ID NO: 148), XTEN_AE72_4A_2 (SEQ ID NO: 149), XTEN_AE72_5A_2 (SEQ ID NO: 150), XTEN_AE72_6B_1 (SEQ ID NO: 151), XTEN_AE72_6B_2 (SEQ ID NO: 152), XTEN_AE72_1A_1 (SEQ ID NO: 153), XTEN_AE72_1A_2 (SEQ ID NO: 154), XTEN_AE144_1A (SEQ ID NO: 155), AE150 (SEQ ID NO: 156), AG150 (SEQ ID NO: 157), AE294 (SEQ ID NO: 158), AG294 (SEQ ID NO: 159), y cualquier combinación de las mismas. En una realización específica, la secuencia XTEN se selecciona del grupo que consiste en AE72, AE144 y AE288. Las secuencias de aminoácidos para ciertas secuencias XTEN de la invención se muestran en la Tabla 8.

Tabla 8. Secuencias XTEN

XTEN	Secuencia de Aminoácidos
AE42-4 (SEQ ID NO: 46)	GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS
AE144-2A (SEQ ID NO: 48)	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG

AG144 (SEQ ID NO:136)	GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTG SPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSAST GTGPGTTPGSGTASSSPGSSTPSGATGSP
AG144_A (SEQ ID NO: 137)	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGASPGTSSTGSPGTTPGSGTASSSPGSSTPSGATGSPGTTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSP
AE288 (SEQ ID NO: 138)	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESA TPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AE288_2 (SEQ ID NO: 139)	GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AG288 (SEQ ID NO: 140)	PGASPGTSSTGSPGASPGTSSTGSPGTTPGSGTASSSPGSSTPSGATGSPGTTPGSGTAS SSPGSSTPSGATGSPGTTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSAS TGTGPGSSPSASTGTGPGASPGTSSTGSPGTTPGSGTASSSPGSSTPSGATGSPGSSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSS PSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTTPGSGTASSSPGSSTPSGATGS
AE576 (SEQ ID NO: 141)	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AG576 (SEQ ID NO: 142)	PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTTPGSGT ASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTTPGS GTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSS TPSGATGSPGASPGTSSTGSPGTTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTTPGSGTAS SSPGSSTPSGATGSPGTTPGSGTASSSPGSSTPSGATGSPGTTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGS
AE864 (SEQ ID NO: 143)	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP

	GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AG864 (SEQ ID NO: 144)	GASPGTSSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTA SSSPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGT SSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSST PSGATGSPGSSTPSGATGSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGT PGSGTASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSPSASTGTGP GTPGSGTASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSTPSGATG SPGSSTPSGATGSPGASPGTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSSTGSPGASPGTSSSTGSPGTPGSG TASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGTPG SGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTG SPGASPGTSSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSSTGSPGSSPSAST GTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSSTGSP
XTEN_AE72_2A_1 (SEQ ID NO: 145)	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPG
XTEN_AE72_2A_2 (SEQ ID NO: 146)	TSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPG
XTEN_AE72_3B_1 (SEQ ID NO: 147)	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSTEPSEGSAPG
XTEN_AE72_3B_2 (SEQ ID NO: 148)	TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPG
XTEN_AE72_4A_2 (SEQ ID NO: 149)	TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPG
XTEN_AE72_5A_2 (SEQ ID NO: 150)	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTE EGSPAGSPTSTEEG
XTEN_AE72_6B_1 (SEQ ID NO: 151)	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSET PGSEPATSGSETPG
XTEN_AE72_6B_2 (SEQ ID NO: 152)	SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPG
XTEN_AE72_1A_1 (SEQ ID NO: 153)	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPG

XTEN_AE72_1A_2 (SEQ ID NO: 154)	TSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPG
XTEN_AE144_1A (SEQ ID NO: 155)	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTS TEEGTSESATPESGPGTSTEPSEGSAPG
AE150 (SEQ ID NO: 156)	GAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSTEPSEGSAPASS
G150 (SEQ ID NO: 157)	GAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTS STGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPS ASTGTGPGTPGSGTASSSPGSSTPSGATGSPASS
AE294 (SEQ ID NO: 158)	GAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTS ESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PASS
AG294 (SEQ ID NO: 159)	GAPPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSG TASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSP SASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATG SASS

[0399] En algunas realizaciones, menos de 100% de los aminoácidos de un XTEN se seleccionan entre glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P), o menos de 100% de la secuencia consiste en los motivos de secuencia de la Tabla 7 o una secuencia XTEN proporcionada en la presente memoria. En tales realizaciones, los restos de aminoácido restantes del XTEN se seleccionan entre cualquiera de los otros 14 L-aminoácidos naturales, pero se pueden seleccionar preferentemente entre aminoácidos hidrófilos de modo que la secuencia XTEN contenga al menos aproximadamente 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o al menos aproximadamente 99% de aminoácidos hidrófilos. El contenido de aminoácidos hidrófobos en el XTEN utilizado en las construcciones de conjugación puede ser menos de 5%, o menos de 2%, o menos de 1% de contenido de aminoácidos hidrófobos. Los restos hidrófobos que son menos favorecidos en la construcción de XTEN incluyen triptófano, fenilalanina, tirosina, leucina, isoleucina, valina y metionina. Además, las secuencias XTEN pueden contener menos de 5% o menos de 4% o menos de 3% o menos de 2% o menos de 1% o ninguno de los siguientes aminoácidos: metionina (p. ej., para evitar la oxidación) o asparragina y glutamina (para evitar la desamidación).

[0400] Las una o más secuencias XTEN pueden insertarse en el extremo C-terminal o en el extremo N-terminal de la proteína terapéutica o insertarse entre dos aminoácidos en la secuencia de aminoácidos de la proteína terapéutica. Por ejemplo, cuando la proteína terapéutica comprende un polipéptido de FVIII, el XTEN puede insertarse entre dos aminoácidos en uno o más sitios de inserción seleccionados de la Tabla 5. Cuando la proteína terapéutica comprende un polipéptido de FIX, el XTEN puede insertarse entre dos aminoácidos en uno o más sitios de inserción seleccionados de la Tabla 5.

[0401] Los ejemplos adicionales de secuencias XTEN que se pueden utilizar de acuerdo con la presente invención y se describen en las publicaciones de patente de EE. UU. Núm. 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1 o 2011/0172146 A1, o las Publicaciones de Patente Internacional Números WO 2010091 122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1, WO 2011028344 A2, WO 2014/011819 A2 o WO 2015/023891.

5. Albúmina o Fragmento, Derivado o Variante de la misma

[0402] En algunas realizaciones, un radical heterólogo comprende albúmina o un fragmento funcional de la misma. La albúmina de suero humano (HSA o HA), una proteína de 609 aminoácidos en su forma de longitud completa, es responsable de una proporción significativa de la presión osmótica del suero y también funciona como portadora de ligandos endógenos y exógenos. El término "albúmina" como se emplea en la presente memoria incluye albúmina de longitud completa o un fragmento funcional, variante, derivado o análogo de la misma. Se describen ejemplos de albúmina o sus fragmentos o variantes en las Publ. de Pat. de EE.UU. Núm. 2008/0194481A1, 2008/0004206 A1, 2008/0161243 A1, 2008/0261877 A1, o 2008/0153751 A1 o en las Publ. de Solicitudes PCT. Núm. 2008/033413 A2, 2009/058322 A1 o 2007/021494 A2, que se incorporan en la presente memoria como referencia en su totalidad.

[0403] En una realización, la proteína terapéutica de la divulgación comprende albúmina, un fragmento o una variante de la misma que se une adicionalmente a un segundo radical heterólogo seleccionado del grupo que consiste en una región constante de inmunoglobulina o una porción de la misma (p. ej., una región Fc), una secuencia PAS, HES, PEG y cualquier combinación de las mismas.

6. Radical de Unión a Albúmina

[0404] En ciertas realizaciones, el radical heterólogo es un radical de unión a albúmina, que comprende un péptido de unión a albúmina, un dominio de unión a albúmina bacteriana, un fragmento de anticuerpo de unión a albúmina o cualquier combinación de los mismos.

[0405] Por ejemplo, la proteína de unión a albúmina puede ser una proteína de unión a albúmina bacteriana, un anticuerpo o un fragmento de anticuerpo que incluye anticuerpos de dominio (véase la Patente de Estados Unidos Núm. 6.696.245). Una proteína de unión a albúmina, por ejemplo, puede ser un dominio de unión a albúmina bacteriana, tal como el de la proteína estreptocócica G (Konig, T. y Skerra, A. (1998) *J. Immunol. Methods* 218, 73-83). Otros ejemplos de péptidos que se unen a la albúmina se que pueden utilizar como compañero de conjugación son, por ejemplo, los que tienen una secuencia consenso Cys-Xaa₁-Xaa₂-Xaa₃-Xaa₄-Cys, en donde Xaa₁ es Asp, Asn, Ser, Thr o Trp; Xaa₂ es Asn, Gln, His, Ile, Leu o Lys; Xaa₃ es Ala, Asp, Phe, Trp o Tyr; y Xaa₄ es Asp, Gly, Leu, Phe, Ser o Thr como describen la Solicitud de Patente de Estados Unidos 2003/0069395 o Dennis et al. (Dennis et al. (2002) *J. Biol. Chem.* 277, 35035-35043).

[0406] El dominio 3 de la proteína G estreptocócica, como describen Kraulis et al., en *FEBS Lett.* 378:190-194 (1996) y Linhult et al., en *Protein Sci.* 11:206-213 (2002) es un ejemplo de un dominio de unión a albúmina bacteriana. Los ejemplos de péptidos que se unen a la albúmina incluyen una serie de péptidos que tiene la secuencia central DICLPRWGCLW (SEQ ID NO: 35). Véase, p. ej., Dennis et al., *J. Biol. Chem.* 2002, 277:35035-35043 (2002). Ejemplos de fragmentos de anticuerpos que se unen a albúmina son descritos por Muller y Kontermann, en *Curr. Opin. Mol. Ther.* 9:319-326 (2007); Roovers et al., en *Cancer Immunol. Immunother.* 56:303-317 (2007) y Holt et al., en *Prot. Ing. Design Sci.*, 21:283-288 (2008), que se incorporan en la presente memoria como referencia en su totalidad. Un ejemplo de tal radical de unión a albúmina es 2-(3-maleimidopropanamido)-6-(4-(4-yodofenil)butanamido)hexanoato (etiqueta "Albu") como describen Trussel et al., *Bioconjugate Chem.* 20:2286-2292 (2009).

[0407] Se pueden utilizar ácidos grasos, en particular ácidos grasos de cadena larga (LCFA) y compuestos de unión a albúmina similares a ácidos grasos de cadena larga para extender la semivida in vivo de las proteínas del factor de coagulación de la divulgación. Un ejemplo de un compuesto de unión a albúmina similar a LCFA es el ácido 16-(1-(3-(9-(((2,5-dioxopirrolidin-1-iloxi)carboniloxi)-metil)-7-sulfo-9H-fluoren-2-ilamino)-3-oxopropil)-2,5-dioxopirrolidin-3-iltio)hexadecanoico (véase, por ejemplo, el documento WO 2010/140148).

7. Secuencia PAS

[0408] En otras realizaciones, el radical heterólogo es una secuencia PAS. Una secuencia PAS, como se emplea en la presente memoria, significa una secuencia de aminoácidos que comprende principalmente restos de alanina y serina o que comprende principalmente restos de alanina, serina y prolina, formando la secuencia de aminoácidos una conformación helicoidal aleatoria en condiciones fisiológicas. Por

consiguiente, la secuencia PAS es un bloque de construcción, un polímero de aminoácidos o un casete de secuencia que comprende, consiste esencialmente en, o consiste en, alanina, serina y prolina que se puede utilizar como parte del radical heterólogo en la proteína quimérica. Sin embargo, el experto en la técnica es consciente de que un polímero de aminoácidos también puede formar una conformación helicoidal aleatoria cuando se añaden residuos distintos de alanina, serina y prolina como constituyente minoritario en la secuencia PAS. El término "constituyente minoritario" como se emplea en la presente memoria significa que se pueden añadir aminoácidos distintos de alanina, serina y prolina en la secuencia PAS hasta un cierto grado, p. ej., hasta aproximadamente 12%, es decir, aproximadamente 12 de 100 aminoácidos de la secuencia PAS, hasta aproximadamente 10%, es decir, aproximadamente 10 de 100 aminoácidos de la secuencia PAS, hasta aproximadamente 9%, es decir, aproximadamente 9 de 100 aminoácidos, hasta aproximadamente 8%, es decir, aproximadamente 8 de 100 aminoácidos, aproximadamente 6%, es decir, aproximadamente 6 de 100 aminoácidos, aproximadamente 5%, es decir, aproximadamente 5 de 100 aminoácidos, aproximadamente 4%, es decir, aproximadamente 4 de 100 aminoácidos, aproximadamente 3%, es decir, aproximadamente 3 de 100 aminoácidos, aproximadamente 2%, es decir, aproximadamente 2 de 100 aminoácidos, aproximadamente 1%, es decir, aproximadamente 1 de 100 de los aminoácidos. Los aminoácidos diferentes de alanina, serina y prolina se pueden seleccionar del grupo que consiste en Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr, y Val.

[0409] En condiciones fisiológicas, el tramo de la secuencia PAS forma una conformación helicoidal aleatoria y, por lo tanto, puede mediar en una mayor estabilidad *in vivo* y/o *in vitro* de la proteína del factor de coagulación. Dado que el dominio helicoidal aleatorio no adopta una estructura o función estable por sí mismo, se conserva esencialmente la actividad biológica mediada por la proteína del factor de coagulación. En otras realizaciones, las secuencias PAS que forman dominios helicoidales aleatorios son biológicamente inertes, especialmente con respecto a la proteólisis en plasma sanguíneo, inmunogenicidad, comportamiento de punto isoelectrico/electrostático, unión a receptores de superficie celular o internalización, pero aún son biodegradables, lo que proporciona claras ventajas sobre polímeros sintéticos tales como PEG.

[0410] Los ejemplos no limitantes de las secuencias PAS que forman una conformación helicoidal aleatoria comprenden una secuencia de aminoácidos seleccionada del grupo que consiste en ASPAAPAPASPAAPAPSAPA (SEQ ID NO: 36), AAPASPAPAAPSAPAPAAPS (SEQ ID NO: 37), APSSPSPSAPSSPSPASPSS (SEQ ID NO: 38), APSSPSPSAPSSPSPASPS (SEQ ID NO: 39), SSPSAPSPSSPASPSPSSPA (SEQ ID NO: 40), AASPAAPSAPPAAASPAAPSAPPA (SEQ ID NO: 41), ASAAAPAAASAAASAPSAAA (SEQ ID NO: 42) y cualquier combinación de las mismas. Se

conocen ejemplos adicionales de secuencias PAS de, por ejemplo, la Publicación de Patente de Estados Unidos Núm. 2010/0292130 A1 y la Publicación de Solicitud PCT Núm. WO 2008/155134 A1.

8. Secuencia HAP

5 [0411] En ciertas realizaciones, el radical heterólogo es un polímero homoaminoácido rico en glicina (HAP). La secuencia HAP puede comprender una secuencia repetitiva de glicina, que tiene al menos 50 aminoácidos, al menos 100 aminoácidos, 120 aminoácidos, 140 aminoácidos, 160 aminoácidos, 180 aminoácidos, 200 aminoácidos, 250 aminoácidos, 300 aminoácidos, 350 aminoácidos, 400 aminoácidos, 450 aminoácidos o 500 aminoácidos de longitud. En una realización, la secuencia HAP es capaz de prolongar la semivida de un radical fusionado o conectado a la secuencia HAP. Los ejemplos no
10 limitantes de la secuencia HAP incluyen, pero no se limitan a $(\text{Gly})_n$, $(\text{Gly}_4\text{Ser})_n$ o $\text{S}(\text{Gly}_4\text{Ser})_n$, en donde n es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20. En una realización, n es 20, 21, 22, 23, 24, 25, 26, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40. En otra realización, n es 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 o 200.

9. Transferrina o Fragmento de la Misma

15 [0412] En ciertas realizaciones, el radical heterólogo es transferrina o un fragmento de la misma. Se puede utilizar cualquier transferrina para producir las proteínas del factor de coagulación de la divulgación. Por ejemplo, el TF (TF) humano de tipo salvaje es una proteína de 679 aminoácidos, de aproximadamente 75 KDa (sin tener en cuenta la glicosilación), con dos dominios principales, N (aproximadamente 330 aminoácidos) y C (aproximadamente 340 aminoácidos), que parecen tener su
20 origen en la duplicación de un gen. Véanse los números de acceso de GenBank NM001063, XM002793, M12530, XM039845, XM 039847 y S95936 (www.ncbi.nlm.nih.gov/), todos los cuales se incorporan en la presente memoria como referencia en su totalidad. La transferrina comprende dos dominios, el dominio N y el dominio C. El dominio N comprende dos subdominios, el dominio N1 y el dominio N2, y el dominio C comprende dos subdominios, el dominio C1 y el dominio C2.

25 [0413] En una realización, el radical heterólogo de transferrina incluye una variante de corte y empalme de transferrina. En un ejemplo, una variante de corte y empalme de transferrina puede ser una variante de corte y empalme de transferrina humana, p. ej., número de Acceso de Genbank AAA61140. En otra realización, la porción de transferrina de la proteína quimérica incluye uno o más dominios de la secuencia de transferrina, p. ej., dominio N, dominio C, dominio N1, dominio N2, dominio C1, dominio
30 C2 o cualquier combinación de los mismos.

10. Receptores de Aclaramiento

[0414] En ciertas realizaciones, el radical heterólogo es un receptor de aclaramiento, un fragmento, una variante o un derivado del mismo. LRP1 es una proteína intrínseca de membrana de 600 kDa que está implicada en el aclaramiento mediado por receptor de una variedad de proteínas, tales como el factor X.

5 Véase, p. ej., Narita et al. Blood 91:555-560 (1998).

11. Factor von Willebrand o Fragmentos del Mismo

[0415] En ciertas realizaciones, el radical heterólogo es el factor von Willebrand (VWF) o uno o más fragmentos del mismo.

10 [0416] El VWF (también conocido como F8VWF) es una gran glicoproteína multimérica presente en el plasma sanguíneo y producida constitutivamente en el endotelio (en los cuerpos de Weibel-Palade), megacariocitos (gránulos α plaquetarios) y tejido conectivo subendotelial. El monómero básico de VWF es una proteína de 2813 aminoácidos. Cada monómero contiene varios dominios específicos con una función específica, los dominios D' y D3 (que juntos se unen al Factor VIII), el dominio A1 (que se une al receptor plaquetario GPIb, heparina y/o posiblemente colágeno), el dominio A3 (que se une al

15 colágeno), el dominio C1 (en el que el dominio RGD se une a la integrina plaquetaria α IIb β 3 cuando está activada) y el dominio "nudo de cisteína" en el extremo C-terminal de la proteína (que comparte VWF con factor de crecimiento derivado de plaquetas (PDGF), el factor de crecimiento transformante β (TGF β) y la gonadotropina coriónica humana β (β HCG)).

20 [0417] La secuencia de aminoácidos del monómero 2813 para VWF humano es referida con el Número de acceso NP000543.2 en Genbank. La secuencia de nucleótidos que codifica el VWF humano es referida con el Número de Acceso NM000552.3 en Genbank. SEQ ID NO: 129 es la secuencia de aminoácidos codificada por SEQ ID NO: 128. El dominio D' incluye los aminoácidos 764 a 866 de SEQ ID NO: 129. El dominio D3 incluye los aminoácidos 867 a 1240 de SEQ ID NO: 44.

25 [0418] En plasma, de 95 a 98% del FVIII circula en un complejo no covalente ajustado con VWF de longitud completa. La formación de este complejo es importante para el mantenimiento de niveles plasmáticos apropiados de F VI 111 in vivo. Lenting et al., Blood. 92 (11): 3983 - 96 (1998); Lenting et al., J. Thromb. Haemost. 5 (7): 1353 - 60 (2007). Cuando el FVIII se activa debido a la proteólisis en las posiciones 372 y 740 de la cadena pesada y en la posición 1689 de la cadena ligera, el VWF unido al FVIII se retira del FVIII activado.

30 [0419] En ciertas realizaciones, el radical heterólogo es Factor de von Willebrand de longitud completa. En otras realizaciones, el radical heterólogo es un fragmento del factor von Willebrand. Como se emplea

en la presente memoria, los términos "fragmento de VWF" o "fragmentos de VWF" utilizados en la presente memoria significan cualquier fragmento de VWF que interactúa con el FVIII y retiene al menos una o más propiedades que normalmente son proporcionadas al FVIII por el VWF de longitud completa, p. ej., previniendo la activación prematura de FVIIIa, previniendo la proteólisis prematura, previniendo la asociación con membranas de fosfolípidos que podrían conducir a un aclaramiento prematuro, previniendo la unión a los receptores de aclaramiento del FVIII que se pueden unir al FVIII desnudo pero no al FVIII unido al FvW, y/o estabilizando las interacciones de la cadena pesada y ligera del FVIII. En una realización específica, el radical heterólogo es un fragmento (VWF) que comprende un dominio D' y un dominio D3 de VWF. El fragmento VWF que comprende el dominio D' y el dominio D3 puede comprender adicionalmente un dominio VWF seleccionado del grupo que consiste de un dominio A1, un dominio A2, un dominio A3, un dominio D1, un dominio D2, un dominio D4, un dominio B1, un dominio B2, un dominio B3, un dominio C1, un dominio C2, un dominio CK, uno o más fragmentos del mismo y cualquier combinación de los mismos. Se describen ejemplos adicionales del polipéptido que tiene actividad de FVIII fusionado con el fragmento de VWF en la Solicitud de Patente Provisional de EE.UU. 61/667.901, presentada el 3 de julio de 2012, y la Publicación de EE. UU. Núm. 2015/0023959 A1, que se incorporan en la presente como referencia en su totalidad.

12. Radicales Conectores

[0420] En ciertas realizaciones, el radical heterólogo es un conector peptídico.

[0421] Como se emplea en la presente memoria, los términos "conectores peptídicos" o "radicales conectores" se refieren a un péptido o secuencia polipeptídica (p. ej., un péptido sintético o secuencia polipeptídica) que conecta dos dominios en una secuencia lineal de aminoácidos de una cadena polipeptídica.

[0422] En algunas realizaciones, se pueden insertar conectores peptídicos entre la proteína terapéutica de la divulgación y un radical heterólogo descrito anteriormente, tal como la albúmina. Los conectores peptídicos pueden proporcionar flexibilidad a la molécula polipeptídica quimérica. Los conectores no se escinden típicamente, sin embargo, tal escisión puede ser deseable. En una realización, estos conectores no se eliminan durante el procesamiento.

[0423] Un tipo de conector que puede estar presente en una proteína quimérica de la divulgación es un conector escindible por proteasa que comprende un sitio de escisión (es decir, un sustrato del sitio de escisión de proteasa, por ejemplo, un factor XIa, Xa o sitio de escisión de trombina) y que pueden incluir conectores adicionales en el extremo N-terminal o C-terminal o en ambos lados del sitio de escisión.

Estos conectores escindibles cuando se incorporan a una construcción de la divulgación dan como resultado una molécula química que tiene un sitio de escisión heterólogo.

[0424] En una realización, una proteína terapéutica codificada por una molécula de ácido nucleico de la presente divulgación comprende dos o más dominios Fc o radicales unidos mediante un conector cscFc para formar una región Fc comprendida en una única cadena polipeptídica. El conector cscFc está flanqueado por al menos un sitio de procesamiento intracelular, es decir, un sitio escindido por una enzima intracelular. La escisión del polipéptido en al menos un sitio de procesamiento intracelular da como resultado un polipéptido que comprende al menos dos cadenas polipeptídicas.

[0425] Se pueden utilizar opcionalmente otros conectores peptídicos en una construcción de la divulgación, por ejemplo, para conectar una proteína del factor de coagulación a una región Fc. Algunos conectores ilustrativos que se pueden utilizar con respecto a la divulgación incluyen, por ejemplo, polipéptidos que comprenden aminoácidos GlySer descritos con más detalle a continuación.

[0426] En una realización, el conector peptídico es sintético, es decir, de origen no natural. En una realización, un conector peptídico incluye péptidos (o polipéptidos) (que pueden ser o no de origen natural) que comprenden una secuencia de aminoácidos que conecta o fusiona genéticamente una primera secuencia lineal de aminoácidos con una segunda secuencia lineal de aminoácidos a la que no está ligada de forma natural o fusionada genéticamente por naturaleza. Por ejemplo, en una realización, el conector peptídico puede comprender polipéptidos no naturales que son formas modificadas de polipéptidos naturales (p. ej., que comprenden una mutación tal como una adición, sustitución o delección). En otra realización, el conector peptídico puede comprender aminoácidos de origen no natural. En otra realización, el péptido conector puede comprender aminoácidos naturales que se encuentran en una secuencia lineal que no se produce en la naturaleza. En otra realización más, el conector peptídico puede comprender una secuencia polipeptídica de origen natural.

[0427] Por ejemplo, en ciertas realizaciones, se puede utilizar un conector peptídico para fusionar radicales Fc idénticos, formando así una región scFc homodimérica. En otras realizaciones, se puede utilizar un conector peptídico para fusionar diferentes restos Fc (p. ej., un radical Fc de tipo salvaje y una variante del radical Fc), formando así una región scFc heterodimérica.

[0428] En otra realización, un conector peptídico comprende o consiste en un conector gly-ser. En una realización, un conector scFc o cscFc comprende al menos una porción de una bisagra de inmunoglobulina y un conector gly-ser. Como se emplea en la presente memoria, el término "conector gly-ser" se refiere a un péptido que consta de restos de glicina y serina. En ciertas realizaciones, dicho conector gly-ser se puede insertar entre otras dos secuencias del conector peptídico. En otras

realizaciones, un conector gly-ser se une a uno o ambos extremos de otra secuencia del conector peptídico. En otras realizaciones más, se incorporan dos o más conectores gly-ser en serie en un conector peptídico. En una realización, un conector peptídico de la divulgación comprende al menos una porción de una región de bisagra superior (p. ej., derivada de una molécula de IgG1, IgG2, IgG3 o IgG4), al menos una porción de una región de bisagra media (p. ej., derivada de una molécula de IgG1, IgG2, IgG3, IgG4) y una serie de residuos de aminoácidos gly/ser.

[0429] Los conectores peptídicos de la divulgación tienen una longitud de al menos un aminoácido y pueden tener longitudes variables. En una realización, un conector peptídico de la divulgación tiene una longitud de aproximadamente 1 a aproximadamente 50 aminoácidos. Como se emplea en este contexto, el término "aproximadamente" indica +/- dos restos de aminoácido. Dado que la longitud del conector debe ser un número entero positivo, la longitud de aproximadamente 1 a aproximadamente 50 aminoácidos de longitud significa una longitud de 1-3 a 48-52 aminoácidos de longitud. En otra realización, un conector peptídico de la divulgación tiene una longitud de aproximadamente 10 a aproximadamente 20 aminoácidos. En otra realización, un conector peptídico de la divulgación tiene una longitud de aproximadamente 15 a aproximadamente 50 aminoácidos. En otra realización, un conector peptídico de la divulgación tiene una longitud de aproximadamente 20 a aproximadamente 45 aminoácidos. En otra forma de realización, un conector peptídico de la divulgación tiene una longitud de aproximadamente 15 a aproximadamente 35 o de aproximadamente 20 a aproximadamente 30 aminoácidos. En otra realización, un conector peptídico de la divulgación tiene aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000 o 2000 aminoácidos de longitud. En una realización, un conector peptídico de la divulgación tiene una longitud de 20 o 30 aminoácidos.

[0430] En algunas realizaciones, el conector peptídico puede comprender al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos 10, al menos 20, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80, al menos 90 o al menos 100 aminoácidos. En otras realizaciones, el conector peptídico puede comprender al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900 o al menos 1000 aminoácidos. En algunas realizaciones, el conector peptídico puede comprender al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 o 2000 aminoácidos. El conector peptídico puede comprender 1-5 aminoácidos, 1-10 aminoácidos, 1-20 aminoácidos, 10-50 aminoácidos, 50-100 aminoácidos, 100-200 aminoácidos, 200-300 aminoácidos, 300-400 aminoácidos, 400-500 aminoácidos, 500-600 aminoácidos, 600-700 aminoácidos, 700-800 aminoácidos, 800-900 aminoácidos, o 900-1000 aminoácidos.

[0431] Se pueden introducir conectores peptídicos en secuencias polipeptídicas utilizando mecanismos conocidas en la técnica. Las modificaciones se pueden confirmarse mediante análisis de secuencia de ADN. Se puede utilizar ADN plasmídico para transformar células anfitrionas para la producción estable de los polipéptidos producidos.

5 **13. Híbridos de Monómero-Dímero**

[0432] En algunas realizaciones, la proteína terapéutica de la divulgación comprende una molécula híbrida de monómero-dímero que comprende un factor de coagulación.

[0433] El término "híbrido de monómero-dímero" utilizado en la presente memoria se refiere a una proteína quimérica que comprende una primera cadena polipeptídica y una segunda cadena polipeptídica, que están asociadas entre sí por un enlace disulfuro, en donde la primera cadena comprende un factor de coagulación, p. ej., FVIII y una primera región Fc y la segunda cadena comprende, consiste esencialmente en, o consiste en una segunda región Fc sin el factor de coagulación. La construcción híbrida de monómero-dímero es por tanto un híbrido que comprende un aspecto de monómero que tiene solo un factor de coagulación y un aspecto de dímero que tiene dos regiones Fc.

15 **14. Secuencias de Control de la Expresión**

[0434] En algunas realizaciones, la molécula de ácido nucleico de la divulgación comprende adicionalmente al menos una secuencia de control de la expresión. Una secuencia de control de la expresión, como se emplea en la presente memoria, es cualquier secuencia reguladora de nucleótidos, tal como una secuencia promotora o una combinación de promotor-potenciador, que facilita la transcripción y traducción eficaces del ácido nucleico codificante al que está conectado operablemente. Por ejemplo, la molécula de ácido nucleico de la divulgación se puede conectar operablemente a al menos una secuencia de control de la transcripción. La secuencia de control de la expresión génica puede ser, por ejemplo, un promotor de mamífero o viral, tal como un promotor constitutivo o inducible.

[0435] Los promotores de mamíferos constitutivos incluyen, pero no se limitan a, los promotores de los siguientes genes: hipoxantina fosforribosil transferasa (HPRT), adenosina desaminasa, piruvato quinasa, promotor de beta-actina y otros promotores constitutivos. Los promotores virales ilustrativos que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores del citomegalovirus (CMV), virus de simio (p. ej., SV40), virus del papiloma, adenovirus, virus de inmunodeficiencia humana (VIH), virus del sarcoma de Rous, citomegalovirus, las repeticiones terminales largas (LTR) del virus de la leucemia de Moloney y otros retrovirus, y el promotor de timidina quinasa del virus del herpes simple. Los expertos en la técnica conocen otros promotores constitutivos. Los promotores útiles como

secuencias de expresión génica de la divulgación también incluyen promotores inducibles. Los promotores inducibles se expresan en presencia de un agente inductor. Por ejemplo, se induce que promotor de la metalotioneína promueva la transcripción y traducción en presencia de ciertos iones metálicos. Los expertos en la técnica conocen otros promotores inducibles.

- 5 **[0436]** En una realización, la divulgación incluye la expresión de un transgén bajo el control de un promotor y/o potenciador específico de tejido. En otra realización, el promotor u otra secuencia de control de la expresión mejoran selectivamente la expresión del transgén en las células hepáticas. En ciertas realizaciones, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en hepatocitos, células sinusoidales y/o células endoteliales. En una realización particular, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en células endoteliales. En ciertas realizaciones, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en las células musculares, el sistema nervioso central, el ojo, el hígado, el corazón o cualquier combinación de los mismos. Los ejemplos de promotores específicos de hígado incluyen, pero no se limitan a, un promotor de tiretina de ratón (mTRR), un
- 10 promotor del factor VIII humano endógeno (F8), el promotor de la alfa-1-antitripsina humana (hAAT), el promotor mínimo de albúmina humana y el promotor de albúmina de ratón. En una realización particular, el promotor comprende un promotor mTTR. El promotor mTTR es descrito por RH Costa et al., 1986, en Mol. Cell. Biol. 6:4697. El promotor F8 es descrito por Figueiredo y Brownlee, 1995, en J. Biol. Chem. 270:11828-11838. En algunas realizaciones, el promotor se selecciona entre un promotor específico de
- 15 hígado (p. ej., α 1-antitripsina (AAT)), un promotor específico del músculo (p. ej., creatina quinasa muscular (MCK), cadena pesada alfa de miosina (α MHC), mioglobina (MB) y desmina (DES)), un promotor sintético (p. ej., SPc5-12, 2R5Sc5-12, dMCK y tMCK) y cualquier combinación de los mismos.
- 20

- [0437]** En una realización, el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor de factor VIII humano endógeno (F8), un promotor de alfa-1-antitripsina humana (hAAT), un promotor mínimo de albúmina humana, el promotor de albúmina de ratón, TTPp, un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV), α 1-antitripsina (AAT), creatina quinasa muscular (MCK), cadena pesada alfa de miosina (α MHC), mioglobina (MB), desmina (DES), SPc5-12, 2R5Sc5-12, dMCK y tMCK, un promotor de fosfoglicerato quinasa (PGK) y cualquier combinación de los mismos.
- 25

- 30 **[0438]** Los niveles de expresión se pueden mejorar adicionalmente para lograr la eficacia terapéutica utilizando uno o más potenciadores. Se pueden proporcionar uno o más potenciadores solos o junto con uno o más elementos promotores. Normalmente, la secuencia de control de la expresión comprende una pluralidad de elementos potenciadores y un promotor específico de tejido. En una realización, un

potenciador comprende una o más copias del potenciador de α -1-microglobulina/bikunina (Rouet et al., 1992, J. Biol. Chem. 267:20765-20773; Rouet et al., 1995, Nucleic Acids Res. 23:395-404; Rouet et al., 1998, Biochem. J. 334:577-584; Ill et al., 1997, Blood Coagulation Fibrinolysis 8:S23-S30). En otra realización, un potenciador se obtiene de sitios de unión de factores de transcripción específicos de hígado, tales como EBP, DBP, HNF1, HNF3, HNF4, HNF6, comprendiendo Enh1 HNF1, (efector)-HNF4, (antisentido)-HNF1, (antisentido)-HNF6, (efector)-EBP, (antisentido)-HNF6 (antisentido).

[0439] En un ejemplo particular, un promotor útil para la divulgación comprende SEQ ID NO: 69 (es decir, promotor ET), que también se conoce como GenBank Núm. AY661265. Véase también Vigna et al., *Molecular Therapy* 11(5):763 (2005). Se describen ejemplos de otros vectores y elementos reguladores de genes adecuados en los documentos WO 02/092134, EP1395293 o las Patentes de Estados Unidos Núm. 6.808.905, 7.745.179 o 7.179.903, que se incorporan en la presente memoria como referencia en su totalidad.

[0440] En una realización, las moléculas de ácido nucleico de la presente divulgación comprenden adicionalmente una secuencia intrónica. En algunas realizaciones, la secuencia intrónica se coloca en 5' con respecto a la secuencia de ácido nucleico que codifica el polipéptido de FVIII. En algunas realizaciones, la secuencia intrónica es una secuencia intrónica de origen natural. En algunas realizaciones, la secuencia intrónica es una secuencia sintética. En algunas realizaciones, la secuencia intrónica se obtiene de una secuencia intrónica de origen natural. En ciertas realizaciones, la secuencia intrónica comprende el intrón T pequeño de SV40. En una realización, la secuencia intrónica comprende SEQ ID NO: 115.

[0441] En algunas realizaciones, la molécula de ácido nucleico comprende adicionalmente un elemento regulador post-transcripcional. En ciertas realizaciones, el elemento regulador postranscripcional comprende un elemento regulador postranscripcional del virus de la hepatitis de la marmota mutado (WPRE). En una realización particular, el elemento regulador postranscripcional comprende SEQ ID NO: 120.

[0442] En algunas realizaciones, la molécula de ácido nucleico comprende un sitio de unión de microARN (miARN). En una realización, el sitio de unión de miARN es un sitio de unión de miARN para miR-142-3p. En otras realizaciones, el sitio de unión de miARN se selecciona entre un sitio de unión de miARN descrito por Rennie et al., *RNA Biol.* 13(6)\554-560 (2016), y STarMirDB, disponible en <http://sfold.wadsworth.org/starmirDB.php>, que se incorporan en la presente memoria como referencia en su totalidad.

[0443] En algunas realizaciones, la molécula de ácido nucleico comprende una o más secuencias de direccionamiento nuclear de ADN (DTS). Una DTS promueve la translocación de moléculas de ADN que contienen tales secuencias al núcleo. En ciertas realizaciones, la DTS comprende una secuencia potenciadora de SV40. En ciertas realizaciones, la DTS comprende una secuencia potenciadora de c-Myc. En algunas realizaciones, las DTS se encuentran entre la primera ITR y la segunda ITR. En algunas realizaciones, la DTS está en posición 3' con respecto a la primera ITR y 5' con respecto a la proteína terapéutica. En otras realizaciones, la DTS está en posición 3' con respecto a la proteína terapéutica y 5' con respecto a la segunda ITR.

[0444] En algunas realizaciones, la molécula de ácido nucleico comprende adicionalmente una secuencia de cola de poli(A) 3'UTR. En una realización, la secuencia de cola de poli(A) 3'UTR comprende poli(A) de bGH. En una realización, la cola de poli(A) 3'UTR comprende un sitio poli(A) de actina. En una realización, la cola de poli(A) 3'UTR comprende un sitio poli(A) de hemoglobina.

[0445] En una realización particular, la secuencia de cola poli(A) 3'UTR comprende SEQ ID NO: 122.

III. Expresión Específica de Tejido

[0446] En ciertas realizaciones, será útil incluir dentro del vector una o más secuencias diana de miARN que, por ejemplo, estén conectadas operablemente al transgén del factor de coagulación. Por tanto, la divulgación también proporciona al menos una diana de secuencia de miARN conectada operablemente a la secuencia de nucleótidos del factor de coagulación o insertada de otro modo dentro de un vector. Más de una copia de una secuencia diana de miARN incluida en el vector puede aumentar la eficacia del sistema. También se incluyen diferentes secuencias diana de miARN. Por ejemplo, los vectores que expresan más de un transgén pueden tener el transgén bajo control de más de una secuencia diana de miARN, que puede ser la misma o diferente. Las secuencias diana de miARN pueden estar en tándem, pero también se incluyen otras configuraciones. El casete de expresión transgénica, que contiene secuencias diana de miARN, también se puede insertar dentro del vector en orientación antisentido. La orientación antisentido puede ser útil en la producción de partículas virales para evitar la expresión de productos génicos que de otro modo pueden ser tóxicos para las células productoras. En otras realizaciones, el vector comprende 1, 2, 3, 4, 5, 6, 7 u 8 copias de la misma secuencia diana de miARN o de una diferente. Sin embargo, en algunas otras realizaciones, el vector no incluirá ninguna secuencia diana de miARN. La elección de incluir o no una secuencia diana de miARN (y cuántas) se guiará por parámetros conocidos tales como la diana de tejido deseada, el nivel de expresión requerido, etc.

[0447] En una realización, la secuencia diana es una diana de miR-223 que se ha informado que bloquea la expresión de manera más eficaz en progenitores comprometidos mieloides y al menos parcialmente en

las HSPC más primitivas. La diana de miR-223 puede bloquear la expresión en células mieloides diferenciadas que incluyen granulocitos, monocitos, macrófagos y células dendríticas mieloides. La diana de miR-223 también puede ser adecuada para aplicaciones de terapia génica que se basan en una expresión transgénica robusta en el linaje linfóide o eritroide. La diana de miR-223 también puede bloquear la expresión de manera muy eficaz en HSC humanas.

[0448] En otra realización, la secuencia diana es una diana de miR142 (tccataaagt aggaacact aca (SEQ ID NO: 43)). En una realización, el vector comprende 4 copias de secuencias diana de miR-142. En ciertas realizaciones, la secuencia complementaria de microARN específicos hematopoyéticos, tales como miR-142 (142T), se incorpora a la región 3' no traducida de un vector, p. ej., vectores lentivirales (LV), haciendo que el transcrito que codifica el transgén sea susceptible a regulación por disminución mediada por miARN. Mediante este método, se puede prevenir la expresión del transgén en células presentadoras de antígeno (APC) de linaje hematopoyético, mientras se mantiene en células no hematopoyéticas (Brown et al., Nat Med 2006). Esta estrategia puede imponer un estricto control postranscripcional sobre la expresión transgénica y, por tanto, permite el suministro estable y la expresión a largo plazo de los transgenes. En algunas realizaciones, la regulación por miR-142 el aclaramiento mediado por el sistema inmunitario de células transducidas y/o induce células T reguladoras específicas de antígenos (T reg) y median una tolerancia inmunológica robusta al antígeno codificado por el transgén.

[0449] En algunas realizaciones, la secuencia diana es una diana de miR181. Chen CZ y Lodish H, Seminars in Immunology (2005) 17(2):155-165 describen miR-181, un miARN expresado específicamente en células B dentro de la médula ósea de ratón (Chen y Lodish, 2005). También divulga que algunos miARN humanos están relacionados con leucemias.

[0450] La secuencia diana puede ser total o parcialmente complementaria al miARN. El término "completamente complementaria" significa que la secuencia diana tiene una secuencia de ácido nucleico que es 100% complementaria a la secuencia del miARN que la reconoce. El término "parcialmente complementaria" significa que la secuencia diana es solo en parte complementaria a la secuencia del miARN que la reconoce, por lo que la secuencia parcialmente complementaria todavía es reconocida por el miARN. En otras palabras, una secuencia diana parcialmente complementaria en el contexto de la presente divulgación es eficaz para reconocer el miARN correspondiente y efectuar la prevención o reducción de la expresión transgénica en células que expresan ese miARN. Los ejemplos de las secuencias diana de miARN se describen en los documentos WO2007/000668, WO2004/094642, WO2010/055413 o WO2010/125471, que se incorporan en la presente memoria como referencia en su totalidad.

[0451] En algunas realizaciones, la expresión del transgén se dirige al hígado. En ciertas realizaciones, la expresión del transgén se dirige a los hepatocitos. En otra realización, la expresión del transgén se dirige a células endoteliales. En una realización particular, la expresión del transgén se dirige a cualquier tejido que exprese naturalmente FVIII endógeno.

5 [0452] En algunas realizaciones, la expresión del transgén se dirige al sistema nervioso central. En ciertas realizaciones, la expresión del transgén se dirige a las neuronas. En algunas realizaciones, la expresión del transgén se dirige a neuronas aferentes. En algunas realizaciones, la expresión del transgén se dirige a neuronas eferentes. En algunas realizaciones, la expresión transgénica se dirige a interneuronas. En algunas realizaciones, la expresión del transgén se dirige a células gliales. En algunas
10 realizaciones, la expresión del transgén se dirige a los astrocitos. En algunas realizaciones, la expresión del transgén se dirige a oligodendrocitos. En algunas realizaciones, la expresión del transgén se dirige a la microglía. En algunas realizaciones, la expresión del transgén se dirige a células ependimarias. En algunas realizaciones, la expresión del transgén se dirige a células de Schwann. En algunas realizaciones, la expresión del transgén se dirige a músculo esquelético.

15 [0453] En algunas realizaciones, la expresión del transgén se dirige al tejido muscular. En algunas realizaciones, la expresión transgénica se dirige al músculo liso. En algunas realizaciones, la expresión transgénica se dirige al músculo cardíaco. En algunas realizaciones, la expresión transgénica se dirige al músculo esquelético.

[0454] En algunas realizaciones, la expresión del transgén se dirige al ojo. En algunas realizaciones, la
20 expresión del transgén se dirige a una célula fotorreceptora. En algunas realizaciones, la expresión del transgén se dirige a células ganglionares de la retina.

IV. Células Anfitrionas

[0455] La divulgación también proporciona una célula anfitriona que comprende una molécula de ácido nucleico o un vector de la divulgación. Como se emplea en la presente memoria, el término
25 "transformación" se utilizará en un sentido amplio para referirse a la introducción de ADN en una célula anfitriona receptora que cambia el genotipo y consecuentemente da como resultado un cambio en la célula receptora.

[0456] "Células anfitrionas" se refiere a células que se han transformado con vectores contruidos utilizando técnicas de ADN recombinante y que codifican al menos un gen heterólogo. Las células
30 anfitrionas de la presente divulgación son preferiblemente de origen mamífero; lo más preferiblemente de origen humano o de ratón. A los expertos en la técnica se les atribuye la capacidad de determinar

preferentemente líneas de células anfitrionas particulares que son las más adecuadas para su propósito. Las líneas de células anfitrionas ilustrativas incluyen, pero no se limitan a, CHO, DG44 y DUXB11 (líneas de ovario de hámster chino, DHFR menos), HELA (carcinoma cervical humano), CVI (línea de riñón de mono), COS (un derivado de CVI con antígeno T de SV40), R1610 (fibroblasto de hámster chino) BALBC/3T3 (fibroblasto de ratón), HAK (línea de riñón de hámster), SP2/0 (mieloma de ratón), P3x63-Ag3.653 (mieloma de ratón), BFA-1c1BPT (células endoteliales bovinas), RAJI (linfocito humano), PER.C6[®], NS0, CAP, BHK21 y HEK 293 (riñón humano). En una realización particular, la célula anfitriona se selecciona del grupo que consiste en: una célula CHO, una célula HEK293, una célula BHK21, una célula PER.C6[®], una célula NS0, una célula CAP y cualquier combinación de las mismas. En algunas realizaciones, las células anfitrionas de la presente divulgación son de origen de insecto. En una realización particular, las células anfitrionas son células SF9. Las líneas de células anfitrionas están disponibles típicamente en servicios comerciales, American Tissue Culture Collection o en la bibliografía publicada.

[0457] La introducción de las moléculas de ácido nucleico o los vectores de la divulgación en la célula anfitriona se puede lograr mediante diversos mecanismos bien conocidos por los expertos en la técnica. Estos incluyen, pero no se limitan a, transfección (incluyendo electroforesis y electroporación), fusión de protoplastos, precipitación con fosfato de calcio, fusión celular con ADN envuelto, microinyección e infección con virus intacto. Véase, Ridgway, AAG "*Mammalian Expression Vectors*" Capítulo 24.2, pág. 470-472 Vectores, Rodríguez y Denhardt, Eds. (Butterworths, Boston, Mass. 1988). Lo más preferiblemente, la introducción del plásmido en el anfitrión se realiza mediante electroporación. Las células transformadas se cultivan en condiciones apropiadas para la producción de las cadenas ligeras y pesadas, y se analizan para la síntesis de proteínas de cadena pesada y/o ligera. Las técnicas de ensayo ilustrativas incluyen ensayo por inmunoabsorción ligado a enzimas (ELISA), radioinmunoensayo (RIA), análisis de clasificación de células activadas por fluorescencia (FACS), inmunohistoquímica y similares.

[0458] Las células anfitrionas que comprenden las moléculas de ácido nucleico aisladas o los vectores de la divulgación se cultivan en un medio de crecimiento apropiado. Como se emplea en la presente memoria, el término "medio de crecimiento apropiado" significa un medio que contiene los nutrientes necesarios para el crecimiento de las células. Los nutrientes necesarios para el crecimiento celular pueden incluir una fuente de carbono, una fuente de nitrógeno, aminoácidos esenciales, vitaminas, minerales y factores de crecimiento. Opcionalmente, los medios pueden contener uno o más factores de selección. Opcionalmente, los medios pueden contener suero bovino de ternero o suero de ternero fetal (FCS). En una realización, el medio no contiene sustancialmente IgG. El medio de crecimiento generalmente seleccionará las células que contienen la construcción de ADN, por ejemplo, selección de fármaco o

deficiencia en un nutriente esencial que se complementa con el marcador seleccionable en la construcción de ADN o se cotransfecta con la construcción de ADN. Las células de mamífero cultivadas generalmente se cultivan en medios que contienen suero o sin suero disponibles comercialmente (p. ej., MEM, DMEM, DMEM/F12). En una realización, el medio es CDOptiCHO (Invitrogen, Carlsbad, CA.).

5 En otra realización, el medio es CD17 (Invitrogen, Carlsbad, CA.). La selección de un medio apropiado para la línea celular particular utilizada está dentro del nivel de los expertos en la técnica.

V. Preparación de polipéptidos

[0459] La divulgación también proporciona un polipéptido codificado por una molécula de ácido nucleico de la divulgación. En otras realizaciones, el polipéptido de la divulgación está codificado por un
10 vector que comprende las moléculas nucleicas de la divulgación. En otras realizaciones más, el polipéptido de la divulgación es producido por una célula anfitriona que comprende las moléculas nucleicas de la divulgación.

[0460] En otras realizaciones, la divulgación también proporciona un método para producir un polipéptido con actividad de factor de coagulación, p. ej., FVIII, que comprende cultivar una célula
15 anfitriona de la divulgación en condiciones en las que se produce un polipéptido con actividad del factor de coagulación, p. ej., FVIII, y se recupera el polipéptido con actividad de factor de coagulación, por ejemplo, FVIII. En algunas realizaciones, la expresión del polipéptido con actividad del factor de coagulación, p. ej., FVIII, aumenta con respecto a una célula anfitriona cultivada en las mismas condiciones pero que comprende una secuencia de nucleótidos de referencia (p. ej., SEQ ID NO: 16, la
20 secuencia del gen de FVIII parental).

[0461] En otras realizaciones, la divulgación proporciona un método para aumentar la expresión de un polipéptido con actividad del factor de coagulación, por ejemplo, FVIII, que comprende cultivar una célula anfitriona de la divulgación en condiciones en las que la actividad de un polipéptido con factor de coagulación, por ejemplo, FVIII, se expresa mediante la molécula de ácido nucleico, en donde la
25 expresión del polipéptido con actividad del factor de coagulación, p. ej., FVIII, aumenta con respecto a una célula anfitriona cultivada en las mismas condiciones que comprende una molécula de ácido nucleico de referencia (p. ej., SEQ ID NO: 16, la secuencia del gen del FVIII parental).

[0462] En otras realizaciones, la divulgación proporciona un método para mejorar el rendimiento de un polipéptido con actividad del factor de coagulación, p. ej., FVIII, que comprende cultivar una célula
30 anfitriona en condiciones en las que la actividad de un polipéptido con factor de coagulación, p. ej., FVIII, es producida por la molécula de ácido nucleico descrita en la presente memoria, en donde el rendimiento de polipéptido con actividad del factor de coagulación, p. ej., FVIII, aumenta con respecto a

una célula anfitriona cultivada en las mismas condiciones que comprende una secuencia de ácido nucleico de referencia (p. ej., SEQ ID NO: 16, la secuencia genética de FVIII parental).

[0463] La proteína terapéutica, por ejemplo, el factor de coagulación, de la divulgación puede sintetizarse en un animal transgénico, tal como un roedor, cabra, oveja, cerdo o vaca. El término "animales transgénicos" se refiere a animales no humanos que han incorporado un gen foráneo en su genoma. Debido a que este gen está presente en los tejidos de la línea germinal, se transmite de padres a hijos. Se introducen genes exógenos en embriones unicelulares (Brinster et al. 1985, Proc. Natl. Acad. Sci. USA 82: 4438). Se conocen en la técnica métodos de producción de animales transgénicos que incluyen transgénicos que producen moléculas de inmunoglobulina (Wagner et al. 1981, Proc. Natl. Acad. Sci. USA 78: 6376; McKnight et al. 1983, Cell 34: 335; Brinster et al. 1983, Nature 306: 332; Ritchie et al. 1984, Nature 312: 517; Baldassarre et al. 2003, Theriogenology 59: 831; Robl et al. 2003, Theriogenology 59: 107; Malassagne et al. 2003, Xenotransplantation 10 (3): 267).

VII. Composición farmacéutica

[0464] Las composiciones que contienen una molécula de ácido nucleico, un polipéptido codificado por la molécula de ácido nucleico, un vector o una célula anfitriona de la presente divulgación pueden contener un portador farmacéuticamente aceptable adecuado. Por ejemplo, pueden contener excipientes y/o agentes auxiliares que faciliten el procesamiento de los compuestos activos en preparaciones diseñadas para ser administradas al sitio de acción.

[0465] En una realización, la presente divulgación está dirigida a una composición farmacéutica que comprende (a) una molécula de ácido nucleico, un vector, un polipéptido o una célula anfitriona descritos en la presente memoria; y (b) un excipiente farmacéuticamente aceptable.

[0466] En algunas realizaciones, la composición farmacéutica comprende adicionalmente un agente de suministro. En ciertas realizaciones, el agente de suministro comprende una nanopartícula lipídica (LNP). En otras realizaciones, la composición farmacéutica comprende adicionalmente liposomas, otras moléculas poliméricas y exosomas.

[0467] Como se emplea en la presente memoria, una "nanopartícula lipídica" se refiere a una nanopartícula que comprende una pluralidad de moléculas lipídicas asociadas físicamente entre sí por fuerzas intermoleculares. Las nanopartículas lipídicas pueden ser, por ejemplo, microesferas (que incluyen vesículas unilaminares y multilaminares, p. ej. liposomas), una fase dispersa en una emulsión, micelas o una fase interna en una suspensión.

[0468] En algunas realizaciones, la presente divulgación proporciona una composición de molécula de ácido nucleico encapsulada que puede incluir un anfitrión de nanopartículas lipídicas que encapsula una molécula de ácido nucleico de la invención. La nanopartícula lipídica puede comprender uno o más lípidos (p. ej., lípidos catiónicos, lípidos no catiónicos y lípidos modificados con PEG). En ciertas realizaciones, las nanopartículas lipídicas de la presente divulgación se formulan para administrar una o más moléculas de ácido nucleico de la invención a una o más células diana. Los ejemplos de lípidos adecuados incluyen, sin limitación, compuestos de fosfatidilo (p. ej., fosfatidiletanolamina, esfingolípidos, fosfatidilcolina, fosfatidilserina, fosfatidilglicerol, gangliósidos y cerebrósidos). Un "lípidio catiónico" se refiere a cualquier especie de lípido que lleve una carga neta positiva a un cierto pH (p. ej., pH fisiológico).

[0469] En ciertas realizaciones, las nanopartículas lipídicas de la presente divulgación tienen una determinada razón N/P. Como se emplea en la presente memoria, "razón N/P" o "razón NP" se refiere a la razón de grupos amina de polímero cargables positivamente respecto a grupos fosfato de ácido nucleico cargados negativamente. El carácter N/P de un complejo de nanopartículas lipídicas/molécula de ácido nucleico puede influir en propiedades tales como la carga superficial neta, la estabilidad y el tamaño. La razón NP de las nanopartículas lipídicas como se describe en la presente memoria puede ser de aproximadamente 1, aproximadamente 2, aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 15, aproximadamente 20, aproximadamente 25, aproximadamente 30, aproximadamente 35, aproximadamente 40, aproximadamente 45, aproximadamente 50, aproximadamente 55, aproximadamente 60, aproximadamente 65, aproximadamente 70, aproximadamente 75, aproximadamente 80, aproximadamente 85, aproximadamente 90, aproximadamente 95, aproximadamente 100, y cualquier razón intermedia. Por ejemplo, la razón NP de las nanopartículas lipídicas como se describe en la presente memoria puede ser de aproximadamente 18, aproximadamente 36 o aproximadamente 72.

[0470] Por consiguiente, en ciertas realizaciones, una composición farmacéutica comprende una molécula de ácido nucleico de la presente divulgación encapsulada en una nanopartícula lipídica y un excipiente farmacéuticamente aceptable.

[0471] La composición farmacéutica se puede formular para administración parenteral (es decir, intravenosa, subcutánea o intramuscular) mediante inyección en bolo. Las formulaciones para inyectables se pueden presentar en forma de dosificación unitaria, p. ej., en ampollas o en recipientes multidosis con un conservante añadido. Las composiciones pueden adoptar formas tales como suspensiones, soluciones o emulsiones en vehículos oleosos o acuosos, y contienen agentes de formulación tales como agentes de

suspensión, estabilizantes y/o dispersantes. Alternativamente, el ingrediente activo puede estar en forma de polvo para reconstituirlo con un vehículo adecuado, p. ej., agua libre de pirógenos.

[0472] Las formulaciones adecuadas para administración parenteral también incluyen soluciones acuosas de los compuestos activos en forma soluble en agua, por ejemplo, sales solubles en agua.

- 5 Además, se pueden administrar suspensiones de los compuestos activos tales como suspensiones inyectables oleosas apropiadas. Los disolventes o vehículos lipófilos adecuados incluyen aceites grasos, por ejemplo, aceite de sésamo, o ésteres de ácidos grasos sintéticos, por ejemplo, oleato de etilo o triglicéridos. Las suspensiones inyectables acuosas pueden contener sustancias que aumentan la viscosidad de la suspensión, incluyendo, por ejemplo, carboximetilcelulosa de sodio, sorbitol y dextrano.
- 10 Opcionalmente, la suspensión también puede contener estabilizantes. Los liposomas también se pueden utilizar para encapsular las moléculas de la divulgación para su administración a células o espacios intersticiales. Los portadores farmacéuticamente aceptables ilustrativos son disolventes fisiológicamente compatibles, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y retardadores de la absorción, agua, solución salina, solución salina tamponada con fosfato,
- 15 dextrosa, glicerol, etanol y similares. En algunas realizaciones, la composición comprende agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol o cloruro de sodio. En otras realizaciones, las composiciones comprenden sustancias farmacéuticamente aceptables tales como agentes humectantes o cantidades minoritarias de sustancias auxiliares tales como agentes humectantes o emulsionantes, conservantes o tampones, que mejoran la vida útil o la eficacia de los ingredientes
- 20 activos.

[0473] Las composiciones de la divulgación pueden estar en una variedad de formas, que incluyen, por ejemplo, líquidos (p. ej., soluciones inyectables e infusibles), dispersiones, suspensiones, formas de dosificación semisólidas y sólidas. La forma preferida depende del modo de administración y la aplicación terapéutica.

- 25 **[0474]** La composición se puede formular como una solución, microemulsión, dispersión, liposoma u otra estructura ordenada adecuada para una alta concentración de fármaco. Se pueden preparar soluciones inyectables estériles incorporando el ingrediente activo en la cantidad requerida en un disolvente apropiado con uno o una combinación de los ingredientes enumerados anteriormente, según se requiera, seguido de esterilización por filtración. Generalmente, las dispersiones se preparan incorporando el
- 30 ingrediente activo en un vehículo estéril que contiene un medio de dispersión alcalino y los otros ingredientes requeridos de los enumerados anteriormente. En el caso de polvos estériles para la preparación de soluciones inyectables estériles, los métodos de preparación preferidos son el secado al vacío y el secado por congelación que produce un polvo del ingrediente activo más cualquier ingrediente

adicional deseado de una solución previamente filtrada en condiciones estériles. La fluidez adecuada de una solución se puede mantener, por ejemplo, mediante el uso de un recubrimiento tal como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersión y mediante el uso de tensioactivos. La absorción prolongada de composiciones inyectables se puede lograr incluyendo en la composición un agente que retrasa la absorción, por ejemplo, sales de monoestearato y gelatina.

[0475] El ingrediente activo se puede formular con una formulación o dispositivo de liberación controlada. Los ejemplos de tales formulaciones y dispositivos incluyen implantes, parches transdérmicos y sistemas de suministro microencapsulados. Se pueden utilizar polímeros biodegradables y biocompatibles, por ejemplo, etileno-acetato de vinilo, polianhídridos, ácido poliglicólico, colágeno, poliortoésteres y ácido poliláctico. Los métodos para la preparación de tales formulaciones y dispositivos son conocidos en la técnica. Véase, por ejemplo, Sustained and Controlled Release Drug Delivery Systems, JR Robinson, ed., Marcel Dekker, Inc., Nueva York, 1978.

[0476] Se pueden preparar formulaciones de depósito inyectables formando matrices microencapsuladas del fármaco en polímeros biodegradables tales como polilactida-poliglicólido. Dependiendo de la razón de fármaco a polímero y de la naturaleza del polímero empleado, se puede controlar la velocidad de liberación del fármaco. Otros polímeros biodegradables ilustrativos son poliortoésteres y polianhídridos. También se pueden preparar formulaciones inyectables de depósito atrapando el fármaco en liposomas o microemulsiones.

[0477] Se pueden incorporar compuestos activos complementarios a las composiciones. En una realización, la molécula de ácido nucleico de la divulgación se formula con un factor de coagulación, o una variante, fragmento, análogo o derivado del mismo. Por ejemplo, el factor de coagulación incluye, pero no se limita a, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, factor XII, factor XIII, protrombina, fibrinógeno, factor von Willebrand o factor tisular soluble recombinante (rsTF) o formas activadas de cualquiera de los anteriores. El factor de coagulación del agente hemostático también puede incluir fármacos antifibrinolíticos, por ejemplo, ácido épsilon-amino-caproico, ácido tranexámico.

[0478] Los regímenes de dosificación se pueden ajustar para proporcionar la respuesta óptima deseada. Por ejemplo, se puede administrar un solo bolo, se pueden administrar varias dosis divididas a lo largo del tiempo, o la dosis se puede reducir o aumentar proporcionalmente según lo indiquen las exigencias de la situación terapéutica. Es ventajoso formular composiciones parenterales en forma de unidad de dosificación para facilitar la administración y uniformidad de la dosificación. Véase, por ejemplo, Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980).

[0479] Además del compuesto activo, la forma de dosificación líquida puede contener ingredientes inertes como agua, alcohol etílico, carbonato de etilo, acetato de etilo, alcohol bencílico, benzoato de bencilo, propilenglicol, 1,3-butilenglicol, dimetilformamida, aceites, glicerol, alcohol tetrahidrofurfurílico, polietilenglicoles y ésteres de ácidos grasos de sorbitán.

5 **[0480]** También se describen ejemplos no limitantes de vehículos farmacéuticos adecuados en Remington's Pharmaceutical Sciences de E. W. Martin. Algunos ejemplos de excipientes incluyen almidón, glucosa, lactosa, sacarosa, gelatina, malta, arroz, harina, tiza, gel de sílice, estearato de sodio, monoestearato de glicerol, talco, cloruro de sodio, leche desnatada en polvo, glicerol, propileno, glicol, agua, etanol, y similares. La composición también puede contener reactivos tamponadores de pH y
10 agentes humectantes o emulsionantes.

[0481] Para la administración oral, la composición farmacéutica puede adoptar la forma de comprimidos o cápsulas preparados por medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (p. ej., jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes
15 (lecitina o goma arábica), vehículos no acuosos (p. ej., aceite de almendras, ésteres oleosos, alcohol etílico o aceites vegetales fraccionados), y conservantes (p. ej., metil- o propil-p-hidroxibenzoatos o ácido sórbico). Las preparaciones también pueden incluir agentes aromatizantes, colorantes y edulcorantes. Alternativamente, la composición puede presentarse como un producto seco para reconstituir con agua u otro vehículo adecuado.

20 **[0482]** Para la administración bucal, la composición puede adoptar la forma de comprimidos o pastillas según protocolos convencionales.

[0483] Para la administración mediante inhalación, los compuestos para su uso de acuerdo con la presente divulgación se administran convenientemente en forma de aerosol nebulizado con o sin excipientes o en forma de aerosol pulverizado de un paquete presurizado o nebulizador, opcionalmente
25 con un propelente, p. ej., diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad medida. Se pueden formular cápsulas y cartuchos de, p. ej., gelatina para uso en un inhalador o insuflador que contengan una mezcla en polvo del compuesto y una base en polvo adecuada tal como lactosa o almidón.

30 **[0484]** La composición farmacéutica también se puede formular para administración rectal como un supositorio o enema de retención, p. ej., que contiene bases de supositorio convencionales tales como manteca de cacao u otros glicéridos.

[0485] En algunas realizaciones, la composición se administra por una ruta seleccionada del grupo que consiste en administración tópica, administración intraocular, administración parenteral, administración intratecal, administración subdural y administración oral. La administración parenteral puede ser intravenosa o subcutánea.

5 VIII. Métodos de tratamiento

[0486] En algunos aspectos, la presente divulgación está dirigida a métodos para tratar una enfermedad o afección en un sujeto que lo necesite, que comprenden administrar una molécula de ácido nucleico, un vector, un polipéptido o una composición farmacéutica descrita en la presente memoria.

- 10 [0487] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético, en donde el casete genético codifica una secuencia diana, en donde la secuencia diana codifica una proteína terapéutica y en donde la molécula de ácido nucleico se utiliza para tratar una enfermedad o afección en un sujeto que lo necesite. En algunas realizaciones, la enfermedad o afección afectan a un órgano seleccionado de músculo, sistema nervioso central (SNC), ocular, hígado, corazón, riñón, páncreas, pulmones, piel, vejiga, tracto urinario y cualquier combinación de los mismos.
- 15 En algunas realizaciones, el sujeto tiene una enfermedad o afección seleccionadas del grupo que consiste en DMD (distrofia muscular de Duchenne), XLMTM (miopatía miotubular ligada al cromosoma X), Parkinson, SMA (atrofia muscular espinal), Ataxia de Friedreich, GUCY2D-LCA (Amaurosis congénita de Leber), XLRS (Retinosquiasis ligada al cromosoma X), AMD (degeneración macular relacionada con la edad), ACHM (acromatopsia), IRD mediada por RPF65, y cualquier combinación de las mismas.
- 20 [0488] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético, en donde el casete genético codifica una secuencia diana, en donde la secuencia diana codifica un miARN, y en donde la molécula de ácido nucleico se utiliza para tratar una enfermedad o afección en un sujeto que lo necesite. En algunas realizaciones, la enfermedad o afección comprenden esclerosis lateral amiotrófica (ELA), enfermedad de Huntington y/o retinosis pigmentaria autosómica dominante.
- 25

- [0489] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete genético, en donde el casete genético codifica una secuencia diana, en donde la secuencia diana codifica un factor de coagulación y en donde la molécula de ácido nucleico se utiliza para tratar una enfermedad o afección de sangrado en un sujeto que lo necesite. La enfermedad o
- 30 afección de sangrado se seleccionan del grupo que consiste en un trastorno de coagulación con sangrado, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, trauma, traumatismo craneoencefálico, sangrado gastrointestinal, hemorragia intracraneal,

hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, sangrado en la vaina del psoas ilíaco y cualquier combinación de los mismos. En otras realizaciones más, el sujeto está programado para someterse a una cirugía. En otras realizaciones más, el tratamiento es profiláctico o bajo demanda.

[0490] La divulgación proporciona un método para tratar un trastorno hemorrágico que comprende administrar a un sujeto que lo necesite una molécula de ácido nucleico, vector o polipéptido de la divulgación. En algunas realizaciones, el trastorno hemorrágico se caracteriza por una deficiencia en un factor de coagulación, por ejemplo, FVIII. En algunas realizaciones, el trastorno hemorrágico es hemofilia. En algunas realizaciones, el trastorno hemorrágico es hemofilia A. En algunas realizaciones del método de tratamiento de un trastorno hemorrágico, la actividad plasmática de un factor de coagulación, por ejemplo, FVIII, a las 24 horas posteriores a la administración aumenta con respecto a un sujeto al que se le administra una molécula de ácido nucleico de referencia (p. ej., SEQ ID NO: 16, la secuencia del gen del FVIII parental), un vector que comprende una molécula de ácido nucleico de referencia o un polipéptido codificado por una molécula de ácido nucleico de referencia.

[0491] La divulgación también se refiere a un método para tratar, mejorar o prevenir un trastorno hemostático en un sujeto que comprende administrar una cantidad terapéuticamente eficaz de una molécula de ácido nucleico aislada de la divulgación o un polipéptido que tiene actividad del factor de coagulación, por ejemplo, FVIII, codificado por la molécula de ácido nucleico de la divulgación. El tratamiento, la mejora y la prevención mediante la molécula de ácido nucleico aislada o el polipéptido codificado pueden ser una terapia de desvío o puente. El sujeto que recibe la terapia de desvío o puente ya puede haber desarrollado un inhibidor de un factor de coagulación, por ejemplo, FVIII, o está sujeto a desarrollar un inhibidor del factor de coagulación.

[0492] Las moléculas de ácido nucleico, vectores o polipéptidos de la divulgación tratan o previenen un trastorno hemostático al promover la formación de un coágulo de fibrina. El polipéptido que tiene actividad de factor de coagulación, por ejemplo, FVIII, codificado por la molécula de ácido nucleico de la divulgación puede activar un miembro de una cascada de coagulación. El factor de coagulación puede participar en la vía extrínseca, la vía intrínseca o ambas.

[0493] Las moléculas de ácido nucleico, vectores o polipéptidos de la divulgación se pueden utilizar para tratar trastornos hemostáticos que se sabe que se pueden tratar con un factor de coagulación. Los trastornos hemostáticos que se pueden tratar mediante los métodos de la divulgación incluyen, entre otros, hemofilia A, hemofilia B, enfermedad de von Willebrand, deficiencia de factor XI (deficiencia de

PTA), deficiencia de factor XII, así como deficiencias o anomalías estructurales en fibrinógeno, protrombina, factor V, factor VII, factor X o factor XIII, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, traumatismo, traumatismo craneoencefálico, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, y sangrado en la vaina del psoas ilíaco.

[0494] En algunas realizaciones, el trastorno hemostático es un trastorno hereditario. En una realización, el sujeto tiene hemofilia A. En otras realizaciones, el trastorno hemostático es el resultado de una deficiencia en un factor de coagulación. En otras realizaciones, el trastorno hemostático es el resultado de una deficiencia de FVIII. En otras realizaciones, el trastorno hemostático puede ser el resultado de un factor de coagulación de FVIII defectuoso.

[0495] En otra realización, el trastorno hemostático puede ser un trastorno adquirido. El trastorno adquirido puede resultar de una enfermedad o afección secundaria subyacente. La afección no relacionada puede ser, como ejemplo, pero no como limitación, cáncer, una enfermedad autoinmunitaria o embarazo. El trastorno adquirido puede resultar de la vejez o de la medicación para tratar un trastorno secundario subyacente (p. ej., quimioterapia contra el cáncer).

[0496] La divulgación también se refiere a métodos para tratar a un sujeto que no tiene un trastorno hemostático o una enfermedad o afección secundaria que dé como resultado la adquisición de un trastorno hemostático. Por tanto, la divulgación se refiere a un método para tratar a un sujeto que necesita un agente hemostático general que comprende administrar una cantidad terapéuticamente eficaz de la molécula de ácido nucleico aislada, vector o polipéptido de la divulgación. Por ejemplo, en una realización, el sujeto que necesita un agente hemostático general se somete a cirugía o está a punto de someterse a ella. La molécula de ácido nucleico aislada, vector o polipéptido de la divulgación se pueden administrar antes o después de la cirugía como agente preventivo. La molécula de ácido nucleico aislada, vector, o polipéptido de la divulgación se pueden administrar durante o después de la cirugía para controlar un episodio de hemorragia aguda. La cirugía puede incluir, pero no se limita a, trasplante de hígado, resección de hígado o trasplante de células madre.

[0497] En otra realización, la molécula de ácido nucleico aislado, vector o polipéptido de la divulgación se pueden utilizar para tratar a un sujeto que tiene un episodio de hemorragia aguda que no tiene un trastorno hemostático. El episodio de hemorragia aguda puede ser el resultado de un traumatismo severo, p. ej., una cirugía, un accidente automovilístico, una herida, un disparo por laceración o cualquier otro acontecimiento traumático que provoque una hemorragia incontrolada.

[0498] La molécula de ácido nucleico aislada, vector o proteína se pueden utilizar para tratar profilácticamente a un sujeto con un trastorno hemostático. La molécula, vector o proteína de ácido nucleico aislado se puede utilizar para tratar un episodio hemorrágico agudo en un sujeto con un trastorno hemostático.

5 **[0499]** En otra realización, la expresión de la proteína del factor de coagulación mediante la administración de la molécula de ácido nucleico aislada o el vector de la divulgación no induce una respuesta inmunitaria en un sujeto. En algunas realizaciones, la respuesta inmunitaria comprende el desarrollo de anticuerpos contra un factor de coagulación. En una realización, la respuesta inmunitaria comprende el desarrollo de anticuerpos contra F VIII. En algunas realizaciones, la respuesta inmunitaria
10 comprende la secreción de citocinas. En algunas realizaciones, la respuesta inmunitaria comprende la activación de células B, células T o tanto células B como células T. En algunas realizaciones, la respuesta inmunitaria es una respuesta inmunitaria inhibidora, en la que la respuesta inmunitaria en el sujeto reduce la actividad de una proteína del factor de coagulación con respecto a la actividad del factor de
15 expresión de una proteína de factor de coagulación mediante la administración de la molécula de ácido nucleico aislada o el vector de la divulgación evita una respuesta inmunitaria inhibidora contra la proteína del factor de coagulación o la proteína del factor de coagulación expresada a partir de la molécula de ácido nucleico aislada o el vector.

[0500] En algunas realizaciones, se administra una molécula de ácido nucleico aislada, un vector o una
20 composición de proteína de la divulgación combinados con al menos otro agente que promueve la hemostasia. Dicho otro agente que favorece la hemostasia es un agente terapéutico con demostrada actividad de coagulación. Como ejemplo, pero no como limitación, el agente hemostático puede incluir FV, FVII, FIX, FX, FXI, FXII, FXIII, protrombina o fibrinógeno o formas activadas de cualquiera de los anteriores. El factor de coagulación o agente hemostático también puede incluir fármacos
25 antifibrinolíticos, por ejemplo, ácido épsilon-amino-caproico, ácido tranexámico.

[0501] En una realización de la divulgación, la composición (p. ej., la molécula de ácido nucleico aislada, vector o polipéptido) es aquella en la que el factor de coagulación está presente en forma activable cuando se administra a un sujeto. Tal molécula activable puede activarse in vivo en el sitio de coagulación después de la administración a un sujeto.

30 **[0502]** En consecuencia, en algunas realizaciones, la presente divulgación proporciona un método para tratar un trastorno hemorrágico en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una

segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica un factor de coagulación, en donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (no AAV). En algunas realizaciones, la presente divulgación proporciona un método para tratar un trastorno hemorrágico en un sujeto que lo necesite, que comprende

5 administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica un factor de coagulación, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187, 188. En algunas realizaciones, la presente divulgación proporciona un método para tratar un

10 trastorno de sangrado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica un factor de coagulación, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos

15 aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0503] Por consiguiente, en algunas realizaciones, la presente divulgación proporciona un método para

20 tratar la hemofilia A en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica el factor VIII, en donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (no AAV). En algunas realizaciones, la presente divulgación proporciona un método para tratar la hemofilia

25 A en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica el factor VIII, en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos

30 aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0504] La divulgación también proporciona un método para tratar un trastorno metabólico del hígado que comprende administrar a un sujeto que lo necesite una molécula de ácido nucleico, vector o polipéptido de la divulgación. En algunas realizaciones, el trastorno metabólico del hígado se selecciona del grupo que consiste en fenilcetonuria (

5 **[0505]** FIG. 7E muestra una inmunotransferencia Western de productos lisados de hígado de ratones con PKU tratados con ADNdh que contiene el transgén de PAH murino y las ITR B19d135 o GPCd165. Los hígados se recogieron el día 81 después del tratamiento y se extrajeron los productos lisados de proteínas. Cada pocillo representa un solo animal. La proteína PAH murina marcada con FLAG se detectó utilizando el anticuerpo anti-FLAG M2 y se incluye un control de carga de GAPDH como comparación.

10 **[0506]**), una enfermedad del ciclo de la urea (p. ej., una deficiencia en transcarbamilasa (OTC) o argininosuccinato sintetasa (ASS)), un trastorno de almacenamiento lisosómico (p. ej., mucopolisacaridosis) y una enfermedad por almacenamiento de glucógeno (p. ej., enfermedad de almacenamiento de glucógeno Tipo I, II, III, IV). Otros trastornos metabólicos del hígado incluyen, sin limitación, enfermedad de Wilson, deficiencia de alfa-1 antitripsina, enfermedad hepática aloinmune
15 gestacional (GALD), defectos de oxidación de ácidos grasos, galactosemia, enfermedades por almacenamiento de lípidos, tirosinemia y trastornos peroxisómicos.

[0507] En algunas realizaciones, la presente divulgación proporciona un método para tratar un trastorno metabólico del hígado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que
20 flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica una enzima metabólica asociada al hígado que es deficiente en el sujeto, en la que la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (no AAV). En algunas realizaciones, la presente divulgación proporciona un método para tratar un trastorno metabólico del hígado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una
25 primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica (p. ej., una proteína requerida para la función metabólica apropiada del hígado), en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188. En algunas realizaciones, la presente divulgación proporciona un método para tratar
30 un trastorno metabólico del hígado en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica una proteína terapéutica (p. ej., una proteína requerida para la función metabólica

apropiada del hígado), en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos expuesta en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0508] En algunas realizaciones, la presente divulgación proporciona un método para tratar una fenilcetonuria (PKU) en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica fenilalanina hidroxilasa (PAH), en la que la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (no AAV). En algunas realizaciones, la presente divulgación proporciona un método para tratar la fenilcetonuria (PKU) en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica fenilalanina hidroxilasa (PAH), en donde la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos establecida en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188. En algunas realizaciones, la presente divulgación proporciona un método para tratar la fenilcetonuria (PKU) en un sujeto que lo necesite, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR) y una segunda ITR que flanquean un casete genético que comprende una secuencia de polinucleótidos heteróloga que codifica fenilalanina hidroxilasa, en la que la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a una secuencia de nucleótidos establecida en SEQ ID NO: 180, 181, 183, 184, 185, 186, 187 o 188, o un derivado funcional de la misma.

[0509] La molécula de ácido nucleico aislada, vector o polipéptido se pueden administrar por vía intravenosa, subcutánea, intramuscular o a través de cualquier superficie mucosa, por ejemplo, por vía oral, sublingual, bucal, sublingual, nasal, rectal, vaginal o por vía pulmonar. La proteína del factor de coagulación se puede implantar o conectar a un soporte sólido de biopolímero que permite la liberación lenta de la proteína quimérica al sitio deseado.

- [0510]** Para la administración oral, la composición farmacéutica puede adoptar la forma de comprimidos o cápsulas preparadas por medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (p. ej., jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes (lecitina o goma arábica), vehículos no acuosos (p. ej., aceite de almendras, ésteres oleosos, alcohol etílico o aceites vegetales fraccionados) y conservantes. (p. ej., metil o propil-p-hidroxibenzoatos o ácido sórbico). Las preparaciones también pueden incluir agentes aromatizantes, colorantes y edulcorantes. Alternativamente, la composición puede presentarse como un producto seco para reconstituir con agua u otro vehículo adecuado.
- 5
- [0511]** Para la administración bucal y sublingual, la composición puede adoptar la forma de comprimidos, pastillas o películas de disolución rápida según protocolos convencionales.
- 10
- [0512]** Para la administración mediante inhalación, el polipéptido que tiene actividad de factor de coagulación para su uso de acuerdo con la presente divulgación se administra convenientemente en forma de aerosol desde un paquete presurizado o nebulizador (p. ej., en PBS), con un propelente adecuado, p. ej., diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad medida. Se pueden formular cápsulas y cartuchos, p. ej., de gelatina para su uso en un inhalador o insuflador que contenga una mezcla en polvo del compuesto y una base en polvo adecuada tal como lactosa o almidón.
- 15
- [0513]** En una realización, la vía de administración de la molécula de ácido nucleico aislada, vector o polipéptido es parenteral. El término parenteral, como se emplea en la presente memoria, incluye administración intravenosa, intraarterial, intraperitoneal, intramuscular, subcutánea, rectal o vaginal. Se prefiere la forma intravenosa de administración parenteral. Aunque todas estas formas de administración se contemplan claramente dentro del alcance de la divulgación, una forma de administración sería una solución para inyección, en particular para inyección o goteo intravenoso o intraarterial. Por lo general, una composición farmacéutica adecuada para inyección puede comprender un tampón (p. ej., tampón de acetato, fosfato o citrato), un tensioactivo (p. ej., polisorbato), opcionalmente un agente estabilizador (p. ej., albúmina humana), etc. Sin embargo, en otros métodos compatibles con las enseñanzas de la presente memoria, la molécula de ácido nucleico aislada, vector, o polipéptido se pueden suministrar directamente al sitio de la población celular adversa, aumentando de ese modo la exposición del tejido enfermo al agente terapéutico.
- 20
- 25
- 30

[0514] Las preparaciones para administración parenteral incluyen soluciones, suspensiones y emulsiones acuosas o no acuosas estériles. Los ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales tales como el aceite de oliva y ésteres orgánicos inyectables tales como el oleato de etilo. Los portadores acuosos incluyen agua, soluciones, emulsiones o suspensiones
5 alcohólicas/acuosas, incluidos medios salinos y tamponados. En la divulgación objeto, los portadores farmacéuticamente aceptables incluyen, pero no se limitan a tampón fosfato 0,01-0,1 M, y preferiblemente 0,05 M o solución salina al 0,8%. Otros portadores parenterales comunes incluyen soluciones de fosfato de sodio, dextrosa de Ringer, dextrosa y cloruro de sodio, lactato de Ringer o aceites fijados. Los portadores intravenosos incluyen reponedores de líquidos y nutrientes, reponedores
10 de electrolitos, tales como los basados en dextrosa de Ringer y similares. También pueden estar presentes conservantes y otros aditivos tales como, por ejemplo, antimicrobianos, antioxidantes, agentes quelantes y gases inertes y similares.

[0515] Más concretamente, las composiciones farmacéuticas adecuadas para uso inyectable incluyen soluciones acuosas estériles (cuando sean solubles en agua) o dispersiones y polvos estériles para la
15 preparación extemporánea de soluciones o dispersiones inyectables estériles. En tales casos, la composición debe ser estéril y debe ser fluida en la medida en que exista una fácil jeringabilidad. Debería ser estable en las condiciones de fabricación y almacenamiento y preferiblemente se conservará frente a la acción contaminante de microorganismos, tales como bacterias y hongos. El vehículo puede ser un disolvente o medio de dispersión que contenga, por ejemplo, agua, etanol, poliol (p. ej., glicerol,
20 propilenglicol y polietilenglicol líquido y similares) y mezclas adecuadas de los mismos. La fluidez adecuada se puede mantener, por ejemplo, mediante el uso de un recubrimiento tal como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersión y mediante el uso de tensioactivos.

[0516] La composición farmacéutica también se puede formular para administración rectal como un
25 supositorio o enema de retención, p. ej., que contiene bases de supositorio convencionales tales como manteca de cacao u otros glicéridos.

[0517] Las dosis eficaces de las composiciones de la presente divulgación para el tratamiento de afecciones varían dependiendo de muchos factores diferentes, incluidos los medios de administración, el sitio diana, el estado fisiológico del paciente, si el paciente es un ser humano o un animal, otros
30 medicamentos administrado, y si el tratamiento es profiláctico o terapéutico. Normalmente, el paciente es un ser humano, pero también se pueden tratar mamíferos no humanos, incluidos mamíferos transgénicos. Las dosis de tratamiento pueden titularse utilizando métodos de rutina conocidos por los expertos en la técnica para optimizar la seguridad y la eficacia.

[0518] La molécula de ácido nucleico, el vector o los polipéptidos de la divulgación se pueden administrar opcionalmente combinados con otros agentes que son eficaces para tratar el trastorno o afección que necesita tratamiento (p. ej., profiláctico o terapéutico).

[0519] Como se emplea en la presente memoria, la administración de moléculas de ácido nucleico

5 aisladas, vectores o polipéptidos de la divulgación juntos o combinados con una terapia coadyuvante significa la administración o aplicación secuencial, simultánea, coextensiva, concurrente, concomitante o contemporánea de la terapia y los polipéptidos descritos. Los expertos en la técnica apreciarán que la administración o aplicación de los diversos componentes del régimen terapéutico combinado se puede programar para mejorar la eficacia global del tratamiento. Un experto en la técnica (p. ej., un médico)

10 sería capaz de discernir fácilmente los regímenes terapéuticos combinados eficaces sin una experimentación indebida basada en la terapia complementaria seleccionada y las enseñanzas de la presente memoria descriptiva.

[0520] Se apreciará además que la molécula de ácido nucleico aislada, vector o polipéptido de la

presente divulgación se pueden utilizar juntos o combinados con un agente o varios agentes (p. ej., para

15 proporcionar un régimen terapéutico combinado). Los agentes ilustrativos con los que se puede combinar un polipéptido o polinucleótido de la divulgación incluyen agentes que representan el patrón actual de atención para un trastorno particular que se está tratando. Tales agentes pueden ser de naturaleza química o biológica. El término "biológico" o "agente biológico" se refiere a cualquier agente farmacéuticamente activo elaborado a partir de organismos vivos y/o sus productos que está destinado a ser utilizado como

20 agente terapéutico.

[0521] La cantidad de agente que se va a utilizar combinado con los polinucleótidos o polipéptidos de la presente divulgación puede variar según el sujeto o se puede administrar de acuerdo con lo que se conoce en la técnica. Véase, p. ej., Bruce A Chabner et al., *Antineoplastic Agents*, en GOODMAN & GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287 ((Joel G. Hardman et al., Eds., 9ª ed.

25 1996). En otra realización, se administra una cantidad de dicho agente compatible con el patrón de atención.

[0522] En una realización, también se describe en la presente memoria un kit que comprende la

molécula de ácido nucleico descrita en la presente memoria e instrucciones para administrar la molécula de ácido nucleico a un sujeto que lo necesite. En otra realización, se describe en la presente memoria un

30 sistema de baculovirus para la producción de la molécula de ácido nucleico proporcionada en la presente memoria. La molécula de ácido nucleico se produce en células de insectos. En otra realización, se

proporciona un sistema de suministro de nanopartículas para construcciones de expresión. La construcción de expresión comprende la molécula de ácido nucleico descrita en la presente memoria.

IX. Terapia Génica

5 **[0523]** Ciertos aspectos de la presente divulgación proporcionan un método para expresar una construcción genética en un sujeto, que comprende administrar la molécula de ácido nucleico aislada de la divulgación a un sujeto que lo necesite. En algunos aspectos, la divulgación proporciona un método para aumentar la expresión de un polipéptido en un sujeto que comprende administrar la molécula de ácido nucleico aislada de la divulgación a un sujeto que lo necesite. En otros aspectos, la divulgación proporciona un método para modular la expresión de un polipéptido en un sujeto que lo necesite, que
10 comprende administrar una molécula de ácido nucleico aislada de la divulgación, por ejemplo, una secuencia de ácido nucleico que comprende un miARN, al sujeto. En algunos aspectos, la divulgación proporciona un método para regular por disminución la expresión de un gen diana en un sujeto que lo necesite que comprende administrar una molécula de ácido nucleico aislada de la descripción, p. ej., una secuencia de ácido nucleico que comprende un miARN, al sujeto.

15 **[0524]** La terapia génica somática se ha explorado como un posible tratamiento para una variedad de afecciones, que incluyen, entre otras, la hemofilia A. La terapia génica es un tratamiento particularmente atractivo para la hemofilia debido a su potencial para curar la enfermedad a través de la producción endógena continua de un factor de coagulación, por ejemplo, FVIII, después de una única administración de vector. La hemofilia A es muy adecuada para un enfoque de reemplazo de genes porque sus
20 manifestaciones clínicas son totalmente atribuibles a la falta de un producto genético único (p. ej., FVIII) que circula en cantidades ínfimas (200 ng/ml) en el plasma.

[0525] Se ha demostrado que el uso de la administración de genes convencional basada en virus induce una respuesta inmunitaria en seres humanos. Las proteínas de la cápside viral pueden desencadenar varios componentes del sistema inmunológico humano. El suministro de genes basado en AAV ha sido
25 atractivo ya que el AAV es un virus común en la población humana, la mayoría de las personas han estado expuestas a AAV y se ha demostrado que el AAV es menos inmunogénico que, por ejemplo, el adenovirus. En consecuencia, la mayoría de las personas ya han desarrollado una respuesta inmunitaria contra las variantes particulares a las que habían estado expuestas anteriormente. Esta respuesta adaptativa preexistente puede incluir NAb y células T que podrían disminuir la eficacia clínica de
30 reinfecciones posteriores con AAV y/o la eliminación de células que han sido transducidas, lo que descalifica a los pacientes con inmunidad anti-AAV preexistente para tratamiento basado en terapia génica con AAV. Las moléculas de ácido nucleico de la presente divulgación encuentran uso en la

terapia génica de base no viral. Como las cápsides virales no son necesarias para el suministro de genes utilizando las moléculas de ácido nucleico de la presente divulgación, no se desarrollará inmunidad a los componentes virales salvo la posterior readministración (o nueva dosificación) de un sujeto. Como tales, las moléculas de ácido nucleico de la presente divulgación permiten la re-dosificación para estrategias de administración de genes a largo plazo.

[0526] Además, como se describe en la presente memoria, las moléculas de ácido nucleico de la presente divulgación comprenden ITR parvovirales que no son AAV que flanquean un casete genético para impulsar la expresión transgénica estable tras la administración. La presencia de las ITR es necesaria para una expresión transgénica estable, como se muestra en la FIG. 5, donde los ácidos nucleicos sin ITR fueron incapaces de efectuar una expresión transgénica estable (véanse, "ADNdh sin ITR" y "minicírculo").

[0527] Se puede producir una proteína del factor de coagulación de la divulgación *in vivo* en un mamífero, por ejemplo, un paciente humano, utilizando un enfoque de terapia génica para el tratamiento de una enfermedad o trastorno hemorrágico seleccionado del grupo que consiste en un trastorno de coagulación con sangrado, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, trauma, traumatismo craneoencefálico, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, y sangrado en la vaina del psoas ilíaco sería terapéuticamente beneficioso. En una realización, la enfermedad o trastorno de sangrado es la hemofilia. En otra realización, la enfermedad o trastorno de sangrado es la hemofilia A.

[0528] Otras afecciones también son adecuadas para el tratamiento utilizando las moléculas de ácido nucleico descritas en la presente memoria. En ciertas realizaciones, los métodos descritos en la presente memoria se utilizan para tratar una enfermedad o afección que afecta a un órgano diana seleccionado entre músculo, sistema nervioso central (SNC), ocular, hígado, corazón, riñón, páncreas, pulmones, piel, vejiga, tracto urinario, o cualquier combinación de los mismos. En ciertas realizaciones, los métodos descritos en la presente memoria se utilizan para tratar una enfermedad o afección seleccionada entre DMD (distrofia muscular de Duchenne), XLMTM (miopatía miotubular ligada al cromosoma X), Parkinson, SMA (atrofia muscular espinal), Ataxia de Friedreich, GUCY2D-LCA (Amaurosis Congénita de Leber), XLRS (Retinosquiosis Ligada a X), AMD (Degeneración Macular Relacionada con la edad), ACHM (Acromatopsia), IRD mediada por RPF65 (Tabla 9).

Tabla 9. Enfermedades y trastornos tratables mediante los métodos descritos en la presente memoria.

Enfermedad	Órgano diana	Gen defectuoso	Terapia génica
DMD (Distrofia muscular de Duchenne)	Músculo	Distrofina ligada a X	Introducción de genes
XLMTM (miopatía miotubular ligada a X)	Músculo	<i>MTM1</i> (miotubularina)	Introducción de genes
Parkinson	SNC	Tirosina hidroxilasa, AADC, ciclohidrolasa	Introducción de genes
SMA (atrofia muscular espinal)	SNC	SMN1	Introducción de genes
Ataxia de Friedreich	SNC	FXN (Frataxina)	Introducción de genes
<i>GUCY2D</i> -LCA Amaurosis Congénita de Leber	Ocular	GUCY2D	Introducción de genes
XLRS Retinosquiasis Ligada a X	Ocular	RS1	Introducción de genes
AMD Degeneración Macular Relacionada con la edad	Ocular	CFH HTRA ARMS CFB/CC2	Introducción de genes
ACHM Acromatopsia	Ocular	CNGA/CNGB	Introducción de genes
IRD mediada por RPF65	Ocular	Prf65	Introducción de genes
Trastornos de almacenamiento lisosómico	Órgano diana	Gen defectuoso	Terapia génica
MLD leucodistrofia metacromática	SNC	ARSA PSAP	Introducción de genes

(Trastorno de almacenamiento lisosómico)			
MPS Mucopolisacaridosis (Trastorno de almacenamiento lisosómico)	Hígado	IDUA (MPS I) IDS (MPS II)	Introducción de genes
PKU Fenilcetonuria (Trastorno de almacenamiento lisosómico)	Hígado	PAH	Introducción de genes
Pompe Enfermedad de almacenamiento de glucógeno tipo II	Corazón, hígado, músculo, SNC	GAA (alfa-glucosidasa ácida)	Introducción de genes
Terapias con Micro ARN	Órgano diana	Gene defectuoso	Terapia génica
ELA Esclerosis lateral amiotrófica	SNC	SOD1 ¹	miARN
Enfermedad de Huntington	SNC	HTT ²	miARN
AdRP Retinosis Pigmentaria Dominante Autosómica	Ocular	RHO ³ (Rodopsina)	miARN

1: La mutación del gen SOD1 representó 20% de los casos de ELA hereditaria. SOD1 de tipo salvaje ha demostrado propiedades antiapoptóticas en cultivos neurales, mientras que se ha observado que SOD1 mutante promueve la apoptosis en las mitocondrias de la médula espinal, pero no en las mitocondrias del hígado, aunque se expresa igualmente en ambas. La regulación por disminución de la expresión de SOD1 mutado podría inhibir la degeneración de la neurona motora en la ELA. 2: La EH es uno de varios trastornos por repetición de trinucleótidos que son causados por la longitud de una sección repetida de un gen que excede un rango normal. *HTT* contiene una secuencia de tres bases de ADN, citosina-adenina-guanina (CAG), repetidas varias veces (es decir ... CAGCAGCAG ...), conocida como repetición de trinucleótidos. CAG es el código genético (codón) de 3 letras para el aminoácido glutamina, por lo que una serie de ellos da como resultado la producción de una cadena de glutamina conocida como tramo de poliglutamina (o tramo polyQ), y la parte repetida del gen, la *región PoliQ*. Generalmente, las personas

tienen menos de 36 glutaminas repetidas en la región polyQ, lo que da como resultado la producción de la proteína citoplasmática Huntingtina. Sin embargo, una secuencia de 36 o más glutaminas da como resultado la producción de una proteína que tiene características diferentes. Esta forma alterada, denominada huntingtina mutante (mHTT), aumenta la tasa de deterioro de ciertos tipos de neuronas.

- 5 Generalmente, el número de repeticiones de CAG está relacionado con cuánto se ve afectado este proceso y representa aproximadamente 60% de la variación de la edad de inicio de los síntomas. La variación restante se atribuye al entorno y a otros genes que modifican el mecanismo de la EH. De 36 a 39 repeticiones dan como resultado una forma de penetrancia reducida de la enfermedad, con un inicio mucho más tardío y una progresión más lenta de los síntomas. En algunos casos, el inicio puede ser tan
- 10 tardío que los síntomas nunca se notan. Con recuentos de repeticiones muy grandes, la EH tiene penetrancia total y puede aparecer antes de los 20 años, cuando se la conoce como EH juvenil, EH rígida acinética o EH variante de Westphal. Esto representa aproximadamente 7% de los portadores de HD. 3: La mayoría de las mutaciones del gen RHO responsables de la retinosis pigmentaria alteran el
- 15 plegamiento o transporte de la proteína rodopsina. Algunas mutaciones hacen que la rodopsina se active constitutivamente en lugar de activarse en respuesta a la luz. Los estudios sugieren que las versiones alteradas de la rodopsina interfieren en las funciones celulares esenciales, provocando que los bastones se autodestruyan (sufran apoptosis). Debido a que los bastones son esenciales para la visión en condiciones de poca luz, la pérdida de estas células conduce a ceguera nocturna progresiva en personas con retinosis pigmentaria.
- 20 **[0529]** En algunas realizaciones, los métodos descritos en la presente memoria se utilizan para tratar un trastorno de almacenamiento lisosómico. En algunas realizaciones, el trastorno de almacenamiento lisosómico se selecciona entre MLD (leucodistrofia metacromática), MPS (mucopolisacaridosis), PKU (fenilcetonuria), enfermedad de almacenamiento de glucógeno tipo II de Pompe o cualquier combinación de las mismas.
- 25 **[0530]** En algunas realizaciones, los métodos descritos en la presente memoria se utilizan en una terapia con microARN (miARN). En algunas realizaciones, el miARN trata una afección causada por expresión en exceso de un gen o una proteína. En algunas realizaciones, el miARN trata una afección causada por la acumulación de una proteína. En algunas realizaciones, el miARN trata una afección causada por la expresión errónea de un gen o proteína. En algunas realizaciones, el miARN trata una afección causada
- 30 por la expresión de un gen mutante. En algunas realizaciones, el miARN trata una afección causada por la expresión de un gen heterólogo. En ciertas realizaciones, la terapia con miARN trata una afección seleccionada entre ELA (esclerosis lateral amiotrófica), enfermedad de Huntington, AdRP (retinosis pigmentaria autosómica dominante) y cualquier combinación de las mismas. En ciertas realizaciones, los métodos de la presente divulgación comprenden el tratamiento dirigido a la ELA mediante la
- 35 administración de una molécula de ácido nucleico descrita en la presente memoria, en donde la molécula de ácido nucleico comprende un casete genético que codifica un miARN, donde el miARN se dirige a la expresión de SOD1. En ciertas realizaciones, el miARN comprende el miARN artificial miR SOD1 descrito por Dirren et al., *Annals of Clinical and Translational Neurology* 2(2): 167-84 (febrero de 2015).

La mutación del gen SOD1 representa 20% de los casos de ELA heredados. SOD1 de tipo salvaje ha demostrado propiedades antiapoptóticas en cultivos neurales, mientras que se ha observado que SOD1 mutante promueve la apoptosis en las mitocondrias de la médula espinal, pero no en las mitocondrias del hígado, aunque se expresa igualmente en ambas. La regulación por disminución de la expresión de SOD1 mutada podría inhibir la degeneración de la neurona motora en la ELA.

[0531] En ciertas realizaciones, los métodos de la presente divulgación comprenden el tratamiento dirigido a la enfermedad de Huntington mediante la administración de una molécula de ácido nucleico descrita en la presente memoria, en donde la molécula de ácido nucleico comprende un casete genético que codifica un miARN, en donde el miARN se dirige a la expresión de HTT. En ciertas realizaciones, el miARN comprende el miARN modificado por ingeniería genética de miHTT descrito por Evers et al., *Molecular Therapy* 26(9):1-15 (epub antes de la impresión de junio de 2018). La enfermedad de Huntington es uno de varios trastornos de repetición de trinucleótidos que son causados por la longitud de una sección repetida de un gen que excede un rango normal. HTT contiene una secuencia de tres bases de ADN, citosina-adenina-guanina (CAG), repetida varias veces (es decir... CAGCAGCAG...), que se conoce como repetición de trinucleótidos. CAG es el código genético de 3 letras (codón) para el aminoácido glutamina, por lo que una serie de estas repeticiones da como resultado la producción de una cadena de glutamina conocida como tramo de poliglutamina (o tramo poliQ), y la parte repetida del gen, la región de PoliQ. Generalmente, las personas tienen menos de 36 glutaminas repetidas en la región poliQ, lo que da como resultado la producción de la proteína citoplasmática huntingtina. Sin embargo, una secuencia de 36 o más glutaminas da como resultado la producción de una proteína que tiene características diferentes. Esta forma alterada, llamada huntingtina mutante (mHTT), aumenta la tasa de deterioro de ciertos tipos de neuronas. Generalmente, el número de repeticiones de CAG está relacionado con cuánto se ve afectado este proceso y representa aproximadamente 60% de la variación de la edad de inicio de los síntomas. La variación restante se atribuye al medio ambiente y a otros genes que modifican el mecanismo de la enfermedad de Huntington. De 36 a 39 repeticiones dan como resultado una forma de penetrancia reducida de la enfermedad, con un inicio mucho más tardío y una progresión más lenta de los síntomas. En algunos casos, el inicio puede ser tan tardío que los síntomas nunca se notan. Con recuentos de repeticiones muy grandes, la enfermedad de Huntington tiene penetrancia total y puede ocurrir antes de los 20 años, cuando se la conoce como enfermedad de Huntington juvenil, rígida acinética o enfermedad de Huntington variante de Westphal. Esto representa aproximadamente 7% de los portadores de la enfermedad de Huntington.

[0532] En ciertas realizaciones, los métodos de la presente divulgación comprenden el tratamiento dirigido a la Retinosis Pigmentaria Autosómica Dominante (AdRP) mediante la administración de una

molécula de ácido nucleico descrita en la presente memoria, en donde la molécula de ácido nucleico comprende un casete genético que codifica un miARN, en donde el miARN se dirige a la expresión de RHO (rodopsina). En ciertas realizaciones, el miARN comprende miR-708 (véase Behrman et al., *JCB* 192(6):919-27 (2011)). La mayoría de las mutaciones del gen RHO responsables de la retinosis

5 pigmentaria alteran el plegamiento o transporte de la proteína rodopsina. Algunas mutaciones hacen que la rodopsina se active constitutivamente en lugar de activarse en respuesta a la luz. Los estudios sugieren que las versiones alteradas de la rodopsina interfieren en funciones celulares esenciales, lo que hace que los bastones se autodestruyan (sufran apoptosis). Debido a que los bastones son esenciales para la visión en condiciones de poca luz, la pérdida de estas células conduce a ceguera nocturna progresiva en

10 personas con retinosis pigmentaria.

[0533] Todos los diversos aspectos, realizaciones y opciones descritos en la presente memoria se pueden combinar en todas y cada una de las variaciones.

[0534] Todas las publicaciones, patentes y solicitudes de patente mencionadas en esta memoria descriptiva se incorporan en la presente memoria como referencia en la misma medida que se

15 incorporaría como referencia cada publicación, patente o solicitud de patente individual indicada específica e individualmente.

[0535] Habiendo descrito en general esta divulgación, se puede obtener una comprensión adicional haciendo referencia a los ejemplos proporcionados en la presente memoria. Estos ejemplos tienen únicamente fines ilustrativos y no se pretende que sean limitantes.

20 EJEMPLOS

Ejemplo 1. Generación de construcciones de expresión de FVIII que portan ITR parvovirales de AAV y no AAV.

Ejemplo 1a. Clonación del gen FVIII con codones optimizados y regiones de repetición terminal invertida (ITR) de AAV en casetes genéticos.

25 [0536] El casete genético de FVIII se generó basándose en el genoma del serotipo 2 de AAV. Sin embargo, las regiones ITR originadas a partir de cualquier serotipo (incluido el sintético) se pueden utilizar en este enfoque (Fig. 1A).

[0537] Se diseñó un plásmido de expresión AAV2-FVIIIco6XTEN que codificaba una secuencia codificante de FVIII con codones optimizados bajo la regulación de un promotor específico de hígado (TTPp) o un promotor ubicuo (CAGp, FIG. 1A y 1B) flanqueado por regiones de repetición terminal

30 invertida (ITR) de AAV (AAV-FVIII) para expresión *in vitro* e *in vivo*, como se muestra en la FIG. 1C.

El casete genético también contiene elementos WPRE y bGHpA para la expresión óptima del transgén (FIG.s 1A-1C). La secuencia de FVIII con codones optimizados flanqueada por ITR se clonó en un esqueleto de plásmido que comprendía un origen de replicación ColE1 y un casete de expresión para beta-lactamasa, que confiere resistencia a ampicilina (FIG. 1C). Se diseñaron mediante ingeniería sitios de reconocimiento para la endonucleasa de restricción PvuII que flanqueaban el casete de expresión para permitir la escisión precisa de la construcción de AAV-FVIII después de la digestión con PvuII (FIG. 1C).

Ejemplo 1b. Clonación del gen FVIII con codones optimizados y regiones de repetición terminal invertida (ITR) de parvovirus no AAV en casetes genéticos.

[0538] Basándose en la relación filogenética entre los miembros de la familia viral *Parvoviridae*, a la que pertenece el AAV (FIG. 2A), se planteó la hipótesis de que otros miembros del género Dependovirus no AAV y los miembros del género Erythrovirus utilizan mecanismos celulares similares para el mantenimiento del ciclo de vida viral y el establecimiento de infección latente persistente. Por lo tanto, las regiones ITR originadas a partir de los genomas de estos virus se podrían utilizar para desarrollar casetes de expresión genética de tipo AAV (pero no basados en AAV). Los siguientes parvovirus se probaron para determinar la idoneidad de sus regiones ITR para el desarrollo de construcciones genéticas para aplicaciones de terapia génica: dependovirus parvovirus del ganso (GPV), cepa B y eritrovirus parvovirus B19 (Fig. 2A).

[0539] La inestabilidad de las regiones ITR parvovirales durante la propagación de vectores plasmídicos en células bacterianas presenta un desafío para la generación y manipulación de construcciones genéticas. Se han generado con éxito algunas construcciones genéticas que contienen las ITR de AAV2 de longitud completa (145 nt), pero estas construcciones son muy inestables y la mayoría de los plásmidos basados en ITR de AAV2 contienen una versión truncada de 130 nt de la región ITR (ilustrada en la Tabla 1). De manera similar, las construcciones de plásmidos que portaban secuencias completas de ITR de B19 y GPV que se generaron exhibieron un alto grado de inestabilidad en el anfitrión bacteriano (datos no mostrados), lo que limita significativamente la utilidad de estas ITR para el desarrollo de vectores genéticos para aplicaciones de terapia génica.

[0540] Anteriormente, se desarrolló un sistema de genética inversa para el rescate del virus B19 recombinante que porta una versión truncada de la ITR (Manaresi, et al. *Virology* 508 (2017): 54-62) (Tabla 2B, ITR ID: B19d135). Por tanto, se utilizó la ITR de B19d135 para generar un plásmido de expresión de FVIII genéticamente estable B19-FVIIIco6XTEN (FIG. 1D). Para utilizar adicionalmente este enfoque para la síntesis de la construcción basada en la ITR de GPV, se compararon las secuencias

de tipo salvaje de longitud completa de las ITR de B19, GPV y AAV2 (FIG. 3A). Basándose en la homología con los primeros 135 y 15 nucleótidos de las secuencias de ITR de B19 y AW2, respectivamente, que son prescindibles para la función de la ITR, se planteó la hipótesis de que los primeros 162 nucleótidos de la ITR de GPV se podrían eliminar para sintetizar construcciones genéticas estables con ITR completamente funcionales (FIG. 3A, secuencias en recuadro). Por lo tanto, de manera similar a las construcciones AAV2-FVIIIco6XTEN y B19-FVIIIco6XTEN que portan versiones truncadas de sus ITR correspondientes, se utilizó GPVd162 (Tabla 2C) para generar una construcción de plásmido de expresión de FVIII estable GPV-FVIIIco6XTEN (FIG. 1E). Notablemente, las ITR tanto de B19 como de GPV de longitud completa son mucho más largas que las ITR de AAV2 de longitud completa (Tabla 1) y no forman la estructura en horquilla distintiva en forma de T de las ITR de AAV (FIG. 2B).

[0541] Los plásmidos que contenían secuencias de ITR de B19 de longitud completa exhibieron un alto grado de inestabilidad en las células anfitrionas bacterianas, ya que podría generarse una construcción de expresión de FVIIIco6XTEN que contenía solo la ITR3'. Utilizando técnicas de clonación molecular convencionales, no se pudieron obtener clones positivos que contuvieran las ITR de B19 de longitud completa tanto 5' como 3'. Para generar una construcción de expresión de FVIIIco6XTEN, B19wt-FVIIIco6XTEN, flanqueada por las ITR de B19 de longitud completa (FIG. 1F), se utilizó la cepa anfitriona específica de E. coli PMC103. PMC103 contiene una delección en el gen *sbcC*, que codifica una exonucleasa que reconoce y elimina las estructuras cruciformes del ADN. Sin estar limitados por la teoría, se pensó que el uso de la cepa PMC103, que carece de *sbcC*, puede permitir la replicación de palíndromos largos (es decir, secuencias que contienen una estructura secundaria compleja) y la clonación satisfactoria de B19wt-FVIIIco6XTEN así como de GPVwt-FVIIIco6XTEN. El plásmido resultante codifica la secuencia de ITR 5' y 3' de tipo salvaje de B19 de 383 pares de bases (Tabla 2D) y el otro plásmido codifica la secuencia de ITR 5' y 3' de GPV de tipo salvaje de 444 pares de bases (Tabla 2F).

[0542] Los plásmidos B19-FVIIIco6XTEN (FIG. 1D; Tabla 2B), GPV-FVIIIco6XTEN (FIG. 1E; Tabla 2C) y B19wt-FVIIIco6XTEN (FIG. 1F, Tabla 2D) que contenían casetes de expresión de FVIII flanqueados por ITR parvovirales no AAV (B19d135, GPVd162 y B19wt) se generaron como se describe en el Ejemplo 1a. Se utilizaron sitios de reconocimiento para la endonucleasa de restricción Lglul para flanquear todos los casetes de expresión de FVIII (FIG.s 1D-1F).

Ejemplo 1c. Preparación de fragmentos de ADN de hebra sencilla que contienen casetes de expresión de FVIII flanqueados con ITR parvovirales de AAV y no AAV.

[0543] Se planteó la hipótesis de que la formación de estructuras en horquilla dentro de las regiones ITR que flanquean el casete de expresión de FVIII impulsaría la transducción persistente de las células diana.

- 5 Para los estudios de prueba de concepto, el plásmido basado en ITR de AAV AAV2-FVIIIco6XTEN y los plásmidos basados en ITR no AAV B19-FVIIIco6XTEN y GPV-FVIIIco6XTEN se digirieron con PvuII y Lgl, respectivamente. Se generaron los fragmentos AAV-FVIII, B19-FVIII o GPV-FVIII de hebra sencilla (hs) con estructuras ITR en horquilla formadas desnaturalizando los productos del fragmento de ADN bicatenario (casete de expresión de FVIII y esqueleto del plásmido) de la digestión de
- 10 PvuII o Lgl a 95°C y a continuación enfriando a 4°C para permitir que las secuencias de ITR palindrómicas se plegaran (FIG. 1A-1B). Los AAVhs-FVIII, B19hs-FVIII resultante, o GPVhs-FVIII resultantes se probaron en el modelo de ratón HemA (hemofilia A) para determinar la capacidad para establecer la transducción persistente de hepatocitos.

Ejemplo 1d. Uso de un sistema de expresión de baculovirus para generar construcciones de expresión de FVIII

- 15 [0544] Se utilizará un sistema de expresión de baculovirus descrito por Li et al., PLoS ONE 8(8): e69879 (2013) para la producción de construcciones de AAV-FVIII, B19-FVIII y GPV-FVIII en una forma de moléculas de ADN de extremo cerrado (ADNec) en células de insecto. Se ha demostrado que la administración sistémica de casetes de expresión de ADNec establece la transducción persistente de
- 20 hepatocitos e impulsa la expresión transgénica estable a largo plazo en el hígado.

Ejemplo 2. La inyección sistémica de construcciones genéticas que comprenden casetes de expresión de FVIII flanqueados por ITR parvovirales de AAV y no AAV da como resultado la expresión de FVIII a largo plazo en ratones HemA.

Ejemplo 2a. Evaluación in vivo de la expresión de FVIII mediada por AAVhs-FVIII.

- 25 [0545] Para validar la capacidad de las regiones ITR de AAV que portan AAVhs-FVIII para mediar la expresión transgénica persistente in vivo, el casete de expresión genética se administró sistémicamente mediante inyección hidrodinámica (HDI) a ratones con hemofilia A (HemA) de 5-12 semanas de edad (4 animales/grupo) a 5 µg, 10 µg, 20 µg de casete de expresión genética de ADNhs (AAVhs-FVIII) (FIG. 4A). La HDI da como resultado la administración primaria del material inyectado en el hígado de
- 30 animales de experimentación. Se recolectaron muestras de plasma de animales experimentales a las 18 horas, 3 días, 2 semanas, 3 semanas, 1 mes, 2 meses, 3 meses y 4 meses después de una única inyección

hidrodinámica de AAVhs-FVIII. La actividad plasmática de FVIII en sangre se analizó mediante un ensayo de actividad cromogénica de FVIII. Los animales de control a los que se habían inyectado 5 µg/ratón del plásmido de expresión parental mostraron niveles elevados de actividad plasmática de FVIII poco después de la administración. Sin embargo, el nivel de FVIII circulante disminuyó rápidamente y se volvió indetectable 15 días después de la inyección (p.i.). En contraste, los animales experimentales a los que se habían inyectado 5, 10 y 20 µg/ratón de AAVhs-FVIII desarrollaron una expresión a largo plazo del transgén con niveles estables de FVIII circulante aproximadamente 8, 16 y 32% del nivel normal de FVIII, respectivamente (FIG. 4A). Se observó una fuerte respuesta a la dosis, lo que sugiere un alto grado de correlación entre la dosis inyectada y el resultado del tratamiento.

Ejemplo 2b. Evaluación in vivo de la expresión de FVIII mediada por B19hs-FVIII y GPVhs-FVIII.

[0546] Para evaluar la expresión in vivo de FVIII de B19hs-FVIII y GPVhs-FVIII que portan las regiones ITR parvovirales no AAV B19d135 y GPVd162, respectivamente, se suministraron sistemicamente 10 o 20 µg/ratón de B19dh-FVIII y 10 o 50 µg/ratón del casete de expresión genética GPVdh-FVIII a través de HDI en ratones con hemofilia A (HemA) de 5-12 semanas de edad. Se recogieron muestras de sangre a los 1, 3, 7, 14, 21, 28, 42, 56, 84, 112, 140 y 168 días p.i. y se analizó la actividad del FVIII en sangre mediante el ensayo de actividad cromogénica de FVIII. Como se observó con la construcción de AAV-FVIII, los animales de control inyectados con 5 µg/ratón del plásmido de expresión de FVIII parental mostraron niveles altos de actividad plasmática de FVIII a las 24 horas p.i. que disminuyeron rápidamente y se volvieron indetectables a los 14 días p.i. Los animales experimentales a los que se había inyectado B19hs-FVIII mostró una actividad plasmática máxima de FVIII a los 3 días p.i. que a continuación disminuyó gradualmente durante el período de 21 días y se estabilizó en torno a 28 días p.i. (FIG. 4B) Los ratones HemA a los que se había inyectado GPVhs-FVIII, por otro lado, desarrollaron niveles estables de actividad plasmática de FVIII en torno al día 112, que se mantuvieron durante el período de observación restante (Fig. 4C). En particular, los animales a los que se habían inyectado 10 µg/ratón de AAVhs-FVIII (FIG. 4A) o GPVhs-FVIII (FIG. 4C) desarrollaron niveles estables muy similares de actividad plasmática de FVIII, lo que sugiere que las regiones ITR de AAV2 y GPV comprenden factores genéticos requeridos para el establecimiento eficaz de la transducción persistente de células diana.

Ejemplo 2c. Evaluación in vivo del requisito de ITR y horquilla para la expresión estable a largo plazo de FVIII en ratones hemA.

[0547] Para comparar la estabilidad y la expresión a largo plazo de casetes de ADN de hebra sencilla con terapias alternativas de ácido nucleico, la construcción del plásmido FVIIIco6XTEN (FIG. 1A) se

digirió con PvuII o AflII para crear ADN lineal de doble hebra con o sin las secuencias de ITR de AAV. El ADN de doble hebra lineal sin las ITR se purificó para generar la construcción 'ADNdh No ITR'. Finalmente, la ligación del ADNdh purificado sin ITR a través de sitios de reconocimiento de AflII solapantes dio como resultado la formación de un minicírculo de ADN. Esta pequeña construcción de ADN circular, similar a un plásmido, carece de cualquier secuencia bacteriana y/o secuencia de ITR. A los ratones HemA se les inyectaron concentraciones equimolares de la construcción de ADN mediante inyección hidrodinámica y se determinaron los niveles de actividad de FVIII a partir de recolecciones de plasma durante 2-4 meses. Todas las construcciones de ADN generaron niveles terapéuticos iniciales de FVIII en el rango normal de 30-60%, sin embargo, solo el ADN monocatenario demostró una persistencia estable de la expresión del transgén al 32% durante 4 meses después de la inyección (FIG. 5). Todo el ADN de doble hebra y el ADN de minicírculo alcanzaron niveles estables de expresión a 6-10% normal los días 14-42, sin embargo, estas mesetas representan solo 10% de la actividad inicial de FVIII observada. Debido a que los niveles elevados y transitorios de expresión de FVIII pueden dar como resultado la formación de anticuerpos anti-fármaco neutralizantes, se requiere una expresión estable para la tolerancia inmunitaria en un entorno de hemofilia A.

Ejemplo 2d. Comparación in vivo de las ITR de B19 de tipo salvaje y derivadas.

[0548] Para comparar el efecto de una ITR derivada de B19 (B19d135, FIG. 1D, Tabla 2C) con la ITR de B19 de longitud completa (Tabla 2C), se generó el casete de expresión de FVIIIco6XTEN flanqueado por las ITR de 248 pares de bases (FIG. 1F). Se inyectaron hidrodinámicamente a ratones con hemofilia A 30 µg de ADN de FVIII de hebra sencilla flanqueado por B19d135 (FIG. 1D), GPVd165 (FIG. 1E) o la ITR de B19 de tipo salvaje (FIG. 1F). Se recogió plasma a los 3, 7, 14, 21, 28 y 35 días de la inyección para todas las cohortes, y se tomaron muestras adicionales los días 42, 55 y 84 para las construcciones de B19d135 y GPVd165 y se analizó la actividad del FVIII mediante un ensayo cromogénico (FIG. 6). En comparación con la ITR de B19 derivado, la ITR de longitud completa dio como resultado un aumento de aproximadamente 2,5 veces en la expresión de FVIII. Además, la expresión de FVIII a partir de la ITR de tipo salvaje fue estable al inicio.

Ejemplo 2e. Evaluación de la re-administración de ADN desnudo de hebra sencilla in vivo.

[0549] Una limitación crítica en las modalidades actuales de terapia génica es la incapacidad de volver a administrar la terapia debido a la formación de anticuerpos anti-fármaco contra la cápside viral del vector de terapia génica. Sin embargo, los sistemas de terapia génica ausentes en las proteínas inmunogénicas podrían volver a dosificarse para valorar al paciente a un nivel terapéutico deseado. Para evaluar si los casetes de hebra sencilla flanqueados por ITR no AAV de los autores de la presente invención se podían

volver a administrar, se inyectaron a ratones hemA 30 µg de ADNhs que contenían las ITR de B19d135 y GPVd165 los días 0 y 35 (FIG. 6). Los ratones a los que se administró GPVd165-FVIII alcanzaron niveles estables de FVIII de aproximadamente 5% de lo normal durante el primer mes de observación. Después de una segunda dosis de ADNhs, los niveles de FVIII aumentaron a 10% antes de disminuir ligeramente, lo que demuestra un aumento de 2 veces en los niveles de FVIII. Los ratones a los que se administró B19d135-FVIII alcanzaron niveles estables de FVIII de 8% durante la primera semana, que aumentaron aproximadamente 3,5 veces a 30% antes de disminuir a 25%. Estos datos demuestran que la re-administración de ADN monocatenario con las ITR no AAV puede aumentar los niveles de expresión estable de FVIII en ratones con hemofilia A.

10 **Ejemplo 3. Generación y evaluación in vivo de construcciones de expresión de FVIII que portan derivados de ITR parvovirales no AAV de B19d135 y GPVd162.**

Ejemplo 3a. Determinación de secuencias esenciales mínimas de ITR B19 y GPV.

[0550] Basándose en la comparación entre las secuencias de ITR de los dependovirus AAV2 y GPV, y el eritrovirus B19 (números de acceso de Gene Bank NC_001401.2, U25749.1 y KY940273.1, respectivamente), se diseñaron las secuencias mínimas de las ITR de parvovirus de GPV y B19 que serían necesarias con o sin secuencias adicionales (espaciadores, inserciones, inversiones, adiciones y/o recombinación con secuencias de tipo salvaje de otras ITR parvovirales) para la transducción persistente de células eucarióticas con construcciones genéticas que portan dichas ITR (FIG.s 3A y 3B). El alineamiento de secuencia de las ITR de AAV2, GPV y B19 reveló las regiones conservadas B19v1 y GPVv1 entre las tres especies virales (presentadas en las Tablas 2A-2C) como secuencias continuas sin regiones espaciadoras de secuencia variable. Asimismo, se diseñaron las variantes de secuencia esencial mínima B19v3 y GPVv3 basándose en la comparación de secuencias entre las ITR de B19 y GPV. Dado que las construcciones de expresión de FVIII que portan la ITR de GPVd162 se comportaron mejor en experimentos in vivo que las construcciones genéticas que portan la ITR de B19d135, se planteó la hipótesis de que la secuencia de B19v3 comprende regiones de secuencia de ITR de B19 mínimas que se conservan entre las secuencias de ITR de B19 y GPV, y la secuencia de GPVv3 comprende regiones de secuencias de ITR de GPV mínimas que están presentes en la secuencia de ITR de GPV y carecen de la secuencia de ITR de B19 (Tablas 2B y 2C). Se generaron las secuencias de B19v2 y GPVv2 excluyendo los primeros 135 y 162 nucleótidos y los 135 y 162 nucleótidos complementarios correspondientes en las regiones del palíndromo de ITR de las secuencias de ITR de B19 y GPV, respectivamente (Tablas 2B y 2C).

Ejemplo 3b. Orientación de las regiones palindrómicas de las ITR de B19 y GPV y sus derivados en construcciones genéticas funcionales.

[0551] Parte de las ITR parvovirales consiste en una región palindrómica autocomplementaria. Se ha demostrado previamente para los parvovirus B19 infecciosos recombinantes que los virus rescatados que
5 portan regiones palindrómicas en orientaciones directas e inversas exhiben propiedades de crecimiento similares (Manaresi, et al. Virology 508 (2017): 54-62). Por lo tanto, se propone que las construcciones de expresión genética que portan las ITR de B19 y GPV y sus derivados sigan siendo funcionales independientemente de si las regiones palindrómicas de tales ITR están en combinación directa, inversa o cualquier combinación posible de combinación de ITR 5' y 3' con respecto al casete de expresión
10 genética. Para validar esta hipótesis, las ITR de B19d135 y GPVd162, así como las ITR de B19 y GPV de tipo salvaje se incorporarán en el casete de expresión de FVIIIco6XTEN en orientaciones directa, inversa, e invertida utilizando secuencias complementarias idénticas, así como inversas para las ITR de la misma especie. El ADN de hebra sencilla de estos plásmidos se generará y probará en ratones con hemofilia A para determinar la expresión de FVIII dirigida al hígado impulsada por el promotor TTPp
15 como se describe en los Ejemplos 2a, 2b y 2d. Además de investigar todas las orientaciones de las ITR de la misma especie, también se generarán combinaciones de ITR de tipo salvaje de GPV y B19 y sus derivados y se probarán para determinar la expresión de FVIII en ratones con hemofilia A. Estos casetes de expresión contendrán una ITR de origen B19 y una ITR de origen GPV para determinar si las secuencias de ITR no homólogas pueden potenciar la concatemerización episómica y la expresión a largo
20 plazo del transgén deseado. A los ratones con hemofilia A se les inyectarán a través de inyección hidrodinámica 10, 20 o 50 µg de ADNhs que contiene los casetes de expresión antes mencionados y el FVIII a partir del plasma murino recogido a intervalos semanales después de la inyección. El efecto sobre la expresión y longevidad de FVIII en ratones a los que se les administraron estos casetes de expresión se comparará directamente con la expresión y longevidad de FVIII en ratones a los que se les administró
25 B19d135, GPVd162 y los casetes de expresión de ITR de tipo salvaje correspondientes (Tablas 2B, 2C, 2D y 2F).

Ejemplo 3c. Inyección sistémica de construcciones genéticas que portan derivados de ITR parvovirales no AAV de B19d135 y GPVd162 en ratones HemA.

[0552] Para evaluar la expresión de FVIII in vivo a partir de construcciones de ADNhs que portan
30 derivados de ITR parvovirales no AAV de B19d135 y GPVd162, se administrarán 5, 10, 20 o 50 µg/ratón de cada casete de expresión genética de ADNhs sistémicamente a través de HDI en ratones HemA de 5-12 semanas de edad. Las muestras de sangre se recolectarán a los 1, 3, 7, 14, 21 y 28 días

p.i., y a continuación una vez al mes durante un período de 4 meses. La actividad del FVIII en sangre se analizará mediante el ensayo de actividad cromogénica del FVIII.

Ejemplo 4. Producción y evaluación in vivo de construcciones de expresión de ADNec que portan derivados de ITR parvovirales no AAV de B19d135 y GPVd162 en células de insecto

5 *Ejemplo 4a. Uso de un sistema de expresión de baculovirus para generar construcciones de expresión de ADNec que portan derivados de ITR de B19d135 y GPVd162*

[0553] De manera similar a las construcciones AAV-FVIII, B19-FVIII y GPV-FVIII descritas en el Ejemplo 1d, el sistema de expresión de baculovirus se utilizará para la producción de expresión de FVIII para construcciones genéticas que portan derivados de ITR parvovirales no AAV de B19d135 y
10 GPVd162 en una forma de ADNec en células de insecto.

Ejemplo 4b. Inyección sistémica de construcciones de expresión de ADNec que portan derivados de ITR parvovirales no AAV de B19d135 y GPVd162 en ratones HemA.

[0554] Para evaluar la expresión de FVIII in vivo a partir de construcciones de ADNec que portan derivados de ITR parvovirales no AAV de B19d135 y GPVd162, se administrarán 5, 10, 20 o 50
15 µg/ratón de cada casete de expresión genética de ADNec sistémicamente a través de HDI en ratones HemA de 5-12 semanas de edad. Las muestras de sangre se recolectarán a los 1, 3, 7, 14, 21 y 28 días p.i., y a continuación una vez al mes durante un período de 4 meses. La actividad del FVIII en sangre se analizará mediante el ensayo de actividad cromogénica del FVIII.

Ejemplo 5. Generación de formulaciones de nanopartículas lipídicas de construcciones de expresión de ADNhs y ADNec de FVIII

20

[0555] Después de producir cada ADNhs o ADNec como se describe en los Ejemplos 1 y 4, cada construcción genética se formulará en nanopartículas lipídicas (LNP) utilizando composiciones lipídicas apropiadas mediante mezcla de microfluidos (LNP-ADNhs y LNP-ADNec). La razón de lípidos a ADN (N/P) se ajustará para optimizar la transducción celular y la expresión de FVIII. Los plásmidos parentales
25 que codifican casetes de expresión de FVIII flanqueados por ITR parvovirales de AAV o no AAV formulados en LNP se utilizarán como controles para la eficacia de la transducción.

Ejemplo 6. Evaluación in vitro e in vivo de LNP-ADNhs y LNP-ADNec

Ejemplo 6a. Evaluación in vitro de la expresión de FVIII mediada por ADNhs y ADNec en hepatocitos cultivados.

[0556] Se formularon construcciones genéticas de expresión de ADNhs o ADNec de FVIII y los correspondientes plásmidos de control parental en LNP, como se describe en el Ejemplo 5, para la administración de genes dirigida. Se sembraron células Huh7 en placas de cultivo de tejidos de 24 pocillos a 1×10^5 células/pocillo y se incubó durante la noche. Al día siguiente, se añadieron LNP-ADNhs o formulaciones a las células a 1000, 500, 250, 125 y 62,5 ng/pocillo. El medio de cultivo se recogió 48 horas después de la transducción después de un cambio de medio 24 horas después de la transducción. La actividad del FVIII en el medio de cultivo se midió mediante el ensayo de actividad cromogénica del FVIII en comparación con un patrón FACT de plasma humano. El plásmido que portaba el casete FVIIIco6XTEN bajo el promotor CAGp y flanqueado por las ITR de AAV se encapsuló en nanopartículas lipídicas a razones N/P de 72, 36 y 18 (FIG. 8A). Tras la transducción de las células Huh7, se midió el FVIII en el medio acondicionado. La transducción de células con la razón N/P de 18 generó un aumento de los niveles de FVIII de razones de 36 y 72, con una dosis máxima de 1 $\mu\text{g/ml}$ que dio como resultado más de 2 UI/ml. Estos datos demuestran la utilidad del suministro de LNP en las células diana del hígado. Para investigar la eficacia de transducción de ADNhs bajo un promotor específico de hígado vía suministro de LNP, se encapsuló el casete FVIIIco6XTEN bajo el promotor TTPp a razones 2 N/p y se transdujeron células Huh7 (FIG. 8B). De acuerdo con nuestros datos anteriores (FIG. 8A), la razón N/P de 18 dio como resultado un aumento de los niveles de actividad de FVIII en comparación con la razón de 36. Además, estos datos demostraron el suministro de LNP de prueba de concepto de ADNhs de FVIII en células hepáticas. Después de 24 horas, aproximadamente 2×10^5 células Huh7 transducidas con 2 $\mu\text{g/ml}$ de FVIIIco6XTEN-AAV de hebra sencilla produjeron 0,33 UI/ml de FVIII.

[0557] Además, se ha demostrado en la bibliografía que las histonas celulares se colocan regularmente a lo largo de los episomas de rAAV, creando una estructura de tipo cromatina que es similar al patrón de nucleosoma de ADN cromosómico celular. Por lo tanto, la capacidad de estas construcciones para establecer estructuras nucleosómicas de tipo cromatina necesarias para la transducción persistente de células diana también se evaluará mediante transferencia Southern.

Ejemplo 6b. Evaluación de la expresión de FVIII a largo plazo mediada por ADNhs y ADNec mediada por LNP en ratones HemA después de la administración intravenosa.

[0558] A los ratones HemA de 5-12 semanas de edad se les administrará LNP-ADNhs, LNP-ADNec o LNP-ADNp (control de plásmido) a 5, 10, 20, 40, 100 $\mu\text{g/ratón}$ mediante inyección IV, N = 4/grupo. Las

muestras de sangre se recolectarán en puntos de tiempo seleccionados a partir de las 48 horas posteriores a la inyección durante hasta 6 meses y la actividad del FVIII en la sangre se analizará mediante el ensayo de actividad cromogénica del FVIII. El perfil de expresión de FVIII en ratones tratados con LNP-ADNhs o LNP-ADNec se comparará con el de ratones tratados con LNP-ADNp para cada construcción genética descrita en los Ejemplos 1 y 4.

Ejemplo 6c. Evaluación in vivo de la expresión de FVIII mediada por ADNhs o ADNec después de una inyección de refuerzo.

[0559] A un subconjunto de ratones tratados con LNP-ADNhs o LNP-ADNec en el Ejemplo 6b se le administrará una inyección intravenosa adicional de refuerzo de las LNP correspondientes a la misma dosis 2 meses después de la inyección inicial. Las muestras de sangre se recolectarán en momentos seleccionados a partir de las 48 horas posteriores a la inyección de refuerzo durante un máximo de 6 meses. La actividad del FVIII en sangre se analizará mediante el ensayo de actividad cromogénica del FVIII. El perfil de expresión de FVIII en ratones tratados con LNP-ADNhs o LNP-ADNec se comparará con el de ratones tratados con el correspondiente LNP-ADNp.

Ejemplo 7. Utilidad de las construcciones de expresión genética que portan ITR de origen B19 o GPV para uso general en terapia génica.

Ejemplo 7a. Generación de construcciones genéticas informadoras que portan ITR de origen B19 o GPV.

[0560] Con el fin de demostrar la utilidad de los sistemas de expresión genética no basados en ITR de AAV como plataforma para uso general en aplicaciones de terapia génica, se generaron construcciones informadoras que comprendían un casete de expresión con proteína verde fluorescente (GFP) o luciferasa (luc) flanqueado por las ITR de B19d135 o GPVd162 basándose en las construcciones descritas en el Ejemplo 1b. Por tanto, los marcos de lectura abiertos (ORF) de FVIII en B19-FVIIIco6XTEN (FIG. 1C) y de GPV-FVIIIco6XTEN (FIG. 1D) se reemplazaron por ORF de GFP o luc mediante técnicas convencionales de clonación molecular.

[0561] También se generaron casetes de expresión flanqueados por las ITR de B19d135 o GPVd162 que contenían el transgén de fenilalanina hidroxilasa (PAH) murina (FIG. 7A), que se utilizaron para evaluar la expresión de PAH y la reducción de las concentraciones de fenilalanina en sangre en un modelo de ratón relevante de fenilcetonuria. Utilizando este modelo, a los ratones PKU (n = 3) se les administraron 200 µg de ADNhs flanqueado por ITR no AAV mediante inyección hidrodinámica para expresión hepática. Se recogieron muestras de sangre los días 3, 7, 14, 28, 42, 56, 70 y 81 y se aisló plasma para la

determinación de la concentración de fenilalanina (Fig. 7B-7C). Los ratones que recibieron el casete de expresión que contenía la ITR de B19d135 mostraron una reducción de los niveles de fenilalanina de 370 µg/ml a 210 µg/ml el día 3 que se mantuvo estable hasta el día 81 (FIG. 7B). Los ratones que recibieron el casete GPVd162 ITR demostraron una reducción de los niveles de fenilalanina en sangre de 350 µg/ml a 310 µg/ml el día 14, que continuó descendiendo hasta un nivel estable de 250 µg/ml el día 42 (FIG. 7C). Estas disminuciones en las concentraciones de fenilalanina en sangre representan una reducción del 45% y del 30% en comparación con las concentraciones antes de la inyección (FIG. 7D). Para confirmar la presencia de proteína PAH murina en el hígado, se realizó una transferencia Western en productos lisados de hígado tomados de ratones tratados el día 81 después de la inyección. Utilizando el anticuerpo anti-etiqueta FLAG para detectar la proteína PAH murina, la FIG. 7E demuestra proteína PAH murina detectable en 5 de 6 animales tratados, con niveles de proteína significativamente más altos observados en ratones tratados con ADNhs que contiene las ITR de B19d135. Estos datos son compatibles con las reducciones de fenilalanina en sangre observadas en las FIG. 7B-7D. Juntas, estas demuestran que el suministro de ADN de hebra sencilla puede dar como resultado la expresión a largo plazo de enzimas hepáticas funcionales.

[0562] Las secuencias de las diversas construcciones de PAH utilizadas en el experimento se exponen en las Tablas 10A y 10B.

Tabla 10A: Construcción de B19-PAH que porta ITR de B19d135 (nucleótidos 1-4146; SEQ ID NO: 197)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 180)	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATC TTAGTGGCACGTCAACCCCAAGCGCTGGCCAGAGCCAACCCTAATTCGGAAGTCCCGC CCACCGGAAGTGACGTACAGGAAATGACGTACAGGAAATGACGTAATTGTCCGCCATC TTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTGGTGTCTTCTTT TAAATTTT
Promotor CAGp (SEQ ID NO:195)	CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCAT TGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC CAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGT ACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCC CCACCCCCAATTTTGTATTTATTTATTTTAAATTATTTTGTGCAGCGATGGGGGCGGGG GGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGG CGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCG AGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGCG

Intrón Sintético (SEQ ID NO:192)	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTAATGACGG CTTGTTCCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGC GGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAG GGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCA TGCCCTTCTTCTTTTCTACAG
Secuencia de PAH murina (SEQ ID NO:196)	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGAC GATGACAAGGCTGCTGTGGTTCTGGAAAATGGCGTGCTGAGCCGGAAGCTGAGCGACTTC GGACAAGAGACAAGCTACATCGAGGACAACAGCAACCAGAATGGCGCCGTGTCTCTGATC TTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCAAGGTGCTGAGACTGTTTCGAGGAAAAC GAGATCAATCTGACCCACATCGAGAGCAGACCCAGCAGACTGAACAAGGACGAGTACGAG TTCTTCACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAGCATCATCAAGAGCCTG AGAAACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGACAAAGAAAAGAACACCGTG CCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTTCGCCAACCAAGATCCTGAGCTAT GGCGCCGAGCTGGACGCTGATCACCTGGCTTTAAGGACCCCGTGTACCGGGCCAGAAGA AAGCAGTTTGCCGATATCGCCTACAACACCAGGACGGCCAGCCTATTCTCGGGTCGAG TACACCGAGGAAGAGAGAAAGACCTGGGGCACCCTGTTTCAGAACCTGAAGGCCCTGTAC AAGACCCACGCCCTGCTACGAGCACAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGC TTCCGCGAGGACAATATCCCTCAGCTCGAAGACGTGTCCAGTTCTGCAGACCTGCACC GGCTTTAGACTGAGGCCTGTTGCCGACTGCTGAGCAGCAGAGATTTTCTCGGCGGCCTG GCCTTCAGAGTGTTCCACTGTACCCAGTACATCAGACACGGCAGCAAGCCCATGTACACC CCTGAGCCTGATATCTGCCACGAGCTGCTGGGACATGTGCCCTGTTCAGCGATAGAAGC TTCGCCCAGTTCAGCCAAGAGATCGGACTGGCTTCTCTGGGAGCCCCGTGACGAGTACATT GAGAAGCTGGCCACCATCTACTGGTTACCGTGGAATTCGGCCTGTGCAAAGAGGGCGAC AGCATCAAGGCTTATGGCGCTGGACTGCTGTCTAGCTTCGGCGAGCTGCAGTACTGTCTG AGCGACAAGCCTAAGCTGCTGCCCCTGGAACGGAAAAGACCGCCTGCCAAGAGTACACA GTGACCGAGTTCCAGCCTCTGTACTACGTGGCCGAGAGCTTCAACGACGCCAAAGAAAAA GTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTAGCGTCAGATACGACCCCTACACA CAGCGCGTGGAAGTGCTGGACAACACACAGCAGCTGAAGATTCTGGCCGACTCCATCAAC AGCGAAGTGGGCATTCTGTGTACGCCCTGCAGAAGATCAAGAGCTGA
WPRE (elemento regulador post- transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGT TGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTC CCGTATGGCTTTTCATTTTCTCCTCCTTGATAAATCCTGGTTGCTGTCTCTTTATGAGGA GTTGTGGCCCGTTGTCAAGCAACGTGGCGTGGTGTGCACTGTGTTTGTGACGCAACCCC CACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCT CCCTATTGCCACGGCGGAACATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCG GCTGTTGGGCACTGACAATTCCGTGGTGTGTGCGGGAAATCATCGTCCTTTCCCTTGGCT GCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCGGC CCTCAATCCAGCGGACCTTCCTTCCCGCGGCTGCTGCCGGCTCTGCGGCCTCTTCCGCG TCTTCGCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTG
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGA CCCTGGAAAGTGCCACTCCCACTGTCTTTCCTAATAAAATGAGGAAATTGCATCGCATT GTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGG ATTGGGAAGACAATAGCAGGCATGCTGGGGA
Repetición terminal invertida ITR 3'	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTC CGGTACAAGATGGCGGACAATTACGTCAATTTCTGTGACGTCAATTCCTGTGACGTCACT TCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGT

(SEQ ID NO: 181)	GCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTG GCCCAGAG
Secuencia de longitud completa (SEQ ID NO: 197)	
CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTC AACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCACCGGAAGTGACGTCACAGGAAAT GACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGC ATCTGATTTGGTGTCTTCTTTTAAATTTTGCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATT TAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCC GCCTCGACGACTAGCGTTTAGTAATGAGACGCACAACTAATATCACAACTGGAAATGTCTATCAATATATA GTTGCTCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGCTTAC ATACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTAT GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCACT TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTG GCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATT ACCATGCATGGTGCAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATT TTGTATTTATTTATTTTAAATTTTGTGTCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCG GGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCT CCGAAAGTTTCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGA GTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCGCCTCGCGCCGCCCCGCCCGGCTCTGACTGA CCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTAAATG ACGGCTTGTTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGC GGCTCGGGGCTGTCCGCGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTTCGGCTTCTGGCGT GTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATGCCTTCTTCTTTTCTTACAGCTCCTGGGCAACG TGCTGGTTATTGTGCTGTCTCATCATTTTTGGCAAAGAATTGGATCGCGAAGCCGCCACCATGGACTACAAAGA CCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGGCTGCTGTGGTTCTGGAA AATGGCGTGCTGAGCCGGAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAACAGCAACCAGA ATGGCGCCGTGTCTCTGATCTTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCAAGGTGCTGAGACTGTTTGA GAAAACGAGATCAATCTGACCACATCGAGAGCAGACCCAGCAGACTGAACAAGGACGAGTACGAGTTCTTC ACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAGCATCATCAAGAGCCTGAGAAACGACATCGGCGCCA CCGTGCACGAGCTGAGCAGAGACAAAGAAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGA CAGATTCGCCAACCAGATCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCTGGCTTTAAGGACCCCGTG TACCGGGCCAGAAGAAAGCAGTTTGCCGATATCGCCTACAACCTACCGGCACGGCCAGCCTATTCTCGGGTGC AGTACACCGAGGAAGAGAGAAAGACCTGGGGCACCGTGTTGAGAACCCTGAAGGCCCTGTACAAGACCCACGC CTGCTACGAGCACAAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGCTTCCGCGAGGACAATATCCCTCAG CTCGAAGACGTGTCCAGTTCTTGCAGACCTGCACCGGCTTTAGACTGAGGCCTGTTGCCGGACTGCTGAGCA GCAGAGATTTTCTCGGCGGCCCTGGCCTTCAGAGTGTTCCACTGTACCCAGTACATCAGACACGGCAGCAAGCC CATGTACACCCCTGAGCCTGATATCTGCCACGAGCTGCTGGGACATGTGCCCCCTGTTACGCGATAGAAGCTTC GCCCAGTTACGCCAAGAGATCGGACTGGCTTCTCTGGGAGCCCCCTGACGAGTACATTGAGAAGCTGGCCACCA TCTACTGGTTACCCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCTGGACTGCT GTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGCTGCCCCCTGGAAGTGGAAAAGACC GCCTGCCAAGAGTACACAGTGACCGAGTTCCAGCCTCTGTACTACGTGGCCGAGAGCTTCAACGACGCCAAAG AAAAAGTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTAGCGTCAGATACGACCCCTACACACAGCGCGT GGAAGTGCTGGACAACACACAGCAGCTGAAGATTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGTGT CACGCCCTGCAGAAGATCAAGAGCTGAGCAAGTAATGAGCGCTGATCATAATCAACCTCTGGATTACAAAATT TGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT TGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGGCTGACGCAACCCCCACTGGT TGGGGCATTGCCACCACTGTGAGCTCCTTTCCGGGACTTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAAC TCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGCTGGTGTGTGTC GGGGAATCATCGTCCTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGACGTCCTTCTGCG	

TACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGC
 GTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCGGCCTCCCCGCTGATCAGCCTCGACTGTGC
 CTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCAC
 TGTCCTTTTCTAATAAAATGAGGAAATTGCATCGCATGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGG
 GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
 CTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATT
 AGGGATAACAGGGTAATGGCGCGGGCCGCAAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGC
 CGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCTGTGACGTCATTTCTGTGACGTC
 ACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGA
 TCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG

Tabla 10B: Construcción de GPV-PAH que porta las ITR de GPVd162 (nucleótidos 1-4214; SEQ ID NO: 198)

Descripción	Secuencia
ITR 5' (SEQ ID NO: 183)	CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTGAGCCCAAT AGTTAAGCCGGAACACGTCACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAA CACGTGACAGGAAGCACGTGACCGGAAGTACGTCACCGGATGTGCGTCACCGGAAGCA TGTGACCGGAAGTTCGCTCACTTCCCCCTCCCTGATTGGCTGGTTTGAACGAACGAA CCTCCAATGAGACTCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG
Promotor CAGp (SEQ ID NO:195)	CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTT CCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGC CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATT GACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT TATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAG TCATCGCTATTACCATGCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCA TCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTAAATTATTTTGTGC AGCGATGGGGGCGGGGGGGGGGGGGGGGGCGCGGCCAGGCGGGGCGGGGCGGGGCGAG GGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCC GAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGC GCGGCGGGCG
Intrón Sintético (SEQ ID NO:192)	GTGAGCGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTAATGAC GGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTT GTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACG GGGCAGGGCGGGGTTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAAC CATGTTTCATGCCTTCTTCTTTTTCCTACAG
Secuencia de PAH (SEQ ID NO:196)	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATG ACGATGACAAGGCTGCTGTGGTTCTGGAAAATGGCGTGCTGAGCCGGAAGCTGAGCGA CTTTCGGACAAGAGACAAGCTACATCGAGGACAACAGCAACCAGAATGGCGCCGTGTCT CTGATCTTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCAAGGTGCTGAGACTGTTTCG AGGAAAACGAGATCAATCTGACCCACATCGAGAGCAGACCCAGCAGACTGAACAAGGA CGAGTACGAGTTCTTACCTACCTGGACAAGCGGAGCAAGCCTGTGCTGGGCAGCATC ATCAAGAGCCTGAGAAACGACATCGGCGCCACCGTGCACGAGCTGAGCAGAGACAAAG AAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTTCGCCAA CCAGATCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCTGGCTTTAAGGACCCC GTGTACCGGGCCAGAAGAAAGCAGTTTGCCGATATCGCCTACAACCTACCGGCACGGCC

	AGCCTATTCTCGGGTCGAGTACACCGAGGAAGAGAGAAAAGACCTGGGGCACCGTGTT CAGAACCCTGAAGGCCCTGTACAAGACCCACGCCTGCTACGAGCACAACCACATCTTC CCACTGCTGGAAGTACTGCGGCTTCCGCGAGGACAATATCCCTCAGCTCGAAGACG TGTCCTCAGTTCTGTCAGACCTGCACCGGCTTTAGACTGAGGCCTGTTGCCGGACTGCT GAGCAGCAGAGATTTCTCGGCGGCCTGGCCTTCAGAGTGTTCCACTGTACCCAGTAC ATCAGACACGGCAGCAAGCCCATGTACACCCCTGAGCCTGATATCTGCCACGAGCTGC TGGGACATGTGCCCCCTGTTTCAGCGATAGAAGCTTCGCCCAGTTTCAGCCAAGAGATCGG ACTGGCTTCTCTGGGAGCCCCCTGACGAGTACATTGAGAAGCTGGCCACCATCTACTGG TTCACCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCTG GACTGCTGTCTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGCT GCCCCTGGAAGTGGAAAAGACCGCCTGCCAAGAGTACACAGTGACCGAGTTCCAGCCT CTGTACTACGTGGCCGAGAGCTTCAACGACGCCAAAAGAAAAGTGCAGGACCTTCGCCG CCACCATTCTCGGCCTTTTAGCGTCAGATACGACCCCTACACACAGCGCGTGGAAGT GCTGGACAACACACAGCAGCTGAAGATTCTGGCCGACTCCATCAACAGCGAAGTGGGC ATTCTGTGTCACGCCCTGCAGAAGATCAAGAGCTGA
WPRe (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120)	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTA TGAGGAGTTGTGGCCCGTTGTGAGCAACGTGGCGTGGTGTGCACTGTGTTTGTGCTGAC GCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCCG CTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCTGCCTTGCCCCGTGCTG GACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTGCGGGGAAATCATCG TCCTTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCT GCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGC TCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGG GCCGCTCCCCGCTG
bGHpA (señal de poliadenilación de la hormona de crecimiento bovina) (SEQ ID NO:122)	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTT GACCCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTAATAAATGAGGAAATTGCATCG CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGG GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
Repetición terminal invertida ITR 3' (SEQ ID NO: 184)	CACTTCCTGGCGCGCAAAATATCCTCTTGCTCCTTGAGTCTCATTGGAGGGTTTCGTTCCG TTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCG GTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCTGTACGTGTTTCCGGT CACGTGACTTCCGGTCATGTGACTTCCGGTGACGTGTTTCCGGCTTAACTATTGGGCT GACCGCGCGGCATGCGCGTGGTCAACCTAACAGCCGGAAACACGTACCG
Secuencia de longitud completa (SEQ ID NO: 198)	
CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGGGTCAGCCCAATAGTTAAGCCGAAA CACGTACCCGGAAGTCACATGACCGGAAGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAAC TACGTACCCGGATGTGCGTCACCGGAAGCATGTGACCGGAACCTGCGTCACTTCCCCCTCCCCTGATTGGCT GGTTTGAACGAACGAACCCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGC GCGCCAGGAAGTGGCGGCA ATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACG CGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAGTAATGAGACGCA CAAATAATATCACAACTGGAAATGTCTATCAATATATAGTTGCTCTAGTTATTAATAGTAATCAATTACG GGGTCAATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGA CCGCCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCA AGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCAGTACATGACCTTATGG	

GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGCATGGTCGAGGTGAGCCCCAC
 GTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTTGTATTTATTTATTTTTTAATTATT
 TTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGAGGGGCGGGG
 CGGGGCGAGGCGGAGAGGTGCGGGCGGACCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAG
 GCGGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGGAGTCGCTGCGCGCTGCCTTCGCCC
 CGTGCCCCGCTCCGCCCGCCGCTCGCGCCGCCGCCCGGCTCTGACTGACCGCGTTACTCCACAGGTGAG
 CGGGCGGGACGGCCCTTCTCCTTCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTTTCTTTCTGTG
 GCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGCTGTCCGCGGG
 GGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAG
 CCTCTGCTAACCATGTTTCATGCCTTCTTCTTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGT
 CTCATCATTTTGGCAAAGAATTGGATCGCGAAGCCGCCACCATGGACTACAAAGACCATGACGGTGATTATA
 AAGATCATGATATCGATTACAAGGATGACGATGACAAGGCTGCTGTGGTTCTGGAAGATGGCGTGCTGAGCC
 GGAAGCTGAGCGACTTCGGACAAGAGACAAGCTACATCGAGGACAACAGCAACCAGAATGGCGCCGTGTCTC
 TGATCTTCAGCCTGAAAGAAGAAGTGGGCGCCCTGGCCAAGGTGCTGAGACTGTTTCGAGGAAAACGAGATCA
 ATCTGACCCACATCGAGAGCAGACCCAGCAGACTGAACAAGGACGAGTACGAGTTCTTCACCTACCTGGACA
 AGCGGAGCAAGCCTGTGCTGGGCAGCATCATCAAGAGCCTGAGAAACGACATCGGCGCCACCGTGACGAGC
 TGAGCAGAGACAAAGAAAAGAACACCGTGCCATGGTTCCCCAGGACCATCCAAGAGCTGGACAGATTGCGCA
 ACCAGATCCTGAGCTATGGCGCCGAGCTGGACGCTGATCACCTGGCTTTAAGGACCCCGTGACCGGGCCA
 GAAGAAAGCAGTTTGCCGATATCGCCTACAACCTACCGGCACGGCCAGCCTATTCTCGGGTCGAGTACACCG
 AGGAAGAGAGAAAGACCTGGGGCACCGTGTTTCAAGACCTGAAGGCCCTGTACAAGACCCACGCCTGCTACG
 AGCACAACCACATCTTCCCACTGCTGGAAAAGTACTGCGGCTTCCGCGAGGACAATATCCCTCAGCTCGAAG
 ACGTGTCACGTTTCTGACAGCTGCACCGGCTTTAGACTGAGGCCTGTTGCCGGACTGCTGAGCAGCAGAG
 ATTTTCTCGGCGGCTGGCCTTCAGAGTGTTCCACTGTACCCAGTACATCAGACACGGCAGCAAGCCCATGT
 ACACCCCTGAGCCTGATATCTGCCACGAGCTGCTGGGACATGTGCCCTGTTTTCAGCGATAGAAGCTTCGCCC
 AGTTCAGCCAAGAGATCGGACTGGCTTCTCTGGGAGCCCCTGACGAGTACATTGAGAAGCTGGCCACCATCT
 ACTGGTTCACCGTGGAATTCGGCCTGTGCAAAGAGGGCGACAGCATCAAGGCTTATGGCGCTGGACTGCTGT
 CTAGCTTCGGCGAGCTGCAGTACTGTCTGAGCGACAAGCCTAAGCTGCTGCCCCTGGAAGTGGAAAAGACCG
 CCTGCCAAGAGTACACAGTGACCGAGTTCAGCCTCTGTACTACGTGGCCGAGAGCTTCAACGACGCCAAAG
 AAAAGTGCGGACCTTCGCCGCCACCATTCCTCGGCCTTTTAGCGTCAGATACGACCCCTACACACAGCGCG
 TGGAAGTGCTGGACAACACACAGCAGCTGAAGATTCTGGCCGACTCCATCAACAGCGAAGTGGGCATTCTGT
 GTCACGCCCTGCAGAAGATCAAGAGCTGAGCAAGTAATGAGCGCTGATCATAATCAACCTCTGGATTACAAA
 ATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG
 CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCT
 CTTTATGAGGAGTTGTGGCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCC
 ACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACG
 GCGGAATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTG
 GTGTTGTGCGGGAAATCATCGTCCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACG
 TCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGG
 CCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTGATCAGC
 CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGG
 TGCCACTCCCCTGTCTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTAT
 TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGC
 GGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTAGC
 ATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCACTTCTGGCGCGCAAAATATCCTCTT
 GTCCTTGAGTCTCATTGGAGGGTTCGTTCTGTTTGAACAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTC
 CGGTACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCTGTACGTGTTTCCG
 GTCACGTGACTTCCGGTATGTGACTTCCGGTGACGTGTTTCCGGCTTAACATTGGGCTGACCGCGCGGCA
 TGCGCGTGGTCAACCTAACAGCCGGAACACGTACCG

Ejemplo 7b. Preparación de construcciones genéticas informadoras de ADNhs que portan ITR de origen B19 o GPV.

5 [0563] Se prepararán construcciones de informador de ADNhs/PAH como se describe en el Ejemplo 1c. Brevemente, los plásmidos se digirán con Lg^{ul}. Los fragmentos de ADNhs con estructuras de ITR en horquilla formadas se generarán desnaturalizando los productos de fragmentos de ADN de doble hebra (casete de expresión informador y esqueleto del plásmido) de la digestión de Lg^{ul} a 95°C y a continuación enfriando a 4°C para permitir que las secuencias de ITR palindrómicas se plieguen. (FIG. 1A). Las construcciones de ADNhs resultantes se probarán en ratones para determinar la capacidad de establecer la transducción persistente de hígado, tejido muscular, fotorreceptores en el ojo y sistema nervioso central (SNC).

Ejemplo 7c. Evaluación in vivo de la expresión del informador mediada por ADNhs.

15 [0564] Para validar la capacidad de las construcciones de informador de ADNhs descritas en el Ejemplo 7b para mediar la expresión transgénica persistente in vivo, se inyectarán 5, 10 o 20 µg/ratón a ratones de 5-12 semanas (4 animales/grupo) de ADNhs informador sistémicamente, localmente para dirigirlas al tejido muscular y células del SNC, y/o subretinalmente para dirigirlas a células fotorreceptoras.

[0565] Para evaluar la expresión de PAH a partir de construcciones de expresión basadas en ITR de B19 y GPV, se utilizará un modelo de ratón de enfermedad relevante. Estas construcciones genéticas serán administradas sistémicamente mediante HDI para dirigirlas al hígado.