

COMPOSICIONES Y MÉTODOS PARA EL TRATAMIENTO DE LA PÉRDIDA AUDITIVA NEUROSENSORIAL MEDIANTE EL USO DE SISTEMAS DE VECTORES DUALES DE ESTEREOCILINA

Listado de secuencias

La presente solicitud contiene un listado de secuencias que se presentó de forma electrónica en formato ASCII y se incorpora a la presente en su totalidad mediante referencia. La copia ASCII, creada el 4 de mayo de 2022, se denomina 51471-011WO2_Sequence_Listing_5_4_22_ST25 y tiene un tamaño de 113,251 bytes.

Campo de la invención

Se describen en la presente composiciones y métodos para el tratamiento de la pérdida auditiva, en particular formas de la enfermedad que están asociadas con mutaciones en la estereocilina (STRC) mediante terapia génica de STRC, en la que la expresión del gen STRC está bajo el control regulador de un promotor de oncomodulina (OCM). La divulgación proporciona sistemas de expresión de dos vectores que incluyen un primer vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo N de una proteína estereocilina y un segundo vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo C de una proteína estereocilina. Estos vectores pueden utilizarse para aumentar la expresión o proporcionar STRC de tipo salvaje a una célula o sujeto, tal como un sujeto que sufre pérdida auditiva (por ejemplo, pérdida auditiva neurosensorial).

Antecedentes

La pérdida auditiva neurosensorial es un tipo de pérdida auditiva causada por defectos en las células del oído interno o en las vías neurales que se proyectan desde el oído interno hasta el cerebro. Aunque la pérdida auditiva neurosensorial a menudo se adquiere y puede ser causada por ruido, infecciones, traumatismo craneoencefálico, fármacos ototóxicos o envejecimiento, también existen formas congénitas de pérdida auditiva neurosensorial asociadas con mutaciones autosómicas recesivas. Una de estas formas de pérdida auditiva neurosensorial autosómica recesiva se asocia con la mutación del gen de la estereocilina (STRC). La estereocilina es una proteína de gran tamaño codificada por el gen STRC del cromosoma 15q15, que contiene 29 exones que abarcan aproximadamente 19 kb del genoma. El gen STRC está duplicado en tándem, donde la segunda copia contiene un codón de parada prematuro en el exón 20, produciendo así un pseudogén STRC. Estudios anteriores han identificado mutaciones en STRC en familias con pérdida auditiva neurosensorial autosómica recesiva no sindrómica (Verpy et ál., Nat. Genet.

29:345-9 (2001)). La expresión de la proteína estereocilina se limita a los estereocilios de los haces ciliados de las células ciliadas del oído interno. Se cree que la proteína estereocilina forma conectores superiores horizontales y coronas tectoriales de unión a la membrana, necesarios para el funcionamiento normal del aparato auditivo (Avan et ál., PNAS 116:25948-57 (2019); Verpy et ál., J. Comp. Neurol. 519:194-210 (2011)). Se ha demostrado que los ratones que carecen de estereocilina presentan haces de células ciliadas anormales con cohesión defectuosa y problemas de audición (Verpy et ál., Nature 456:255-8 (2008)).

En los últimos años, los esfuerzos para tratar la pérdida auditiva se han centrado cada vez más en la terapia génica como una posible solución; sin embargo, el gen STRC es demasiado grande para permitir el tratamiento con enfoques estándar de terapia génica. Existe la necesidad de nuevos tratamientos para tratar la pérdida auditiva neurosensorial relacionada con STRC.

Compendio de la Invención

La presente invención proporciona composiciones y métodos para tratar la pérdida auditiva neurosensorial en un sujeto, tal como un sujeto humano. Las composiciones y métodos de la divulgación pertenecen a sistemas de vectores duales para el suministro de un polinucleótido que codifica una proteína estereocilina a un sujeto que tiene o está en riesgo de desarrollar pérdida auditiva neurosensorial (por ejemplo, un sujeto con una mutación en STRC). Por ejemplo, mediante el uso de las composiciones y métodos descritos en la presente, un primer vector de ácido nucleico y un segundo vector de ácido nucleico que codifican cada uno una parte de una proteína funcional de estereocilina puede administrarse a un sujeto mediante terapia génica viral. Las composiciones y métodos descritos en la presente también se pueden usar para aumentar la expresión de una proteína estereocilina de tipo salvaje en una célula ciliada coclear (por ejemplo, una célula ciliada externa).

En un primer aspecto, la invención proporciona un sistema de dos vectores que comprende (a) un primer vector de ácido nucleico que comprende un promotor de oncomodulina (OCM) que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con cualquiera de las SEQ ID NO: 1-3 unido de forma operativa a un primer polinucleótido que codifica una parte de extremo N de una proteína estereocilina; y (b) un segundo vector de ácido nucleico que incluye un segundo polinucleótido que codifica una parte de extremo C de una proteína estereocilina.

En algunas modalidades, el primer polinucleótido se superpone parcialmente con el segundo polinucleótido. En algunas modalidades, el primer polinucleótido y el segundo polinucleótido tienen una región de superposición que tiene una longitud de al menos 200 bases (b) (por ejemplo, al menos 200 b, 300 b, 400 b, 500 b, 600 b, 700 b, 800 b, 900 b, 1,0 kilobase (kb), 1,1 kb, 1,2 kb, 1,3

kb, 1,4 kb, 1,5 kb o más). En estas modalidades, cuando se introduce en una célula de mamífero, el primer y segundo vector de ácido nucleico se someten a una recombinación homóloga para formar un polinucleótido recombinado que codifica una proteína estereocilina de longitud completa. En algunas modalidades, el primer vector de ácido nucleico incluye un polinucleótido que incluye la

5 secuencia de nucleótidos 225 a 4574 de la SEQ ID NO: 43. En algunas modalidades, el segundo vector de ácido nucleico incluye un polinucleótido que incluye la secuencia de nucleótidos 211 a 4219 de la SEQ ID NO: 44.

En algunas modalidades, el primer vector de ácido nucleico incluye una secuencia señal donante de empalme ubicada en el extremo 3' del primer polinucleótido y el segundo vector de ácido

10 nucleico incluye una secuencia señal aceptora de empalme ubicada en el extremo 5' del segundo polinucleótido. En algunas modalidades, el primer y segundo polinucleótido no se superponen.

En algunas modalidades, el primer vector de ácido nucleico incluye una secuencia señal donante de empalme posicionada en el extremo 3' del primer polinucleótido y una primera región recombinogénica posicionada 3' de la secuencia señal donante de empalme y el segundo vector

15 de ácido nucleico incluye una segunda región recombinogénica, una secuencia señal aceptora de empalme 3' de la región recombinogénica, y el segundo polinucleótido 3' de la secuencia señal aceptora de empalme. En algunas modalidades, el primer y segundo polinucleótido no se superponen. En algunas modalidades, la primera y la segunda región recombinogénica son la misma. En algunas modalidades, la primera región recombinogénica y la segunda región

20 recombinogénica son un fragmento del gen AP. En algunas modalidades, el fragmento del gen AP incluye o consiste en la secuencia de cualquiera de las SEQ ID NO: 47-52. En algunas modalidades, el fragmento del gen AP incluye o consiste en la secuencia de la SEQ ID NO: 50. En algunas modalidades, el primer vector de ácido nucleico incluye además una secuencia señal de degradación ubicada a 3' de la región recombinogénica; y el segundo vector de ácido nucleico

25 incluye además una secuencia señal de degradación ubicada entre la región recombinogénica y la secuencia señal aceptora de empalme. En algunas modalidades, el primer vector de ácido nucleico incluye un polinucleótido que incluye la secuencia de nucleótidos 225 a 4454 de la SEQ ID NO: 45 y el segundo vector de ácido nucleico incluye un polinucleótido que incluye la secuencia de nucleótidos 257 a 3597 de la SEQ ID NO: 46.

En algunas modalidades, el segundo vector de ácido nucleico incluye además un promotor de OCM que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la secuencia de ácidos nucleicos de cualquiera de las SEQ ID NO: 1-3 unida de forma operativa al segundo polinucleótido, donde el promotor está ubicado a 5' del segundo polinucleótido. En algunas

30 modalidades, el promotor de OCM en el segundo vector de ácido nucleico es el mismo (es decir,

35

tiene la misma secuencia de nucleótidos) que el promotor de OCM en el primer vector de ácido nucleico. En algunas modalidades, el promotor de OCM en el segundo vector de ácido nucleico tiene una secuencia de nucleótidos diferente que el promotor de OCM en el primer vector de ácido nucleico.

5 En algunas modalidades, el primer vector de ácido nucleico incluye además un polinucleótido que codifica una inteína de extremo N (N-inteína) ubicada 3' del primer polinucleótido. En algunas modalidades, el segundo vector de ácido nucleico incluye además un polinucleótido que codifica una inteína de extremo C (C-inteína) ubicada entre el promotor de OCM y el segundo polinucleótido. En algunas, la N-inteína y C-inteína son componentes de un sistema de transempalme de inteína dividida.

10 En algunas modalidades, el primer y/o segundo vector incluyen una señal de degradación de inteína. En algunas modalidades, la señal de degradación es un N-degrón y/o un C-degrón. En algunas modalidades, el N-degrón y/o el C-degrón son independientemente un degrón CL1, PB29, SMN, CIITA, u ODC. En algunas modalidades, la señal de degradación es una señal de degradación de la dihidrofolato reductasa de *E. coli* (ecDHFR). En algunas modalidades, la señal de degradación es un dominio de degradación de FKBP12 (Banaszynski et ál., Cell 126:995-1004, 2006). En algunas modalidades, la señal de degradación es un dominio de degradación de PEST (Rechsteiner and Rogers, Trends Biochem Sci. 21:267-271, 1996). En algunas modalidades, la señal de degradación es una señal de ubiquitinación de etiqueta UbR (Chassin et ál., Nat Commun. 10:2013, 2019). En algunas modalidades, la señal de degradación es una mutación desestabilizada de ELRBD humano (Miyazaki et ál., J. Am. Chem. Soc., 134:3942-3945, 2012).

20 En algunas modalidades, el primer y segundo vector, cuando se introducen en una célula de mamífero, producen una primera y segunda proteína de fusión, respectivamente, donde la primera proteína de fusión incluye la parte de extremo N de estereocilina y la N-inteína ubicada 3' de este, y donde la segunda proteína de fusión incluye la C-inteína y la parte de extremo C de estereocilina ubicada 3' de esta. En algunas modalidades, el extremo C de la N-inteína de la primera proteína de fusión y el extremo N de la C-inteína de la segunda proteína de fusión son capaces de formar un enlace peptídico, produciendo así un polipéptido que incluye, del extremo N al extremo C, la parte del extremo N de estereocilina, N-inteína, C-inteína, y la parte de extremo C de estereocilina, donde la N-inteína y la C-inteína unidas son capaces de autoexcitarse y ligar el extremo C de la parte de extremo N de estereocilina y el extremo N de la parte de extremo C de estereocilina, produciendo así una proteína estereocilina de longitud completa.

30 En algunas modalidades, el sistema de transempalme de inteína dividida deriva de un gen *DnaE* de una o más bacterias. En algunas modalidades, la una o más bacterias se seleccionan del grupo que consiste en *Nostoc punctiforme* (Npu), *Synechocystis sp.* PCC6803 (Ssp), *Fischerella sp.*

PCC9605 (Fsp), *Scytonema tolypothrichoides* (Sto), bacteria Cyanobacteria SW_9_47_5, *Nodularia spumigena* (Nsp), *Nostoc flagelliforme* (Nfl), *Crocospaera watsonii* (Cwa) WH8502, *Chroococcidiopsis cubana* (Ccu) CCALA043, *Trichodesmium erythraeum* (Ter), *Rhodothermus marinus* (Rma), *Saccharomyces cerevisiae* (Sce), *Saccharomyces castellii* (Sca), *Saccharomyces unisporus* (Sun), *Zygosaccharomyces bisporus* (Zbi), *Torulaspora pretoriensis* (Tpr), *Mycobacteria tuberculosis* (Mtu), *Mycobacterium leprae* (Mle), *Mycobacterium smegmatis* (Msm), *Pyrococcus abyssi* (Pab), *Pyrococcus horikoshii* (Pho), *Coxiella burnetti* (Cbu), *Coxiella neoformans* (Cne), *Coxiella gattii* (Cga), *Histoplasma capsulatum* (Hca) y cloroplasto *Porphyra purpurea* (Ppu). En algunas modalidades, el sistema de transempalme de la inteína dividida deriva de estudios de alineación de secuencias múltiples de *DnaE* para identificar un diseño de consenso (por ejemplo, Cfa) para diseñar una inteína dividida con la estabilidad y actividad deseables.

En algunas modalidades, la N-inteína tiene una secuencia de cualquiera de las SEQ ID NO: 8, 10, 13, 15, 17-22, 27, 29, 31, 33, 35, 37 y 39, y la C-inteína tiene una secuencia de cualquiera de las SEQ ID NO: 9, 11, 12, 14, 16, 23-26, 28, 30, 32, 34, 36, 38 y 40. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 8 y la C-inteína tiene la secuencia de la SEQ ID NO:9. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 8 y la C-inteína tiene la secuencia de la SEQ ID NO:11. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 8 y la C-inteína tiene la secuencia de la SEQ ID NO:12. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 10 y la C-inteína tiene la secuencia de la SEQ ID NO:9. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 10 y la C-inteína tiene la secuencia de la SEQ ID NO:11. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 10 y la C-inteína tiene la secuencia de la SEQ ID NO:12. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 13 y la C-inteína tiene la secuencia de la SEQ ID NO:14. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 15 y la C-inteína tiene la secuencia de la SEQ ID NO:16. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 17 y la C-inteína tiene la secuencia de la SEQ ID NO:23. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 20 y la C-inteína tiene la secuencia de la SEQ ID NO:24. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 21 y la C-inteína tiene la secuencia de la SEQ ID NO:25. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 22 y la C-inteína tiene la secuencia de la SEQ ID NO:26. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 27 y la C-inteína tiene la secuencia de la SEQ ID NO:28. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 29 y la C-inteína tiene la secuencia de la SEQ ID NO:30. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 31 y la C-inteína tiene la secuencia de la SEQ ID NO:32. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 33

y la C-inteína tiene la secuencia de la SEQ ID NO:34. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 35 y la C-inteína tiene la secuencia de la SEQ ID NO:36. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 37 y la C-inteína tiene la secuencia de la SEQ ID NO:38. En algunas modalidades, la N-inteína tiene la secuencia de la SEQ ID NO: 39 y la C-inteína tiene la secuencia de la SEQ ID NO:40. En algunas modalidades, la N-inteína tiene la secuencia de cualquiera de las SEQ ID NO: 17-22 y la C-inteína tiene la secuencia de cualquiera de las SEQ ID NO: 23-26.

En algunas modalidades, el sistema de transempalme de inteínas divididas incluye una o más inteínas que realizan el transempalme de proteínas solo al entrar en contacto con un ligando. En algunas modalidades, el ligando se selecciona del grupo que consiste en 4-hidroxitamoxifeno, un péptido, una proteína, un polinucleótido, un aminoácido o un nucleótido.

En algunas modalidades, el primer vector de ácido nucleico incluye además un polinucleótido que codifica un péptido señal. En algunas modalidades, el polinucleótido que codifica un péptido señal se coloca a 5' del polinucleótido que codifica la parte de extremo N de la proteína estereocilina. En algunas modalidades, el polinucleótido que codifica un péptido señal se coloca a 3' del polinucleótido que codifica la parte de extremo N de la proteína estereocilina. En algunas modalidades, el segundo vector de ácido nucleico incluye además un polinucleótido que codifica un péptido señal. En algunas modalidades, el polinucleótido que codifica un péptido señal se coloca a 5' del polinucleótido que codifica la parte de extremo C de la proteína estereocilina. En algunas modalidades, el polinucleótido que codifica un péptido señal se coloca a 3' del polinucleótido que codifica la parte de extremo C de la proteína estereocilina.

En algunas modalidades, ni el primero ni el segundo polinucleótido codifican una proteína estereocilina de longitud completa. En algunas modalidades, cada uno del primer y segundo polinucleótido codifica alrededor de la mitad de la secuencia de la proteína estereocilina.

En algunas modalidades, el segundo vector de ácido nucleico incluye además una secuencia poli(A) 3' del segundo polinucleótido.

En algunas modalidades, el primer y segundo vector de ácido nucleico no incluye regiones no traducidas (UTR) de STRC. En algunas modalidades, el primer y segundo vector de ácido nucleico incluye UTR de STRC. En algunas modalidades, el primer vector de ácido nucleico incluye una 5' UTR de STRC 5' del primer polinucleótido. En algunas modalidades, el segundo vector de ácido nucleico incluye una 3' UTR de STRC 3' del segundo polinucleótido.

En algunas modalidades, el primer y segundo polinucleótido que codifican la proteína estereocilina no incluyen intrones (por ejemplo, el primer y segundo polinucleótido son partes de ADNc de STRC). En algunas modalidades, el primer y segundo polinucleótido que codifica la proteína estereocilina incluye intrones.

En algunas modalidades, el promotor de OCM tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 1. En algunas modalidades, el promotor de OCM tiene la secuencia de la SEQ ID NO: 1. En algunas modalidades, el promotor de OCM tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 2. En algunas modalidades, el promotor de OCM tiene la secuencia de la SEQ ID NO: 2. En algunas modalidades, el promotor de OCM tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 3. En algunas modalidades, el promotor de OCM tiene la secuencia de la SEQ ID NO: 3.

En algunas modalidades, el sistema de dos vectores es capaz de dirigir la expresión específica de la célula ciliada externa (OHC) coclear de una proteína estereocilina de longitud completa en una OHC de mamífero. En algunas modalidades, la OHC de mamífero es una OHC humana. En algunas modalidades, la OHC de mamífero es una OHC murina.

En algunas modalidades, la proteína estereocilina es una proteína estereocilina humana que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 4. En algunas modalidades, la proteína estereocilina tiene la secuencia de la SEQ ID NO: 4. En algunas modalidades, la proteína estereocilina humana se codifica por un polinucleótido que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 6. En algunas modalidades, el polinucleótido tiene al menos 85 % (por ejemplo, al menos 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 6 que codifica la proteína estereocilina de la SEQ ID NO: 4. En algunas modalidades, la proteína estereocilina humana está codificada por un polinucleótido que tiene la secuencia de la SEQ ID NO: 6. En algunas modalidades, la proteína STRC es una proteína estereocilina murina que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 5. En algunas modalidades, la proteína estereocilina murina tiene la secuencia de la SEQ ID NO: 5. En algunas modalidades, la proteína estereocilina murina se codifica por un polinucleótido que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 7. En algunas modalidades, el polinucleótido tiene al menos 85 % (por ejemplo, al menos 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 7 que codifica la proteína estereocilina

de la SEQ ID NO: 5. En algunas modalidades, la proteína estereocilina murina está codificada por un polinucleótido que tiene la secuencia de la SEQ ID NO:7.

En algunas modalidades, el primer y segundo vector son vectores virales, plásmidos, cósmidos o cromosomas artificiales. En algunas modalidades, el primer y segundo vector son vectores virales.

5 En algunas modalidades, los vectores virales son vectores de virus adenoasociados (AAV), vectores de adenovirus o vectores de lentivirus. En algunas modalidades, el primer y segundo vector son vectores de AAV. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene un AAV1, AAV2, AAV2quad(Y-F), AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, rh10, rh39, rh43, rh74, Anc80, Anc80L65, DJ/8, DJ/9, 7m8, PHP.B, PHP.eb, o
10 cápside PHP.S. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV1. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV9. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV6. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV8. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una
15 cápside Anc80. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside Anc80L65. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside DJ/9. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside 7m8. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV2. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene
20 una cápside PHP.B. En algunas modalidades, cada uno del primer y segundo vector de AAV tiene una cápside AAV2quad(Y-F).

En otro aspecto, la invención proporciona una composición farmacéutica que contiene el sistema de dos vectores del aspecto y modalidades anteriores. En algunas modalidades, la composición incluye además un portador, diluyente o excipiente farmacéuticamente aceptable.

25 En otro aspecto, la invención proporciona una célula (por ejemplo, una célula de mamífero, por ejemplo, una célula humana, tal como una de OHC, por ejemplo, una OHC que tiene una mutación patogénica en el gen STRC) que incluye el sistema de dos vectores de cualquiera de los aspectos y modalidades anteriores. En algunas modalidades, la célula es una OHC de mamífero. En algunas modalidades, la OHC de mamífero es una OHC humana.

30 En otro aspecto, la divulgación proporciona un método para expresar una proteína estereocilina en una célula de mamífero al poner en contacto la célula con el sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores. En algunas modalidades, la célula es una célula ciliada coclear. En algunas modalidades, la célula es una OHC. En algunas modalidades, la célula es una célula de ser humano. En algunas modalidades, la célula está en un
35 sujeto (por ejemplo, el contacto ocurre in vivo).

En otro aspecto, la invención proporciona un método para tratar a un sujeto que tiene o está en riesgo de desarrollar pérdida auditiva neurosensorial al administrar a un oído interno del sujeto una cantidad eficaz del sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores. En algunas modalidades, la pérdida auditiva neurosensorial es una pérdida auditiva neurosensorial genética. En algunas modalidades, la pérdida auditiva genética es una pérdida auditiva recesiva autosómica. En algunas modalidades de cualquiera de los aspectos anteriores, la pérdida auditiva está asociada con la pérdida de las OHC o la disfunción de las OHC. En algunas modalidades, la pérdida auditiva se asocia con una desviación anormal del haz de estereocilios de las OHC o una conectividad deteriorada entre los haces ciliados de las OHC y la membrana tectorial.

En otro aspecto, la invención proporciona un método para aumentar la expresión de STRC (por ejemplo, expresión de STRC de tipo salvaje, por ejemplo, para producir proteína estereocilina de tipo salvaje) en un sujeto que lo necesite, el método incluye administrar a un oído interno del sujeto una cantidad terapéuticamente eficaz del sistema de dos vectores o la composición farmacéutica de los aspectos y las modalidades anteriores.

En otro aspecto, la invención proporciona un método para prevenir o reducir el daño a OHC o la muerte en un sujeto que lo necesite, que incluye administrar a un oído interno del sujeto una cantidad eficaz del sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores.

En otro aspecto, la invención proporciona un método para aumentar la supervivencia de la OHC en un sujeto que la necesite, que incluye administrar a un oído interno del sujeto una cantidad eficaz del sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores.

En otro aspecto, la invención proporciona un método para aumentar o mejorar la fijación de los haces ciliados de OHC a la membrana tectorial en un sujeto que lo necesite, que incluye administrar a un oído interno del sujeto una cantidad eficaz del sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores.

En algunas modalidades de cualquiera de los aspectos anteriores, el sujeto tiene una mutación en STRC. En algunas modalidades de cualquiera de los aspectos anteriores, se ha identificado que el sujeto tiene una mutación en STRC. En algunas modalidades de cualquiera de los aspectos anteriores, el método incluye además identificar que el sujeto tiene una mutación en STRC antes de administrar el sistema de dos vectores o la composición farmacéutica. En algunas modalidades de cualquiera de los aspectos anteriores, el sujeto padece sordera, autosómica recesiva 16 (DFNB16). En algunas modalidades de cualquiera de los aspectos anteriores, se ha identificado que el sujeto tiene DFNB16.

En algunas modalidades de cualquiera de los aspectos anteriores, el método incluye además evaluar la audición del sujeto antes de administrar el sistema de dos vectores o la composición farmacéutica (por ejemplo, evaluar la audición mediante pruebas estándar, tal como audiometría, respuesta auditiva del tronco encefálico (ABR), electrococcleografía (ECOG) u emisiones otoacústicas).

En algunas modalidades de cualquiera de los aspectos anteriores, el método incluye además evaluar la audición del sujeto después de administrar el sistema de dos vectores o la composición farmacéutica (por ejemplo, evaluar la audición mediante pruebas estándar, tales como audiometría, ABR, ECOG u emisiones otoacústicas).

En algunas modalidades de cualquiera de los aspectos anteriores, el sistema de dos vectores o la composición farmacéutica se administra de forma local. En algunas modalidades, el sistema de dos vectores o la composición farmacéutica se administra al oído del sujeto (por ejemplo, administrado al oído interno, por ejemplo, en la perilinfa o endolinfa, tal como en o a través de la ventana oval, la ventana redonda o el canal horizontal, o mediante inyección transtimpánica o intratimpánica). En algunas modalidades, los vectores del sistema de dos vectores se administran simultáneamente. En algunas modalidades, los vectores del sistema de dos vectores se administran secuencialmente.

En algunas modalidades de cualquiera de los aspectos anteriores, el vector o composición de ácido nucleico se administra en una cantidad suficiente para prevenir o reducir la pérdida auditiva, retrasar el desarrollo de la pérdida auditiva, ralentizar la progresión de la pérdida auditiva, mejorar la audición, mejorar la discriminación del habla, mejorar la función de las células ciliadas, prevenir o reducir el daño de las células ciliadas, prevenir o reducir la muerte de las células ciliadas, promover o aumentar la supervivencia de las células ciliadas, mejorar la fijación del haz ciliado OHC a la membrana tectorial, o aumentar la expresión de STRC en una célula ciliada.

En algunas modalidades de cualquiera de los aspectos anteriores, el sujeto es un ser humano. En otro aspecto, la invención proporciona un kit que contiene un sistema de dos vectores o la composición farmacéutica de los aspectos y modalidades anteriores.

Definiciones

Tal como se usa en la presente, el término "alrededor de" se refiere a un valor que se encuentra dentro del 10 % por encima o por debajo del valor que se describe.

Tal como se utiliza en la presente, "administración" se refiere a proporcionar o dar a un sujeto un agente terapéutico (por ejemplo, un sistema de dos vectores que contiene un promotor de oncomodulina (OCM) unido de forma operativa a un polinucleótido que codifica una proteína estereocilina), por cualquier vía eficaz. A continuación se describen en la presente vías de

administración de ejemplo.

Tal como se utiliza en la presente, la expresión "administrar al oído interno" se refiere a proporcionar o administrar un agente terapéutico descrito en la presente a un sujeto por cualquier vía que permita la transducción de las células del oído interno. Las vías de ejemplo de administración al oído interno incluyen la administración en la perilinfa o endolinfa, tal como a o a través de la ventana oval, ventana redonda o canal semicircular (por ejemplo, canal horizontal), o por inyección transtimpánica o intratimpánica, por ejemplo, administración a una OHC.

Tal como se usa en la presente, el término "tipo de célula" se refiere a un grupo de células que comparten un fenotipo que es estadísticamente separable en base a los datos de expresión génica. Por ejemplo, las células de un tipo de célula común pueden compartir estructuras similares y/o características funcionales, tales como patrones similares de activación de genes y perfiles de presentación de antígenos. Las células de un tipo de célula común pueden incluir aquellas que son aisladas de un tejido común (por ejemplo, tejido epitelial, tejido neural, tejido conectivo o tejido muscular) y/o aquellas que son aisladas de un órgano común, sistema de tejido, vaso sanguíneo u otra estructura y/o región en un organismo.

Tal como se usa en la presente, el término "célula ciliada coclear" se refiere a un grupo de células especializadas en el oído interno que están involucradas en la detección del sonido. Hay dos tipos de células ciliadas cocleares: células ciliadas internas y células ciliadas externas. El daño a las células ciliadas cocleares y las mutaciones genéticas que alteran la función de las células ciliadas cocleares están implicados en la pérdida auditiva y la sordera.

Tal como se usa en la presente, los términos "mutación conservadora", "sustitución conservadora" y "sustitución conservadora de aminoácidos" se refieren a una sustitución de uno o más aminoácidos por uno o más aminoácidos diferentes que exhiben propiedades fisicoquímicas similares, tales como polaridad, carga electrostática y volumen estérico. Estas propiedades se resumen para cada uno de los veinte aminoácidos naturales en la tabla 1 a continuación.

Tabla 1. Propiedades fisicoquímicas representativas de los aminoácidos naturales

Aminoácido	Código de 3 letras	Código de 1 letra	Polaridad de cadena lateral	Carácter electrostático a pH fisiológico (7.4)	Volumen estérico [†]
Alanina	Ala	A	no polar	neutro	pequeño
Arginina	Arg	R	polar	catiónico	grande
Asparagina	Asn	N	polar	neutro	intermedio

Aminoácido	Código de 3 letras	Código de 1 letra	Polaridad de cadena lateral	Carácter electrostático a pH fisiológico (7.4)	Volumen estérico[†]
Ácido aspártico	Asp	D	polar	aniónico	intermedio
Cisteína	Cys	C	no polar	neutro	intermedio
Ácido glutámico	Glu	E	polar	aniónico	intermedio
Glutamina	Gln	Q	polar	neutro	intermedio
Glicina	Gly	G	no polar	neutro	pequeño
Histidina	His	H	polar	Ambas formas neutra y catiónica en equilibrio a pH 7.4	grande
Isoleucina	Ile	I	no polar	neutro	grande
Leucina	Leu	L	no polar	neutro	grande
Lisina	Lys	K	polar	catiónico	grande
Metionina	Met	M	no polar	neutro	grande
Fenilalanina	Phe	F	no polar	neutro	grande
Prolina	Pro	P	no polar	neutro	intermedio
Serina	Ser	S	polar	neutro	pequeño
Treonina	Thr	T	polar	neutro	intermedio
Triptófano	Trp	W	no polar	neutro	voluminoso
Tirosina	Tyr	Y	polar	neutro	grande
Valina	Val	V	no polar	neutro	intermedio
[†] según el volumen en Å ³ : 50-100 es pequeño, 100-150 es intermedio, 150-200 es grande y >200 es voluminoso					

De esta tabla se aprecia que las familias conservadoras de aminoácidos incluyen (i) G, A, V, L e I; (ii) D y E; (iii) C, S y T; (iv) H, K y R; (v) N y Q; y (vi) F, Y y W. Por lo tanto, una mutación o sustitución conservadora es una que sustituye un aminoácido por un miembro de la misma familia de aminoácidos (por ejemplo, una sustitución de Ser por Thr o Lys por Arg).

Tal como se utiliza en la presente, el término "secuencia señal de degradación" se refiere a una

secuencia (por ejemplo, una secuencia de nucleótidos que puede traducirse en una secuencia de aminoácidos) que media en la degradación de un polipéptido en el que está contenida. Las secuencias señal de degradación pueden incluirse en los vectores de ácido nucleico de la invención para reducir o impedir la expresión de partes de proteínas de estereocilina que no hayan sufrido recombinación y/o empalme.

El término "deriva" y "derivado" tal como se usa en la presente se refieren a un ácido nucleico, péptido o proteína o una variante o análogo de estos que comprende una o más mutaciones y/o modificaciones químicas en comparación con un ácido nucleico, péptido o proteína de tipo salvaje de longitud completa correspondiente. Los ejemplos no taxativos de modificaciones químicas que implican ácidos nucleicos incluyen, por ejemplo, modificaciones del resto de base, resto de azúcar, resto de fosfato, estructura de fosfato-azúcar o una combinación de estos.

Tal como se usa en la presente, los términos "cantidad eficaz", "cantidad terapéuticamente eficaz" y "cantidad suficiente" de una composición, construcción de vector o vector viral descrito en este documento se refieren a una cantidad suficiente para, cuando se administra al sujeto, que incluye una mamífero, por ejemplo un ser humano, produce resultados beneficiosos o deseados, incluidos los resultados clínicos y, como tal, una "cantidad eficaz" o sinónimo de los mismos depende del contexto en el que se está aplicando. Por ejemplo, en el contexto del tratamiento de la pérdida auditiva neurosensorial, es una cantidad de la composición, construcción de vector o vector viral suficiente para lograr una respuesta al tratamiento en comparación con la respuesta obtenida sin la administración de la composición, construcción de vector o vector viral. La cantidad de una composición dada descrita en la presente que corresponderá a dicha cantidad variará dependiendo de varios factores, tal como el agente dado, la formulación farmacéutica, la vía de administración, el tipo de enfermedad o trastorno, la identidad del sujeto (por ejemplo, edad, sexo, peso) o el hospedador que se está tratando, y similares, pero, no obstante, un experto en la técnica puede determinarlo de forma rutinaria. Además, tal como se usa en la presente, una "cantidad terapéuticamente eficaz" de una composición, construcción de vector o vector viral de la presente divulgación es una cantidad que da como resultado un resultado beneficioso o deseado en un sujeto en comparación con un control. Como se define en este documento, una cantidad terapéuticamente eficaz de una composición, construcción de vector o vector viral de la presente divulgación puede ser determinada fácilmente por un experto en la materia mediante métodos de rutina conocidos en la técnica. El régimen de dosificación se puede ajustar para proporcionar la respuesta terapéutica óptima.

Tal como se usa en la presente, el término "endógeno" se refiere a una molécula (por ejemplo, un polipéptido, ácido nucleico o cofactor) que se encuentra naturalmente en un organismo particular (por ejemplo, un ser humano) o en una ubicación particular dentro de un organismo (por ejemplo,

un órgano, un tejido o una célula, tal como una célula humana, por ejemplo, una OHC).

Tal como se usa en la presente, el término "expresa" se refiere a uno o más de los siguientes eventos: (1) producción de una plantilla de ARN a partir de una secuencia de ADN (por ejemplo, mediante transcripción); (2) procesamiento de una transcripción de ARN (por ejemplo, mediante empalme, edición, formación de casquete 5' y/o procesamiento de extremo 3'); (3) traducción de un ARN en un polipéptido o proteína; y (4) modificación postraduccional de un polipéptido o proteína.

Tal como se usa en la presente, el término "exógeno" describe una molécula (por ejemplo, un polipéptido, ácido nucleico o cofactor) que no se encuentra naturalmente en un organismo particular (por ejemplo, un ser humano) o en una ubicación particular dentro de un organismo (por ejemplo, un órgano, un tejido o una célula, tal como una célula humana, por ejemplo, una OHC humana). Los materiales exógenos incluyen aquellos que se proporcionan de una fuente externa a un organismo o a la materia cultivada extraída de allí.

Como se usa en este documento, el término "exón" se refiere a una región dentro de la región codificante de un gen, cuya secuencia de nucleótidos determina la secuencia de aminoácidos de la proteína correspondiente. El término exón también se refiere a la región correspondiente del ARN transcrito de un gen. Los exones se transcriben en pre-ARNm y pueden incluirse en el ARNm maduro dependiendo del corte y empalme alternativo del gen. Los exones que se incluyen en el ARNm maduro después del procesamiento se traducen en proteína, en donde la secuencia del exón determina la composición de aminoácidos de la proteína.

Tal como se usa en la presente, el término "heterólogo" se refiere a una combinación de elementos que no se encuentran en la naturaleza. Por ejemplo, un transgén heterólogo se refiere a un transgén que no es expresado naturalmente por el promotor al que está unido de forma operativa.

Tal como se usa en la presente, los términos "creciente" y "decreciente" se refieren a la modulación que da como resultado, respectivamente, cantidades mayores o menores de función, expresión o actividad de una métrica con respecto a una referencia. Por ejemplo, después de la administración de una composición en un método descrito en este documento, la cantidad de un marcador de una métrica (por ejemplo, expresión transgénica) como se describe en este documento puede aumentarse o disminuirse en un sujeto en al menos 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 98 % o más en relación con la cantidad de marcador antes de la administración. Generalmente, la métrica se mide después de la administración en el momento en que la administración ha tenido el efecto mencionado, por ejemplo, al menos una semana, un mes, 3 meses o 6 meses, después de que ha comenzado un régimen de tratamiento.

Tal como se usa en la presente, el término "inteína", también denominado "intrón proteico", se

refiere a una parte de una proteína que tiene típicamente 100-900 residuos de aminoácidos de largo y que es capaz de autoextirparse y ligar los fragmentos proteicos flanqueantes ("exteínas") con un enlace peptídico. Las inteínas se producen durante el empalme de proteínas. El término "inteína" engloba cuatro clases diferentes de inteínas, que incluye la maxi-inteína, la mini-inteína, la inteína de transempalme y la inteína de alanina. Las maxi-inteínas se refieren a las regiones de empalme de extremo N y C de una proteína que contiene un dominio de endonucleasa. Los dominios de endonucleasa, también conocidos como "genes de endonucleasa homing" o "HEG" se refieren a una clase de endonucleasas codificadas como genes independientes dentro de intrones, como fusiones de proteínas con otras proteínas o como inteínas autoempalmables. Por lo general, las HEG hidrolizan muy pocas y a menudo dirigidas regiones de ADN. Una vez que una HEG hidroliza un fragmento de ADN, el gen que codifica la HEG suele incorporarse al sitio de escisión, aumentando así su frecuencia alélica. Las mini-inteínas se refieren a los dominios de empalme de extremo N y C que carecen el dominio de endonucleasa. Las inteínas de transempalme se refieren a inteínas que se dividen en dos o más dominios que, a su vez, se dividen en los extremos N y extremos C. Las inteínas de alanina se refieren a inteínas que tienen una unión de empalme de una alanina en lugar de una cisteína o serina. Una inteína de una proteína precursora puede venir en dos genes; en tales casos, la inteína se designa como "inteína" dividida.

Tal como se usa en la presente, el término "intrón" se refiere a una región dentro de la región codificante de un gen, cuya secuencia de nucleótidos no se traduce en la secuencia de aminoácidos de la proteína correspondiente. El término intrón también se refiere a la región correspondiente del ARN transcrito de un gen. Los intrones se transcriben en pre-ARNm, pero se eliminan durante el procesamiento y no se incluyen en el ARNm maduro.

Tal como se usa en la presente, el término "expresión específica de las células ciliadas externas" o "expresión específica de las OHC" se refiere a la producción de una transcripción de ARN o polipéptido principalmente en las OHC cocleares en comparación con otros tipos de células de la cóclea (por ejemplo, neuronas ganglionares espirales, glía u otros tipos de células cocleares). La expresión específica de OHC de un transgén puede confirmarse comparando la expresión del transgén (por ejemplo, expresión de ARN o proteína) entre varios tipos celulares de la cóclea (por ejemplo, OHC frente a no OHC) mediante cualquier técnica estándar (por ejemplo, RT PCR cuantitativa, inmunohistoquímica, análisis de transferencia Western o medición de la fluorescencia de un indicador (por ejemplo, GFP) unida de forma operativa a un promotor). Un promotor específico de OHC induce la expresión (por ejemplo, expresión de ARN o proteína) de un transgén al que está unido de forma operativa que es al menos 50 % superior (por ejemplo, 50 %, 75 %, 100 %, 125 %, 150 %, 175 %, 200 % superior o más) en OHC en comparación con al menos 2 (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) de los siguientes tipos de células del oído interno: células

ciliadas internas, células límite, células de la falange interna, células del pilar interno, células del pilar externo, células de Deiter de la primera fila, células de Deiter de la segunda fila, células de Deiter de la tercera fila, células de Hensen, células de Claudius, células del surco interno, células del surco externo, células de prominencia en espiral, células radicales, células interdentes, células basales de la estría vascular, células intermedias de la estría vascular, células marginales de la estría vascular, neuronas del ganglio espiral, células de Schwann. Un promotor específico de OHC induce la expresión (por ejemplo, expresión de ARN o proteína) de un transgén al que está unido en forma operativa que es al menos un 50 % mayor (por ejemplo, 50 %, 75 %, 100 %, 125 %, 150 %, 175 %, 200 % mayor o más) en las OHC de la cóclea en comparación con otras células de la cóclea.

Tal como se usa en la presente, "localmente" o "administración local" significa la administración en un sitio particular del cuerpo destinado a un efecto local y no a un efecto sistémico. Ejemplos de administración local son administración epicutánea, inhalatoria, intraarticular, intratecal, intravaginal, intravítrea, intrauterina, intralesional, administración en ganglios linfáticos, administración intratumoral, administración en el oído interno y administración en una membrana mucosa del sujeto, donde la administración está destinada a tener un efecto local y no uno sistémico.

Tal como se usa en la presente, el término "unido/a de forma operativa" hace referencia a una primera molécula unida a una segunda molécula, en donde las moléculas están dispuestas de manera que la primera molécula afecta la función de la segunda molécula. Las dos moléculas pueden ser o no parte de una única molécula contigua y pueden ser adyacentes o no. Por ejemplo, un promotor está unido de forma operativa a una molécula de polinucleótido transcribible si el promotor modula la transcripción de la molécula de polinucleótido transcribible de interés en una célula. Además, dos porciones de un elemento regulador de la transcripción están operativamente unidas entre sí si están unidas de modo que la funcionalidad de activación de la transcripción de una parte no se vea afectada adversamente por la presencia de la otra parte. Dos elementos reguladores de la transcripción pueden unirse de forma operativa entre sí por medio de un polinucleótido enlazador (por ejemplo, un polinucleótido no codificante intermedio) o pueden unirse de forma operativa entre sí sin nucleótidos intermedios presentes.

Tal como se usa en la presente, el término "plásmido" se refiere a una molécula de ADN de cadena doble circular extracromosómico en la que se pueden ligar segmentos de ADN adicionales. Un plásmido es un tipo de vector, una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Determinados plásmidos pueden replicarse de forma autónoma en una célula hospedadora en la que se introducen (por ejemplo, plásmidos bacterianos que tienen un origen bacteriano de replicación y plásmidos episómicos de mamíferos). Otros vectores (por

ejemplo, vectores mamíferos no episomales) se pueden integrar al genoma de una célula hospedadora luego de introducirse en ella y, por lo tanto, se replica junto con el genoma hospedador. Determinados plásmidos son capaces de dirigir la expresión de genes a los que se enlazan de manera operativa.

5 Como se usa en este documento, el término "polinucleótido" se refiere a un polímero de nucleósidos. Típicamente, un polinucleótido se compone de nucleósidos que se encuentran naturalmente en ADN o ARN (por ejemplo, adenosina, timidina, guanosina, citidina, uridina, desoxiadenosina, desoxitimidina, desoxiguanosina y desoxicitidina) unidos por enlaces fosfodiéster. El término abarca moléculas que comprenden nucleósidos o análogos de nucleósidos
10 que contienen bases modificadas química o biológicamente, cadenas principales modificadas, etc., se encuentren o no en ácidos nucleicos de origen natural, y tales moléculas pueden ser preferidas para ciertas aplicaciones. En los casos en los que la presente solicitud hace referencia a un polinucleótido, se entiende que se proveen tanto ADN y ARN, y en cada caso, se proveen formas de cadena sencilla o de cadena doble (y complementos de cada molécula de cadena simple). Tal
15 como se usa en la presente, "secuencia de polinucleótidos" puede hacer referencia al material polinucleótido en sí mismo y/o a la información de secuencia (es decir, la sucesión de letras utilizadas como abreviaturas de bases) que caracterizan bioquímicamente un ácido nucleico específico. Una secuencia de polinucleótidos presentada en la presente se presenta en una dirección 5' a 3', a menos que se indique lo contrario.

20 Tal como se usa en la presente, el término "promotor" se refiere a un sitio de reconocimiento en el ADN que está unido por una ARN polimerasa. La polimerasa impulsa la transcripción del transgén. Un promotor representativo de la divulgación es el promotor de oncomodulina (OCM), tal como un promotor de OCM que tiene una secuencia de ácido nucleico de cualquiera de las SEQ ID NO: 1-3 o una variante de estas que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia
25 con la secuencia de ácidos nucleicos de cualquiera de las SEQ ID NO: 1-3.

El "porcentaje (%) de identidad de secuencia de aminoácidos" con respecto a una secuencia de polipéptidos o polinucleótidos de referencia se define como el porcentaje de ácidos nucleicos o aminoácidos en una secuencia candidata que son idénticos a los ácidos nucleicos o aminoácidos
30 en la secuencia de polipéptidos o polinucleótidos de referencia, después de alinear las secuencias e introducir espacios, si fuera necesario, para lograr el porcentaje máximo de identidad de secuencia. La alineación a efectos de determinar el porcentaje de identidad de secuencia de aminoácidos o ácidos nucleicos puede lograrse de diversas maneras que se encuentran dentro de las capacidades del experto en la técnica, por ejemplo, mediante el uso de software informático
35 disponible para el público, tal como el software BLAST, BLAST-2, o Megalign. Los expertos en la

técnica pueden determinar parámetros adecuados para alinear las secuencias, lo que incluye cualquier algoritmo necesario para lograr la alineación máxima en la longitud completa de las secuencias que se comparan. Por ejemplo, los valores porcentuales de identidad de secuencia se pueden generar mediante el uso del programa informático de comparación de secuencias BLAST.

A modo de ilustración, el porcentaje de identidad de secuencia de una secuencia de aminoácidos o ácidos nucleicos A dada, para, con o contra una secuencia de aminoácidos o ácidos nucleicos B dada, (que se puede expresar de forma alternativa como una secuencia de aminoácidos o ácidos nucleicos A dada, que tiene un determinado porcentaje de identidad de secuencia para, con o contra una secuencia de aminoácidos o ácidos nucleicos B dada) se calcula como se indica a continuación:

$$100 \text{ multiplicado por (la fracción } X/Y)$$

donde X es la cantidad de nucleótidos o aminoácidos puntuados como emparejamientos idénticos por un programa de alineación de secuencias (por ejemplo, BLAST) en la alineación de este programa de A y B, y donde Y es la cantidad total de ácidos nucleicos en B. Se entenderá que cuando la longitud de la secuencia de ácido nucleico o de aminoácidos A no es igual a la longitud de la secuencia de ácido nucleico o aminoácidos B, el porcentaje de identidad de la secuencia de A respecto a B no será igual al porcentaje de identidad de secuencia de B respecto a A.

Tal como se usa en la presente, el término "composición farmacéutica" se refiere a una mezcla que contiene un agente terapéutico, opcionalmente en combinación con uno o más excipientes, diluyentes, y/o portadores farmacéuticamente aceptables, para ser administrados a un sujeto, tal como un mamífero, por ejemplo, un ser humano, con el fin de prevenir, tratar o controlar una enfermedad o afección particular que afecte o pueda afectar al sujeto.

Tal como se usa en la presente, la frase "farmacéuticamente aceptable" hace referencia a los compuestos, materiales, composiciones y/o formas de dosificación que son adecuados para ponerse en contacto con los tejidos de un sujeto, tal como un mamífero (por ejemplo, un ser humano) sin toxicidad excesiva, irritación, respuesta alérgica y otros problemas o complicaciones, con una relación beneficio/riesgo correspondiente razonable.

Tal como se usa en la presente, el término "región recombinogénica" se refiere a una región de homología que media la recombinación entre dos secuencias diferentes.

Tal como se usa en la presente, el término "secuencia reguladora" incluye promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación) que controlan la transcripción o traducción de los polinucleótidos que codifican STRC. Tales secuencias reguladoras se describen en, por ejemplo, Goeddel; Gene Expression Technology: Methods in Enzymology 185 (Academic Press, San Diego, CA, 1990); que se incorpora a la presente mediante referencia.

Tal como se usa en la presente, el término "muestra" se refiere a una muestra (por ejemplo, sangre, componente sanguíneo (por ejemplo, suero o plasma), orina, saliva, líquido amniótico, líquido cefalorraquídeo, tejido (por ejemplo, placentario o dérmico), líquido pancreático, muestra de vellosidades coriónicas y células) aisladas de un sujeto.

En la presente, los términos "estereocilina" y "STRC" (también conocido como DFNB16) se refieren a una proteína codificada por el gen STRC y al gen que codifica esta proteína, respectivamente. En humanos, STRC está duplicado en tándem, donde la segunda copia contiene un codón de parada prematuro en el exón 20, produciendo así un pseudogén STRC. En el contexto de la presente divulgación, STRC no se refiere al pseudogén STRC. Estudios anteriores han identificado mutaciones en la copia de longitud completa de STRC en pacientes humanos con pérdida auditiva neurosensorial autosómica recesiva no sindrómica (Verpy et ál., Nat. Genet. 29:345-9 (2001)). La expresión de la proteína estereocilina se limita a los estereocilios de los haces ciliados de las células ciliadas. Se cree que la estereocilina forma conectores superiores horizontales y coronas tectoriales de unión a la membrana, necesarios para el funcionamiento normal del aparato auditivo (Avan et ál., PNAS 116:25948-57 (2019); Verpy et ál., J. Comp. Neurol. 519:194-210 (2011)). Se ha demostrado que los ratones que carecen de estereocilina presentan haces de células ciliadas anormales con cohesión defectuosa y problemas de audición (Verpy et ál., Nature 456:255-8 (2008)). La presente divulgación proporciona polinucleótidos que codifican la proteína estereocilina de longitud completa, que, cuando se incorporan a los sistemas de vectores descritos en la presente, pueden utilizarse como agente terapéutico para el tratamiento de la pérdida auditiva (por ejemplo, pérdida auditiva neurosensorial) en sujetos que lo necesiten. Los términos "estereocilina" y "STRC" también se refieren a variantes de la proteína estereocilina de tipo silvestre y ácidos nucleicos que codifican la misma, respectivamente, tales como proteínas variantes que tienen al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o 99,9 % de identidad, o más) con la secuencia de aminoácidos de una proteína estereocilina de tipo salvaje (por ejemplo, SEQ ID NO: 4 o 5) o polinucleótidos que tengan al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 99,9 % de identidad, o más) con la secuencia de ácido nucleico de un gen STRC de tipo salvaje (por ejemplo, SEQ ID NO: 6 o 7), siempre que el análogo de STRC codificado conserve la función terapéutica de STRC de tipo salvaje.

Tal como se usa en la presente, el término "elemento regulador de la transcripción" se refiere a un polinucleótido que controla, al menos en parte, la transcripción de un gen de interés. Los elementos reguladores de la transcripción pueden incluir promotores, potenciadores y otros polinucleótidos (por ejemplo, señales de poliadenilación) que controlan o ayudan a controlar la

transcripción génica. Se describen ejemplos de elementos reguladores de la transcripción, por ejemplo, en Lorence, Recombinant Gene Expression: Reviews and Protocols (Humana Press, Nueva York, NY, 2012).

Como se usa en este documento, los términos "sujeto" y "paciente" se refieren a un animal (por ejemplo, un mamífero, como un ser humano). Un sujeto a ser tratado de acuerdo con los métodos descritos en la presente puede ser uno a quien se le ha diagnosticado pérdida auditiva (por ejemplo, pérdida auditiva asociada con una mutación en STRC), o uno en riesgo de desarrollar esta afección. El diagnóstico puede llevarse a cabo mediante cualquier método o técnica conocido en la técnica. Un experto en la técnica entenderá que un sujeto a tratar de acuerdo con la presente descripción puede haber sido sometido a pruebas estándar o puede haber sido identificado, sin exámenes, como en riesgo debido a la presencia de uno o más factores de riesgo asociados con la enfermedad o afección.

Tal como se usa en la presente, los términos "transducción" y "transducir" se refieren a un método para introducir una construcción de vector o una parte de esta en una célula. Cuando la construcción del vector está contenida en un vector viral tal como, por ejemplo, un vector de AAV, la transducción se refiere a la infección viral de la célula y la posterior transferencia e integración de la construcción del vector o parte de esta en el genoma celular.

Como se usa en este documento, "tratamiento" y "tratar" en referencia a una enfermedad o afección, se refieren a un enfoque para obtener resultados beneficiosos o deseados, por ejemplo, resultados clínicos. Los resultados beneficiosos o deseados pueden incluir, de modo no taxativo, el alivio o la mejora de uno o más síntomas o afecciones; la disminución de la extensión de la enfermedad o afección; la estabilización (es decir, el no empeoramiento) del estado de la enfermedad, trastorno o afección; la prevención de la propagación de la enfermedad o afección; el retraso o ralentización del progreso de la enfermedad o afección; la mejora o paliación de la enfermedad o afección; y la remisión (ya sea parcial o total), ya sea detectable o indetectable. "Mejorar" o "paliar" una enfermedad o afección significa que la extensión y/o las manifestaciones clínicas indeseables de la enfermedad, trastorno o afección se reducen y/o el curso temporal de la progresión se ralentiza o alarga, en comparación con la extensión o el evolución temporal en ausencia de tratamiento. El "tratamiento" también puede significar prolongar la supervivencia en comparación con la supervivencia esperada de no recibirse el tratamiento. Aquellos que necesitan el tratamiento incluye aquellos que ya presentan la afección o el trastorno, así como aquellos propensos a presentar la afección o el trastorno o aquellos en los que la afección o el trastorno ha de prevenirse.

Como se usa en este documento, el término "vector" se refiere a un vector de ácido nucleico, por ejemplo, un vector de ADN, como un plásmido, cósmido o cromosoma artificial, un vector de ARN,

un virus o cualquier otro replicón adecuado (por ejemplo, vector viral). Se han desarrollado diversos vectores para el suministro de polinucleótidos que codifican proteínas exógenas en una célula procariota o eucariota. Ejemplos de tales vectores de expresión se describen en, por ejemplo, Gellissen, *Production of Recombinant Proteins: Novel Microbial and Eukaryotic Expression Systems* (John Wiley & Hijos, Marblehead, MA, 2006). Los vectores de expresión adecuados para usar con las composiciones y métodos descritos en la presente contienen una secuencia de polinucleótidos así como, por ejemplo, elementos de secuencia adicionales usados para la expresión de proteínas y/o la integración de estas secuencias de polinucleótidos en el genoma de una célula de mamífero. Ciertos vectores que pueden usarse para la expresión de STRC tal como se describe en la presente incluyen vectores que contienen secuencias reguladoras, tales como regiones promotoras y potenciadoras, que dirigen la transcripción génica. Otros vectores útiles para la expresión de STRC contienen secuencias de polinucleótidos que mejoran la velocidad de traducción de STRC o mejoran la estabilidad o exportación nuclear del ARNm que resulta de la transcripción génica. Estos elementos de secuencia incluyen, por ejemplo, regiones no traducidas 5' y 3' y un sitio de señal de poliadenilación para dirigir la transcripción eficaz del gen transportado en el vector de expresión. Los vectores de expresión adecuados para usar con las composiciones y métodos descritos en la presente también pueden contener un polinucleótido que codifica un marcador para la selección de células que contienen dicho vector. Los ejemplos de un marcador adecuado incluyen genes que codifican resistencia a antibióticos, tales como ampicilina, cloranfenicol, kanamicina o nourseotricina. Tal como se usa en la presente, el término "de tipo salvaje" se refiere a un genotipo con la frecuencia más alta para un gen particular en un organismo dado.

Breve descripción de las figuras

Las Figuras 1A-1B son una serie de imágenes fluorescentes de cócleas de ratón transducidas con un vector de virus adenoasociado (AAV) que expresa proteína verde fluorescente (GFP) bajo el control del promotor ubicuo del citomegalovirus (CMV) (Figura 1A), o un vector AAV que expresa GFP bajo el control de un promotor de oncomodulina (OCM) (SEQ ID NO: 1; Figura 1B). Se muestra la fluorescencia de GFP nativa. Mediante el uso de un promotor ubicuo, el AAV-CMV-GFP indujo la expresión de GFP en muchos tipos celulares de la cóclea, incluidas las células ciliadas internas (IHC), las células ciliadas externas (OHC), las neuronas ganglionares espirales, las células mesenquimales y la glía (Figura 1A). Utilizando un promotor específico de OHC, AAV-OCM (SEQ ID NO: 1)-GFP indujo la expresión de GFP exclusivamente en OHC (Figura 1B).

Las Figuras 2A y 2B son una serie de micrografías de secciones de parafina individuales de una vuelta basal de cóclea de dos primates no humanos (*Macaca fascicularis*) a los que se administró

un vector AAV que expresa H2B-GFP bajo el control del promotor OCM de la SEQ ID NO: 1. La Figura 2A es una micrografía de una sección de parafina de un primer animal y la Figura 2B es una micrografía de una sección de parafina de un segundo animal. El panel A de la Figura 2A y el panel B de la Figura 2B (imágenes superiores) muestran una conversión en escala de grises de la zona alrededor del órgano de Corti con los núcleos teñidos con hematoxilina originalmente en azul y el anticuerpo H2B-GFP teñido originalmente en rojo. El panel A' de la Figura 2A y el panel B' de la Figura 2B (imágenes inferiores) muestran la señal restante tras eliminar la señal de hematoxilina azul; los núcleos H2B-GFP positivos (rojos) siguen siendo visibles como colores más oscuros tras la conversión a escala de grises. Las barras de escala representan 100 μm . Las células ciliadas internas (IHC) y las células ciliadas externas (OHC) están resaltadas a modo de orientación.

Las Figuras 3A-3C son una serie de imágenes fluorescentes de la expresión de la estereocilina en el órgano de Corti del ratón en un ROI de 200 μm^2 a 16 kHz. La Figura 3A muestra la tinción del anticuerpo contra la estereocilina en las puntas de los estereocilios de las células ciliadas externas (OHC) en un ratón CBA/CaJ de tipo salvaje. Tal como se muestra en la Figura 3B, los animales STRC inactivados (KO) de 232 pb carecían de señal para el anticuerpo. La Figura 3C muestra la tinción del anticuerpo de estereocilina en un ratón KO de STRC de 232 pb al que se le administraron vectores Anc80 duales, en los que el primer vector llevaba un promotor CMV y los nucleótidos 1-3200 del ADNc de STRC murino y el segundo vector llevaba los nucleótidos 2201-5430, creando una superposición de 1000 pb entre los dos ADNc en los dos vectores. La expresión de la proteína estereocilina de-novo pudo observarse en las puntas de los estereocilios OHC y en el cuerpo de las células ciliadas internas del órgano de Corti en ratones KO de STRC tratados de 232 pb.

Las Figuras 4A-4C son una serie de gráficos que muestran la mejora de la función auditiva en ratones KO de STRC de 232 pb tratados con Anc80-CMV-mStrc y su correlación con la expresión de OHC STRC. Los oídos contralaterales no tratados mostraron DPOAE casi ausentes y umbrales ABR muy elevados, indicativos de la pérdida de la función OHC (Figuras 4A-4B, círculos abiertos), mientras que los animales con KO de STRC de 232 pb tratados mostraron una recuperación de los umbrales auditivos (Figuras 4A-4B, círculos rellenos). El mejor respondedor de los animales tratados (Figuras 4A-4B, cuadrados negros) mostró umbrales de audición cercanos al tipo salvaje (Figuras 4A-4B, triángulos). Se observó que una fracción elevada de las OHC de ratones KO de STRC de 232 pb que expresaban estereocilina tras el tratamiento con AAV-Anc80-CMV-mStrc favorecía la recuperación de la audición (Figura 4C).

Las Figuras 5A-5B son una imagen y un gráfico que muestran que la transfección de células HEK293T con un sistema de inteína dividida en dos vectores condujo a la reconstitución de la estereocilina de longitud completa. La Figura 5A es una imagen representativa de un análisis de

transferencia Western contra la proteína estereocilina y beta-actina. La Figura 5B es una cuantificación por densitometría de la intensidad de la banda de estereocilina de longitud completa en relación con la actina e indica la expresión relativa de la proteína estereocilina de longitud completa para el control negativo (GFP), el control positivo de longitud completa y la construcción de inteína Npu.

Las Figuras 6A-6B son mapas de los plásmidos P959 y P724, respectivamente, utilizados para crear un sistema de vector dual superpuesto para expresar estereocilina (STRC) bajo el control de un promotor murino de OCM.

Las Figuras 7A-7B son mapas de los plásmidos P960 y P726, respectivamente, utilizados para crear un sistema de vector híbrido dual para expresar STRC bajo el control de un promotor murino de OCM.

Descripción detallada

En la presente se describen composiciones y métodos para el tratamiento de la pérdida auditiva neurosensorial en un sujeto (tal como un sujeto mamífero, por ejemplo, un ser humano) mediante la administración de un primer vector de ácido nucleico que contiene un promotor, tal como un promotor de oncomodulina (OCM), y un polinucleótido que codifica una parte de extremo N de una proteína estereocilina (STRC) (por ejemplo, una proteína STRC de tipo salvaje (WT)) y un segundo vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo C de una proteína STRC y una secuencia de poliadenilación (poli(A)). Cuando se introducen en una célula de mamífero, tal como una célula ciliada externa coclear (OHC), los polinucleótidos codificados por los dos vectores de ácido nucleico pueden combinarse para formar un polinucleótido que codifica la proteína de la STRC de longitud completa. La divulgación también presenta sistemas de expresión de dos vectores (por ejemplo, vectores duales superpuestos, vectores de transempalme, vectores híbridos duales y vectores de transempalme de inteína dividida) que contienen los polinucleótidos antes mencionados. Las composiciones y métodos descritos en la presente pueden utilizarse para expresar polinucleótidos que codifican STRC específicamente en las OHC y, por lo tanto, las composiciones descritas en la presente pueden administrarse a un sujeto (tal como un sujeto mamífero, por ejemplo, un ser humano) para tratar trastornos causados por disfunción de las OHC, tales como pérdida auditiva (por ejemplo, pérdida auditiva neurosensorial) y neuropatía auditiva.

Estereocilina

La estereocilina (también conocida como DFNB16) es una proteína de gran tamaño codificada por el gen STRC del cromosoma 15q15, que contiene 29 exones que abarcan aproximadamente 19 kb

del genoma. En gen STRC está duplicado en tándem, donde la segunda copia contiene un codón de parada prematuro en el exón 20, produciendo así un pseudogén STRC. Estudios anteriores han identificado dos mutaciones de cambio de marco y una gran eliminación en la copia de longitud completa de STRC en dos familias con pérdida auditiva neurosensorial autosómica recesiva no

5 sindrómica (Verpy et ál., Nat. Genet. 29:345-9 (2001)). La expresión de la proteína estereocilina está limitada a la estereocilina en haces ciliados de las células ciliadas del oído interno y se cree que forma conectores superiores horizontales y coronas tectoriales de unión a la membrana, necesarios para el funcionamiento normal del aparato auditivo (Avan et ál., PNAS 116:25948-57 (2019); Verpy et ál., J. Comp. Neurol. 519:194-210 (2011)). Se ha demostrado que los ratones que

10 carecen de estereocilina presentan haces de células ciliadas anormales con cohesión defectuosa y problemas de audición (Verpy et ál., Nature 456:255-8 (2008)).

Las composiciones y métodos descritos en la presente se pueden usar para tratar la pérdida auditiva neurosensorial mediante la administración de un primer vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo N de una proteína estereocilina y un

15 segundo vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo C de una proteína estereocilina. La secuencia codificante de STRC de longitud completa es demasiado grande para incluirla en el tipo de vector que se usa comúnmente para la terapia génica (por ejemplo, un vector de virus adenoasociado (AAV), que se cree que tiene un límite de empaquetamiento de 5 kb). Las composiciones y métodos descritos en la presente superan este

20 problema al dividir la secuencia codificante STRC entre dos vectores de ácido nucleico diferentes, de forma que la secuencia STRC de longitud completa pueda reconstituirse en una célula. Estas composiciones y métodos se pueden usar para tratar sujetos que tienen una o más mutaciones en el gen STRC, por ejemplo, una mutación STRC que reduce la expresión de STRC, reduce la función STRC o está asociada con la pérdida auditiva (por ejemplo, Un sujeto que tiene DFNB16).

25 Cuando el primer y segundo vector de ácido nucleico se administran en una composición, los polinucleótidos que codifican las partes de extremo N y de extremo C de la estereocilina pueden combinarse dentro de una célula (por ejemplo, una célula humana, por ejemplo, una célula ciliada coclear) para formar un solo polinucleótido que contenga la secuencia codificante de STRC de longitud completa (por ejemplo, mediante recombinación homóloga y/o empalme).

30 Los vectores de ácido nucleico utilizados en las composiciones y métodos descritos en la presente incluyen secuencias de polinucleótidos que codifican estereocilina de tipo salvaje, o una variante de esta, tales como secuencias de polinucleótidos que, cuando se combinan, codifican una proteína que tiene al menos un 85 % de identidad de secuencia (por ejemplo, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia) con la secuencia de aminoácidos de estereocilina de mamífero de tipo salvaje (por ejemplo, humano o ratón). Los polinucleótidos

35

utilizados en los vectores de ácido nucleico descritos en la presente codifican una parte de extremo N y una parte de extremo C de una secuencia de aminoácidos de estereocilina de la Tabla 2 siguiente (por ejemplo, dos partes que, cuando se combinan, codifican una secuencia de aminoácidos de estereocilina de longitud completa de la Tabla 2, por ejemplo, la SEQ ID NO: 4 o la SEQ ID NO: 5).

De acuerdo con los métodos descritos en la presente, a un sujeto se le puede administrar una composición que contiene un primer vector de ácido nucleico y un segundo vector de ácido nucleico que contiene una parte de extremo N y extremo C, respectivamente, de una secuencia de polinucleótidos que codifica la secuencia de aminoácidos de la SEQ ID NO: 4 o la SEQ ID NO: 5, o una secuencia de polinucleótidos que codifica una secuencia de aminoácidos que tiene al menos 85 % de identidad de secuencia (por ejemplo, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más, identidad de secuencia) con la secuencia de aminoácidos de la SEQ ID NO: 4 o la SEQ ID NO: 5, o una secuencia de polinucleótidos que codifica una secuencia de aminoácidos que contiene una o más sustituciones conservadoras de aminoácidos con la SEQ ID NO: 4 o la SEQ ID NO: 5 (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más sustituciones conservadoras de aminoácidos), siempre que el análogo de estereocilina codificado retenga la función terapéutica de STRC de tipo salvaje. En algunas modalidades, no más del 10 % de los aminoácidos en la parte de extremo N de la proteína estereocilina y no más del 10 % de los aminoácidos en la parte de extremo C de la proteína estereocilina pueden ser reemplazados con sustituciones conservadoras de aminoácidos. La proteína estereocilina puede estar codificada por un polinucleótido que tiene la secuencia de la SEQ ID NO:5 o la SEQ ID NO: 6. La proteína estereocilina también puede estar codificada por un polinucleótido que tenga variantes de un solo nucleótido (SNV) que se ha hallado que no son patogénicas en sujetos humanos. La proteína estereocilina puede ser una proteína estereocilina humana o puede ser un homólogo de la proteína estereocilina humana de otra especie de mamífero (por ejemplo, ratón, rata, vaca, caballo, cabra, oveja, burro, gato, perro, conejo, conejillo de indias u otro mamífero).

Tabla 2. Secuencias de STRC

SEQ ID NO.	Nombre de secuencia	Secuencia
4	proteína estereocilina humana de tipo	MALSLWPLLLLLLLLLLSFAVTLAPTGPHSLDPGLSFLKSLLSTLD QAPQGSLSRSRFFFTFLANISSSFEPGRMGEGPVGEPPPLQPPAL RLHDFLVTLRGSPDWEPMGLLLGDMLALLGQEQTPrDFLVHQAG

SEQ ID NO.	Nombre de secuencia	Secuencia
5	salvaje, ID de UniProt: Q7RTU9	VLGGLVEVLLGALVPGGPPTPTRPPCTRDGPSCVLAADWLPSL LLLLEGTRWQALVQVQPSVDPTNATGLDGREAAAPHFLQGLLGLL TPTGELGSKEALWGGLLRTVGAPLYAAFQEGLLRVTHSLQDEVF SILGQPEPDTNGQCQGGNLQQLLLWGVHRNLSWDVQALGFLSG SPPPPPALLHCLSTGVPLPRASQPSAHISPRQRRRAITVEALCENHL GPAPPYSISNFSIHLLCQHTKPATPQPHPSTTAICQTAVWYAVSW 10 APGAQGWLQACHDQFPDEFDAICSNLSFSALSGSNRRLVKRLC AGLLPPPTSCPEGLPPVPLTPDIFWGCFLNETLWAERLCGEASL QAVPPSNQAWVQHVCQGPTPDVTASPPCHIGPCGERCPDGGSF LVMVCANDTMYEVLVPFWPWLQAGQCRISRGGNDTCFLEGLLGPL LPSLPPLGPSPLCLTPGPFLLGMLSQ LPRCQSSVPALAHPTRLHY 15 LLRLLTFLLGPGAGGAEAQGMLGRALLSSLPDNCSEFWDAFRPE GRRSVLRTIGEYLEQDEEQPTPSGFPTVNPSSGISKMELLACFS PVLWDLQREKSVWALQILVQAYLHMPPENLQQVLVSAEREAQ GFLTMLQGKLQGKLQVPPSEEQALGRLTALLQRYPRLTSQLFI DLSPLIPFLAVSDLMRFPSSLANDSVLAAIRDYSPGMRPEQKEAL 20 AKRLLAPELFGVPAWPQELLWAVLPLLPHLPLENFLQLSPHQIQ ALEDSPAAAGLPGHARHVLRLVNQSVQDGEEQVRRLGPLAC FLSPEELQSLVPLSDPTGPVERGLLECAANGTLSPEGRVAYELLG VLRSSGGAVLSPRELRVWAPLFSQLGLRFLQELSEPQLRAMLPVL QGTSVTPAQAVLLLGRLLPRHDLSEELCSLHLLLPGLSPQTLQAI 25 PRRVLVGACSCLAPELSRLSACQTAALLQTFRVKDGVKNMGTTG AGPAVCIPGQPIPTTWPDCLLPLLPLKLLQLDSLALLANRRRYWEL PWSEQQAQFLWKKMQVPTNLTLRNLQALGTLAGGMSCEFLQQI NSMVDLFLEVHMIYQLPTRVRGSLRACIWAELQRRMAMPEPEWT TVGPELNGLDSKLLLDLPIQLMDRLSNESIMLVVELVQRAPEQLLA 30 LTPLHQAALAERALQNLAPKETPVSGEVLETGLPLVGFLGTESTR QIPLQILLSHLSQLQGFCLETGFWLLQESVLGKPELWSQ DEVEQAGRLVFTLSTEAISLIPREALGPETLERLLEKQQSWEQSR VGQLCREPQLAAKKAALVAGVVRPAAEDLPEVPNCADVRGTFP AAWSATQIAEMELSDFEDCLTLFAGDPGLGPEELRAAMGKAKQL

SEQ ID NO.	Nombre de secuencia	Secuencia
5		WGPPRGFRPEQILQLGRLLIGLGDRELQELILVDWGVLSTLGQID GWSTTQLRIVVSSFLRQSGRHVSHLDFVHLTALGYTLCGLRPEEL QHISSWEFSQAALFLGTLHLQCSEEQLEVLAHLLVLPGGFGPISN WGPEIFTEIGTIAAGIPDLALSALLRGQIQGVTPLAISVIPPKFAVV FSPIQLSSLTSAQAVAVTPEQMAFLSPEQRRRAVAWAQHEGKESP EQQGRSTAWGLQDWSRPSWSLVLTISFLGHLL
10	5	proteína estereocilina murina (NP_536707.2)
15		MALSLQPQLLLLLSLLPQEVTSAPTGPQSLDAGLSLLKSFVATLDQ APQRSLSQSRFSAFLANISSSFQLGRMGEGPVGEPPLQPPALR LHDFLVTLRGSPDWEPMGLLGDVLALLGQEQTTPRDFLVHQAGV LGGLVEALLGALVPGGPPAPTRPPCTRDGPSDCVLAADWLPSLM LLEGTRWQALVQLQPSVDPTNATGLDGREPAPHFLQGLLGLLT PAGELGSEEALWGGLLRTVGAPLYAAFQEGLLRVTHSLQDEVFSI MGQPEPDASGQCQGGNLQQLLLWGMRRNNLSWDARALGFLSGS PPPPALLHCLSRGVPLPRASQPAAHISPRQRRRAISVEALCENHS GPEPPYSISNFSIYLLCQH IKPATPRPPPTTPRPPPTTPQPPPTTT QPIPDTTQPPPVTPRPPPTTPQPPPSTAVICQTAVWYAVSWAPG 20 ARGWLQACHDQFPDQFLDMICGNLSFSALSGPSRPLVKQLCAGL LPPPTSCPPGLIPVPLTPEIFWGCFLNETLWAERLCVEDSLQAVP PRNQAWVQHVCRGPTLDATDFPPCRVGPCGERCPDGGSFLLMV CANDTLYEALVPFWAWLAGQCRISRGGNDTCFLEGMLGPLLPSL PPLGPSPLCLAPGPFLGMLSQLP RCQSSVPALAHPTRLHYLLRL 25 LTFLLGPGTGGAETQGMLGQALLSSLPDNCSFWDAFRPEGRRS VLRTVGEYLQREEPTPPGLDSSLSLGSMSKMELLSCFSPVLWD LLQREKSVWALRTL VKAYLRMPPEDLQQLVLSAEMEEAAQGFLT MLRSWAKLKVQPSEEQAMGRLTALLQRYPRLTSQLFIDMSPLIP FLAVPDLMRFPSSLANDSVLAAIRDHSSGMKPEQKEALAKRLLA 30 PELFGEVPDWPQELLWAALPLLPHLPLESFLQLSPHQIQALED SW PVADLGPGHARHVLRSLVNQSMEDGEEQVLRRLGSLACFLSPEEL QSLVPLSDPMGPVEQGLLECAANGTLSPEGRVAYELLGVLRSSG GTVLSPREL RVWAPLFPQLGLRFLQELSETQLRAMLPALQGASVT PAQAVLLFGRLLPKHDL SLEELCSLHPLL PGLSPQTLQAIPKRVLV
35		

SEQ ID NO.	Nombre de secuencia	Secuencia
5		GACSCLGPELSRLSACQIAALLQTFRVKDGVKNMGAAGAGSAVCI PGQPTTWPDCLLPLLPLKLLQLDAAALLANRRLYRQLPWSEQQA QFLWKKMQVPTNLSLRNLQALGNLAGGMTCEFLQQISSMVDFLD VVHMLYQLPTGVRESLRACIWTELQRRMTMPEPELTTLGPELSEL DTKLLLDLPIQLMDRLSNDSIMLVVEMVQGAPEQLLALTPLHQTAL AERALKNLAPKETPISKEVLETLGPLVGFLGIESTRRIPLPILLSHLS 10 QLQGFCLGETFATELGWLLLQEPVLGKPELWSQDEIEQAGRLVF TLSAEAISSIPREALGPETLERLLGKHQSWEQSRVGHLCGESQLA HKKAALVAGIVHPAAEGLQEPVPNCADIRGTFPAAWSATQISEME LSDFEDCLSLFAGDPGLGPEELRAAMGKAKQLWGPPRGFRPEQI 15 LQLGRLLIGLGERELQELTLVDWGVLSLQIDGWSSMQLRAVVS SFLRQSGRHVSHLDFIYLTALGYTVCGLRPEELQHISWEFSQAA LFLGSLHLPCSEEQLEVLAYLLVLPGGFGPVSNWGPEIFTEIGTIA AGIPDLALSALLRGQIQGLTPLAISVIPAPKFAVVFNPQLSSLTRGQ AVAVTPEQLAYLSPEQRRAVAWAQHEGKEIPEQLGRNSAWGLY DWFQASWALALPVSIFGHLL
20	6	Polinucleótido
25		que codifica la estereocilina humana WT de longitud completa (de NM_153700.2),
30		codifica la proteína de la SEQ ID NO: 4 (incluye codón de parada).
35		ATGGCTCTCAGCCTCTGGCCCCTGCTGCTGCTGCTGCTGCTG CTGCTGCTGCTGCTCCTTTGCAGTGACTCTGGCCCCTACTGGGC CTCATTCCCTGGACCCTGGTCTCTCCTTCTGAAGTCATTGCT CTCCACTCTGGACCAGGCTCCCCAGGGCTCCCTGAGCCGCTC ACGGTTCTTTACATTCTGGCCAACATTTCTTCTTCTTTGAGC CTGGGAGAATGGGGGAAGGACCAGTAGGAGAGCCCCCACCT CTCCAGCCGCCTGCTCTGCGGCTCCATGATTTTCTAGTGACAC TGAGAGGTAGCCCCGACTGGGAGCCAATGCTAGGGCTGCTAG GGGATATGCTGGCACTGCTGGGACAGGAGCAGACTCCCCGAG ATTCCTGGTGCACCAGGCAGGGGTGCTGGGTGGACTTGTGG AGGTGCTGCTGGGAGCCTTAGTTCCTGGGGGCCCCCTACCC CAACTCGGCCCCCATGCACCCGTGATGGGCCGTCTGACTGTG TCCTGGCTGCTGACTGGTTGCCTTCTCTGCTGCTGTTGTTAGA GGGCACACGCTGGCAAGCTCTGGTGCAGGTGCAGCCAGTGT GGACCCACCAATGCCACAGGCCTCGATGGGAGGGAGGCAG

SEQ ID NO.	Nombre de secuencia	Secuencia
5		CTCCTCACTTTTTGCAGGGTCTGTTGGGTTTGCTTACCCCAAC AGGGGAGCTAGGCTCCAAGGAGGCTCTTTGGGGCGGTCTGCT ACGCACAGTGGGGGCCCCCTCTATGCTGCCTTTCAGGAGGG GCTGCTCCGTGTCACTCACTCCCTGCAGGATGAGGTCTTCTCC ATTTTGGGGCAGCCAGAGCCTGATACCAATGGGCAGTGCCAG GGAGGTAACCTTCAACAGCTGCTCTTATGGGGCGTCCGGCAC 10 AACCTTTCCTGGGATGTCCAGGCGCTGGGCTTTCTGTCTGGAT CACCACCCCCACCCCCTGCCCTCCTTCACTGCCTGAGCACGG GCGTGCCTCTGCCCAGAGCTTCTCAGCCGTCAGCCCACATCA GCCCACGCCAACGGCGAGCCATCACTGTGGAGGCCCTCTGTG AGAACCACTTAGGCCCAGCACCAACCCTACAGCATTTCCAATT 15 CTCCATCCACTTGCTCTGCCAGCACACCAAGCCTGCCACTCCA CAGCCCCATCCCAGCACCACTGCCATCTGCCAGACAGCTGTG TGATATGCAGTGTCTGGGCACCAGGTGCCCAAGGCTGGCTA CAGGCCTGCCACGACCAGTTTCCTGATGAGTTTTTGATGCGA TCTGCAGTAACCTCTCCTTTTCAGCCCTGTCTGGCTCCAACCG 20 CCGCCTGGTGAAGCGGCTCTGTGCTGGCCTGCTCCCACCCCC TACCAGCTGCCCTGAAGGCCTGCCCCCTGTTCCCCTCACCCC AGACATCTTTTGGGGCTGCTTCTTGAGAATGAGACTCTGTGG GCTGAGCGACTGTGTGGGGAGGCAAGTCTACAGGCTGTGCCC CCCAGCAACCAGGCTTGGGTCCAGCATGTGTGCCAGGGCCCC 25 ACCCCAGATGTCACTGCCTCCCCACCATGCCACATTGGACCCT GTGGGGAACGCTGCCCCGATGGGGGCAGCTTCCTGGTGATG GTCTGTGCCAATGACACCATGTATGAGGTCCTGGTGCCCTTCT GGCCTTGGCTAGCAGGCCAATGCAGGATAAGTCGTGGGGGCA ATGACACTTGCTTCCTAGAAGGGCTGCTGGGCCCCCTTCTGC 30 CCTCTCTGCCACCACTGGGACCATCCCCACTCTGTCTGACCCC TGGCCCCCTTCCTCCTTGGCATGCTATCCCAGTTGCCACGCTGT CAGTCCTCTGTCCCAGCTCTTGCTCACCCACACGCCTACACT ATCTCCTCCGCCTGCTGACCTTCCTCTTGGGTCCAGGGGCTG GGGGCGCTGAGGCCCAGGGGATGCTGGGTGCGGGCCCTACTG 35

SEQ ID NO.	Nombre de secuencia	Secuencia
5		CTCTCCAGTCTCCCAGACAACTGCTCCTTCTGGGATGCCTTTC GCCCAGAGGGCCGCGCAGTGTGCTACGGACGATTGGGGAA TACCTGGAACAAGATGAGGAGCAGCCAACCCCATCAGGCTTT GAACCCACTGTCAACCCCAGCTCTGGTATAAGCAAGATGGAG CTGCTGGCCTGCTTTAGTCCTGTGCTGTGGGATCTGCTCCAGA GGGAAAAGAGTGTTTGGGCCCTGCAGATTCTAGTGCAGGCGT
10		ACCTGCATATGCCCCCAGAAAACCTCCAGCAGCTGGTGCTTTC AGCAGAGAGGGAGGCTGCACAGGGCTTCCTGACACTCATGCT GCAGGGGAAGCTGCAGGGGAAGCTGCAGGTACCACCATCCG AGGAGCAGGCCCTGGGTGCGCTGACAGCCCTGCTGCTCCAG CGGTACCCACGCCTCACCTCCCAGCTCTTCATTGACCTGTCAC
15		CACTCATCCCTTTCTTGGCTGTCTCTGACCTGATGCGCTTCCC ACCATCCCTGTTAGCCAACGACAGTGTCTGGCTGCCATCCG GGATTACAGCCCAGGAATGAGGCCTGAACAGAAGGAGGCTCT GGCAAAGCGACTGCTGGCCCCTGAACTGTTTGGGGAAGTGCC
20		TGCCTGGCCCCAGGAGCTGCTGTGGGCAGTGCTGCCCCTGCT CCCCACCTCCCTCTGGAGAACTTTTTGCAGCTCAGCCCTCAC CAGATCCAGGCCCTGGAGGATAGCTGGCCAGCAGCAGGTCTG GGGCCAGGGCATGCCCCGCCATGTGCTGCGCAGCCTGGTAAA
25		CCAGAGTGTCCAGGATGGTGAGGAGCAGGTACGCAGGCTTGG GCCCCTCGCCTGTTTCCTGAGCCCTGAGGAGCTGCAGAGCCT AGTGCCCCTGAGTGATCCAACGGGGCCAGTAGAACGGGGGCT GCTGGAATGTGCAGCCAATGGGACCCTCAGCCCAGAAGGACG
30		GGTGGCATATGAACTTCTGGGTGTGTTGCGCTCATCTGGAGG AGCGGTGCTGAGCCCCCGGGAGCTGCGGGTCTGGGCCCCCTC TCTTCTCTCAGCTGGGCCTCCGCTTCCTTCAGGAGCTGTCAGA GCCCCAGCTTAGAGCCATGCTTCCTGTCCTGCAGGGAAGTAG
35		TGTTACACCTGCTCAGGCTGTCCTGCTGCTTGGACGGCTCCTT CCTAGGCACGATCTATCCCTGGAGGAACTCTGCTCCTTGCACC TTCTGCTACCAGGCCTCAGCCCCCAGAACTCCAGGCCATCC CTAGGCGAGTCCTGGTTCGGGGCTTGTTCTGCTGGCCCCCTG

SEQ ID NO.	Nombre de secuencia	Secuencia
5		A A C T G T C A C G C C T C T C A G C C T G C C A G A C C G C A G C A C T G C T G C A G A C C T T T C G G G T T A A G A T G G T G T T A A A A T A T G G G T A C A A C A G G T G C T G G T C C A G C T G T G T G T A T C C C T G G T C A G C C T A T T C C C A C C A C C T G G C C A G A C T G C C T G C T T C C C C T G C T C C C A T T A A A G C T G C T A C A A C T G G A T T C C T T G G C T C T T C T G G C A A A T C G A A G A C G C T A C T G G G A G C T G C C C T G G T C T G A G C A G C A G G C A C A G T T T C T 10 C T G G A A G A A G A T G C A A G T A C C C A C C A A C C T T A C C C T C A G G A A T C T G C A G G C T C T G G G C A C C C T G G C A G G A G G C A T G T C C T G T G A G T T T C T G C A G C A G A T C A A C T C C A T G G T A G A C T T C C T T G A A G T G G T G C A C A T G A T C T A T C A G C T G C C C A C T A G A G T T C G A G G G A G C C T G A G G G C C T G T A T C T G G G C A G A G C T A C A G C G G A G G A T G G C A A T 15 G C C A G A A C C A G A A T G G A C A A C T G T A G G G C C A G A A C T G A A C G G G C T G G A T A G C A A G C T A C C T G G A C T T A C C G A T C C A G T T G A T G G A C A G A C T A T C C A A T G A A T C C A T T A T G T T G G T G G T G G A G C T G G T G C A A A G A G C T C C A G A G C A G C T G C T G G C A C T G A C C C C C C T C C A C C A G G C A G C C C T G G C A G A G A G G G C A C T A C A A A A C C T G G C T C 20 C A A A G G A G A C T C C A G T C T C A G G G G A A G T G C T G G A G A C C T T A G G C C C T T T G G T T G G A T T C C T G G G G A C A G A G A G C A C A C G A C A G A T C C C C C T A C A G A T C C T G C T G T C C C A T C T C A G T C A G C T G C A A G G C T T C T G C C T A G G A G A G A C A T T T G C C A C A G A G C T G G G A T G G C T G C T A T T G C A G G A G T C T G T T C T T G G G A A A C C A G A G T T G T G G A G C 25 C A G G A T G A A G T A G A G C A A G C T G G A C G C C T A G T A T T C A C T C T G T C T A C T G A G G C A A T T T C C T T G A T C C C C A G G G A G G C C T T G G G T C C A G A G A C C C T G G A G C G G C T T C T A G A A A A G C A G C A G A G C T G G G A G C A G A G C A G A G T T G G A C A G C T G T G T A G G G A G C C A C A G C T T G C T G C C A A G A A A G C A G C C C T G G T A G C A G G G G T G G T G C G A C C A 30 G C T G C T G A G G A T C T T C C A G A A C C T G T G C C A A A T T G T G C A G A T G T A C G A G G G A C A T T C C C A G C A G C C T G G T C T G C A A C C C A G A T T G C A G A G A T G G A G C T C T C A G A C T T T G A G G A C T G C C T G A C A T T A T T T G C A G G A G A C C C A G G A C T T G G G C C T G A G G A A C T G C G G G C A G 35 C C A T G G G C A A A G C A A A A C A G T T G T G

SEQ ID NO.	Nombre de secuencia	Secuencia
5		GGGTCCCCCCCCGGGGATTTTCGTCCTGAGCAGATCCTGCAGCT TGGTAGGCTCTTAATAGGTCTAGGAGATCGGGAACCTACAGGA GCTGATCCTAGTGGACTGGGGAGTGCTGAGCACCTGGGGCA GATAGATGGCTGGAGCACCACTCAGCTCCGCATTGTGGTCTC CAGTTTCCTACGGCAGAGTGGTCGGCATGTGAGCCACCTGGA CTTCGTTTCATCTGACAGCGCTGGGTTATACTCTCTGTGGACTG 10 CGGCCAGAGGAGCTCCAGCACATCAGCAGTTGGGAGTTCAGC CAAGCAGCTCTCTTCCTCGGCACCCTGCATCTCCAGTGCTCTG AGGAACAACCTGGAGGTTCTGGCCCCACCTACTTGTACTGCCTG GTGGGTTTGGCCCAATCAGTAACTGGGGGCCTGAGATCTTCA CTGAAATTGGCACCATAGCAGCTGGGATCCCAGACCTGGCTC 15 TTTGAGCACTGCTGCGGGGACAGATCCAGGGCGTTACTCCTC TTGCCATTTCTGTTCATCCCTCCTCCTAAATTTGCTGTGGTGTTT AGTCCCATCCAACCTATCTAGTCTCACCAGTGCTCAGGCTGTGG CTGTCACTCCTGAGCAAATGGCCTTTCTGAGTCCTGAGCAGCG ACGAGCAGTTGCATGGGCCCAACATGAGGGAAAGGAGAGCCC 20 AGAACAGCAAGGTCTGAAGTACAGCCTGGGGCCTCCAGGACTG GTCACGACCTTCCTGGTCCCTGGTATTGACTATCAGCTTCCTT GGCCACCTGCTATGA
25	7	Polinucleótido que codifica la estereocilina de tipo salvaje murina de longitud completa (de NM_080459.2), codifica la proteína de la
30	SEQ ID NO: 5	(incluye codón
35		ATGGCTCTGAGCCTCCAGCCCCAGCTGCTCCTTCTCCTGTGCG TCCTGCCGCAGGAAGTGACTTCAGCCCCTACTGGGCCTCAGT CTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCGTAGCCAC TCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTCACGGTTC TCTGCGTTCCTGGCCAACATTTCTTCATCCTTCCAGCTTGGGA GGATGGGGGAGGGACCGGTGGGAGAGCCCCCACCTCTCCAG CCCCCTGCACTTCGACTTCATGATTTCTCGTGACACTGAGAG GTAGCCCAGACTGGGAGCCAATGCTAGGGCTTCTGGGAGATG TGCTGGCACTCCTGGGACAGGAACAGACTCCCCGGGACTTTT TGGTGCACCAGGCAGGTGTACTGGGTGGACTTGTAGAGGCAT TGTTGGGAGCGTTAGTTCCTGGAGGCCCCCCCTGCCCCCACTC GACCCCCATGCACCCGTGATGGCCCTTCTGACTGTGTCCTGG

SEQ ID NO.	Nombre de secuencia	Secuencia
5	de parada).	CTGCTGATTGGTTGCCTTCTCTGATGTTGTTATTAGAGGGTACA CGCTGGCAGGCCCTGGTGCAGTTGCAGCCCAGTGTGGACCCA ACCAATGCCACAGGTCTTGATGGTAGAGAGCCAGCTCCTCACT TTTTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGTT GGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCACAGT GGGGGCCCCCTCTATGCTGCCTTCCAGGAGGGGCTACTGCG 10 AGTCACTCATTCTCTGCAAGATGAGGTCTTTTCTATTATGGGAC AGCCAGAGCCTGATGCCAGTGGGCAGTGCCAGGGAGGCAAC CTTCAACAGCTGCTTTTATGGGGCATGCGGAACAACCTTTCTT GGGACGCCCCGAGCACTGGGTTTTCTATCTGGATCACCACTC CACCCCCTGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTC 15 TGCCCAGGGCTTCCCAGCCTGCGGCTCACATCAGCCCTCGAC AGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGAGAACCCT CAGGCCCAGAGCCACCCTACAGCATCTCCAATTCTCCATCTA CTTGCTCTGCCAGCACATCAAGCCTGCCACCCCGCGGCCCCC TCCTACCACCCACAGGCCTCCTCCTACCACCCACAGCCCCC 20 TCCTACCACTACACAGCCCATTCTGACACTACACAGCCCCCT CCTGTCACCCCAAGGCCTCCTCCTACCACCCACAACCCCT CCTAGCACAGCTGTCATCTGCCAGACAGCTGTATGGTACGCA GTCTCGTGGGCACCAGGTGCCCCGAGGTTGGCTCCAAGCCTGC CATGATCAGTTTCCTGATCAATTTCTGGATATGATCTGCGGCAA 25 CCTCTCATTTTCAGCCCTGTCTGGCCCCAGTCGTCCTTTGGTA AAGCAGCTCTGTGCTGGCTTGCTCCACCCCCCACTAGCTGT CCACCAGGCCTGATCCCTGTGCCCCTCACCCAGAAATATTCT GGGGCTGTTTCCTGGAGAATGAGACACTGTGGGCTGAACGGT TGTGTGTGGAGGACAGTCTGCAGGCTGTGCCCCCGAGGAA 30 CCAGGCTTGGGTTTCAGCATGTGTGTGCGGGGCCCCACCTTGGA CGCCACTGATTTTCCACCGTGCCGCGTTGGACCCTGTGGGGA ACGCTGCCCAGATGGGGGCAGCTTCCTGCTCATGGTCTGTGC CAATGACACTCTGTATGAAGCCTTGTTCCCTTCTGGGCTTGG CTAGCAGGCCAATGCAGAATTAGTCGTGGAGGAAATGATACTT 35

SEQ ID NO.	Nombre de secuencia	Secuencia
5		GCTTTCTAGAAGGCATGCTGGGCCCCTTGTTGCCCTCTCTGCC CCCTCTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTT CTGCTTGGCATGTTATCCCAGTTGCCACGCTGTCAGTCCTCCG TGCCAGCCCTCGCCCACCCACGCGCCTACATTACCTCCTGC GCCTACTGACCTTCCTTCTGGGTCCAGGGACTGGGGGTGCCG AGACGCAGGGGATGTTAGGTCAAGCCCTGCTGCTCTCTAGTC 10 TCCCAGACAACTGTTCAATTCTGGGATGCCTTCCGCCCAGAGG GCCGGAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGC GGGAAGAGCCAACCCACAGGCTTAGACTCCTCCCTCAGCC TCGGCTCTGGTATGAGCAAGATGGAGCTTCTGTCCTGCTTCAG 15 TCCTGTACTGTGGGATCTACTCCAGAGAGAGAAGAGCGTTTGG GCCCTGAGGACCCTGGTGAAGGCCTACCTGCGCATGCCTCCA GAAGACCTTCAGCAGCTTGTGCTTTCAGCAGAGATGGAGGCT GCACAGGGCTTCCTGACGCTCATGCTTCGTTCCCTGGGCTAAG CTGAAGGTTCAACCATCCGAGGAGCAGGCCATGGGCCGCTG ACAGCCTTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAA 20 CTCTTTATCGACATGTCACCGCTCATCCCCTTCCTGGCTGTCC CTGACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAACGACAG TGTCCTGGCTGCCATCAGGGATCACAGCTCAGGAATGAAGCC TGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCCCCTGA GCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAGCTGCTGTG 25 GGCAGCCCTGCCTCTGCTTCCCCATCTGCCTCTGGAGAGCTTT CTCCAGCTCAGCCCTCACCAGATCCAGGCCCTGGAGGATAGC TGGCCAGTAGCAGATCTTGGGCCGGGACACGCCCCGACATGTG CTTCGTAGCCTAGTAAACCAGAGCATGGAGGATGGGGAGGAG CAGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCCTGAGTCCT 30 GAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCCAATGGGG CCTGTAGAACAGGGTCTGCTGGAATGTGCGGCCAATGGGACC CTCAGCCCAGAAGGACGGGTGGCATATGAACTTCTGGGAGTG TTGCGTTCATCTGGAGGAACTGTCTTAAGCCCCCGAGAGCTGA GGGTCTGGGCACCTCTCTTTCCCCAGCTGGGCCTCCGCTTCC 35

SEQ ID NO.	Nombre de secuencia	Secuencia
5		TGCAGGAGCTCTCAGAGACCCAGCTTAGAGCCATGCTTCCTG CCCTACAGGGAGCCAGTGTACACCTGCCCAGGCTGTTCTGT TGTTTGGAAGGCTCCTTCCTAAGCATGATCTGTCCCTGGAGGA ACTCTGCTCCCTGCACCCTCTCCTGCCAGGTCTCAGCCCCCA GACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGGTGCTTGT TCCTGCCTGGGCCCTGAACTGTCAAGGCTTTCAGCTTGCCAGA 10 TTGCAGCTCTGCTGCAGACCTTTCGGGTAAAAGATGGTGTTAA AAATATGGGTGCAGCAGGTGCCGGCTCAGCCGTGTGCATTCC TGGGCAGCCCACCACTTGGCCAGACTGCCTGCTTCCCCTGCT CCCATTAAGCTGCTACAGCTGGACGCTGCAGCTCTTCTGGCA AACCGAAGACTCTATCGGCAGCTGCCTTGGTCTGAGCAACAG 15 GCACAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTGA GCCTGAGGAATCTGCAGGCTCTGGGCAACTTGGCAGGAGGCA TGACCTGCGAGTTTCTGCAGCAGATCAGCTCAATGGTTGACTT TCTTGATGTGGTACACATGCTCTACCAGCTGCCCCACTGGTGTT CGAGAGAGCCTGCGGGCCTGTATCTGGACAGAGCTACAGCGG 20 AGGATGACAATGCCAGAGCCAGAGCTGACCACCCTAGGGCCA GAACTGAGTGAAGTTGACACAAAGCTACTCCTGGACTTGCCGA TCCAGCTGATGGACAGATTGTCCAATGATTCCATTATGTTGGT GGTGGAGATGGTCCAAGGCGCTCCAGAGCAGCTGCTGGCACT GACCCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTAA 25 AAACCTGGCTCCAAAGGAGACCCCAATCTCCAAAGAAGTGCTG GAGACACTGGGCCCCTTGTTGGATTCTGGGAATAGAGAGC ACGCGACGGATCCCTTTACCCATTCTACTGTCTCATCTCAGTC AGCTGCAGGGCTTCTGCCTAGGAGAGACATTTGCCACAGAGC TGGGATGGCTGCTGTTGCAGGAGCCTGTTCTTGAAAACCAG 30 AATTGTGGAGCCAGGATGAAATAGAGCAAGCTGGACGCCTAG TATTAATCTGTCTGCTGAGGCTATTTCTCGATCCCCAGGGA GGCTTTGGGCCCAGAGACACTGGAGAGGCTTCTGGGAAAGCA TCAAAGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGGA GTCACAGCTTGCCCAAGAAAGCAGCTCTGGTAGCTGGGAT

SEQ ID NO.	Nombre de secuencia	Secuencia
5		TGTGCATCCAGCTGCTGAGGGTCTCCAAGAGCCTGTACCAAA CTGTGCAGACATACGGGGAACCTTCCCAGCGGCCTGGTCTGC GACACAAATCTCAGAGATGGAACCTCTCAGACTTTGAAGACTGC CTGTCACTATTTGCTGGAGATCCAGGACTTGGTCCTGAGGAAC TACGGGCAGCCATGGGCAAGGCCAAGCAGTTGTGGGGTCCC CCTCGAGGATTCCGTCCTGAGCAGATCTTGAGCTGGGCCGT 10 CTCCTGATAGGTCTAGGAGAACGGGAACTGCAGGAGCTTACC TTGGTGGACTGGGGTGTGCTGAGCAGCCTGGGGCAAATAGAT GGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAGTTTC CTAAGGCAGAGTGGTCGGCATGTGAGCCACCTGGACTTCATTT ATCTGACAGCACTGGGTACACAGTCTGTGGATTGCGACCAGA 15 GGAGTTACAGCACATCAGCAGTTGGGAGTTTAGCCAAGCAGC TCTCTTCCTGGGTAGCTTGATCTCCCGTGCTCTGAGGAACAG CTGGAAGTTCTGGCCTATCTCCTTGTGTTGCCTGGTGGCTTTG GCCCAGTCAGTAACTGGGGGCCTGAGATCTTCACTGAAATTG GCACAATAGCAGCTGGCATCCCAGACCTGGCTCTTTCAGCATT 20 ACTGCGGGGACAGATCCAAGGCCTGACTCCTCTTGCCATTTCT GTCATTCTGCTCCCAAGTTTGAGTGGTCTTCAACCCCATCC AGTTATCTAGTCTCACCAGGGGTCAGGCCGTAGCTGTTACTCC TGAACAGCTGGCCTATCTGAGTCCTGAGCAGCGGCGAGCAGT TGCATGGGCCCCAACACGAAGGAAGGAGATCCCAGAGCAGCT 25 GGGTCGAAACTCAGCCTGGGGTCTCTACGACTGGTTCCAAGC CTCCTGGGCCCTGGCATTGCCCGTCAGCATTTTTGGCCACCTA TTATGA

Expresión de estereocilina en células de mamífero

30 Las mutaciones en STRC se han relacionado con la pérdida auditiva neurosensorial. Las composiciones y métodos descritos en la presente se pueden usar para inducir o aumentar la expresión de la estereocilina de WT mediante la administración a un sujeto o al poner en contacto una célula con un primer vector de ácido nucleico que contiene un polinucleótido que codifica una parte de extremo N de una proteína estereocilina y un segundo vector de ácido nucleico que

35 contiene un polinucleótido que codifica una parte de extremo C de una proteína estereocilina. Para

utilizar vectores de ácido nucleico para aplicación terapéutica en el tratamiento de la pérdida auditiva neurosensorial, pueden dirigirse al interior de la célula y, en particular, a tipos celulares específicos. Se ha establecido una amplia gama de métodos para el suministro de proteínas a células de mamíferos y para la expresión estable de genes que codifican proteínas en células de mamíferos.

Polinucleótidos que codifican estereocilina

Una plataforma que puede usarse para lograr concentraciones intracelulares terapéuticamente eficaces de estereocilina en células de mamíferos es mediante la expresión estable del gen que codifica la estereocilina (por ejemplo, mediante la integración en el genoma nuclear o mitocondrial de una célula de mamífero, o por formación de concatémeros episomales en el núcleo de una célula de mamífero). El gen es un polinucleótido que codifica la secuencia de aminoácidos primaria de la proteína correspondiente. Para introducir los genes exógenos en una célula de mamífero, los genes pueden incorporarse en un vector. Los vectores se pueden introducir en una célula mediante una variedad de métodos, que incluyen transformación, transfección, transducción, captación directa, bombardeo con proyectiles y encapsulación del vector en un liposoma. Ejemplos de métodos adecuados para transfectar o transformar células incluyen precipitación con fosfato cálcico, electroporación, microinyección, infección, lipofección y absorción directa. Tales métodos se describen con más detalle, por ejemplo, en Green, et ál., *Molecular Cloning: A Laboratory Manual*, cuarta edición (Cold Spring Harbor University Press, Nueva York 2014); y Ausubel, et ál., *Current Protocols in Molecular Biology* (John Wiley & Sons, Nueva York 2015), cuyas divulgaciones se incorporan a la presente mediante referencia.

La STRC también se puede introducir en una célula de mamífero al dirigir vectores que contienen partes de un gen que codifica una proteína estereocilina a fosfolípidos de la membrana celular. Por ejemplo, los vectores pueden dirigirse a los fosfolípidos en la superficie extracelular de la membrana celular al unir la molécula del vector a una proteína VSV-G, una proteína viral con afinidad por todos los fosfolípidos de la membrana celular. Tal construcción puede producirse usando métodos bien conocidos por los expertos en la materia.

El reconocimiento y la unión del polinucleótido que codifica una proteína estereocilina por la ARN polimerasa de mamífero es importante para la expresión génica. Como tal, se pueden incluir elementos de secuencia dentro del polinucleótido que exhiben una alta afinidad por los factores de transcripción que reclutan la ARN polimerasa y promueven el ensamblaje del complejo de transcripción en el sitio de inicio de la transcripción. Tales elementos de secuencia incluyen, por ejemplo, un promotor de mamífero, cuya secuencia puede ser reconocida y unida por factores de iniciación de la transcripción específicos y finalmente ARN polimerasa. Se han descrito ejemplos

de promotores de mamíferos en Smith, et ál., Mol. Sys. Biol., 3:73, publicación en línea, cuya divulgación se incorpora a la presente mediante referencia.

Los polinucleótidos adecuados para su uso en las composiciones y métodos descritos en la presente incluyen aquellos que codifican una proteína estereocilina corriente abajo de un promotor de mamífero (por ejemplo, un polinucleótido que codifica una parte del extremo N de una estereocilina corriente abajo de un promotor de mamífero). Los promotores que son útiles para la expresión de una proteína estereocilina en células de mamíferos incluyen los promotores específicos de OHC, tal como un promotor de oncomodulina (OCM) (por ejemplo, un polinucleótido que tiene al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, de identidad de secuencia) con respecto a cualquiera de las secuencias de promotor de OCM enumeradas en la Tabla 3 (por ejemplo, cualquiera de las SEQ ID NO: 1-3)).

Promotores de oncomodulina

Los presentes inventores han descubierto una región de 1.140 pares de bases (pb) ubicada corriente arriba del sitio de inicio de la traducción de la OCM que es suficiente para impulsar la expresión génica en las OHC. Las composiciones y métodos descritos en la presente pueden, por lo tanto, utilizarse para expresar estereocilina en las OHC para tratar a sujetos que tienen o corren el riesgo de desarrollar pérdida auditiva (por ejemplo, pérdida auditiva neurosensorial asociada con una mutación en STRC, tal como DFNB16). Dado que los promotores de OCM descritos en la presente (por ejemplo, un promotor de OCM que tiene al menos un 85 % de identidad de secuencia con la SEQ ID NO: 1) pueden utilizarse para inducir la expresión génica específica de las OHC, pueden reducir o eliminar la expresión fuera del objetivo en otras células del oído interno (por ejemplo, en células distintas de las OHC), mejorando así la seguridad y eficacia de la terapia génica al dirigir la expresión de la STRC a las células en las que se expresa endógenamente y reducir la toxicidad asociada con la expresión fuera del objetivo.

Las composiciones y métodos descritos en la presente incluyen un promotor OCM enumerado en la Tabla 3 (por ejemplo, cualquiera de las SEQ ID NO: 1-3) que sea capaz de expresar estereocilina específicamente en las OHC, tal como una secuencia polinucleotídica que tenga al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia) con cualquiera de las SEQ ID NO: 1-3.

Las secuencias de promotor de OCM de ejemplo se enumeran en la Tabla 3.

Tabla 3. Secuencias de promotor de OCM

SEQ ID NO:	Descripción de la secuencia promotora	Secuencia del promotor
1	Secuencia promotora de OCM murina (1140 pb)	AGCAGGTTTGTTACAGAAACCTTAGTTAAGGTTTGTT GAGGGTTTTTTTTCTCTCTCTCTCTTAATTGGCTG TCCCAATCCATCCTTCTATAAATAGAAAAGAGAGACA GGGAGTGTGTGTGGTTTCATTACTAAGGTAAAGACA CTTGAGCTACACACACTTGATCCCTGAACATGAAAT CTAAGAGGTTGAACGATCACAGTTTCAGGACTATAT AAGGTGGTGAAAGACCATCTGCTTCGTTTTTCTGTTT GTTCTACAACCTCTTCCCTCCGCTTGATTTTAACTC TAAATTGGTGAGTAGCTGGTGGGCTCACCAGACTCC GAGATCCTCTTCTCTGCACGCACTGTATTAGACTTG GCACCCGGGAGGATTTTCACCTCTGCTGCATGGGC TAATCTTCCACAAGGGATCTGTGGTATTGCAATCTC GGGTTGATGCATGACGGTGATGTTGTGTTTATAGCA TGGCTAAGGTTTAGCTGCCTATGATGATTGGTTAGG GAAGGATAATTTTGTCTAGAAGATTGGACTTTAGGG AAAAAAAACCCCACTTTTATTTGCTTTTAGAATTTTAA AAGACTGGGCCATGTAGCTCAGGCTGGTTTGGAGT TCATTATGTAGTCAAGGATGCTCTTGGACTCTTTAGC ATCCTCCTCCTCCTCTTTTTCTCCTCCTCCTTCTTG TTCTTCTTCTTGTTCTTCTCCTCTTCCCCTTCTTCCC CCTTCTCTCCCTCTTCCCTCCTCTTCCCTCCTTCTT GTTCTTCTTCTCCTCTTCCCCTTCTCCTCTCCCCCTTG CCTCCTCTTCTTCTCCTCCTCCTCCTCTTCTTCTTT CTGAGTACCAAGATTGCAAGTGTGCACACGATGACC AGCTTGGTCTTTCTTTGTCTTTTTTTTTTAACCTCAAT TTTGGAGTGAATTCAAGAGCAACCATGTAGTCAAGA GGTGGCTGGAGTCTTTTCTGTATCTGGGTTTGGTTT AGTACTCTGCCCCATCACTTAACAGGTCCTTATGGC CACATCTTAAAAAATTCTAGAGATACACGGTGTCTG GTGAGTGGCTGAGAATGTGTGGTCTTCCCATTCTC TGTCACCGTGGCTCACATCTTGTTTCCTCTGTTCCG CCAGGTAGAAA

SEQ ID NO:	Descripción de la secuencia promotora	Secuencia del promotor
2	Secuencia promotora de OCM humana que contiene un polinucleótido localizado -2 kb a +0,5 kb del TSS del gen OCM humano	TTTTACCACAATAATTTAAAAAGAACAGTCTAGCACAG TGCTGGCCATATAAAGGCTCAATAAATGTTTGCTGA AAGTTAAAAAAAAAAAAAAAAAAAAAAAAAAGCCAGGC GCAGTGGTTCATTCCTGTAATCCCAGCACTTTAGGA GGATGAGGTGGGAGAATTACTTGAGCCCAGGAGTT CGAGACCAGCCTAGGCAACATGGCAAACCCTGTC AAAACCCTGTCTCTCCAAAAAATATGCATATTTAAAA AATTAGCCAGGCATGGTGGTGTGTGCCTGTAGTACC AGCTACTCGGGAGACTGAAGTGGGAGGATCGCTTG AGCCTGGGAGGTCAAGGCTGCAATGAGCTGAGATC GTGCCACTGCACTCCAGCCGGGGCAACAGAGCAAG ACCCTGTCACAACAGAAACAAAATCTTGAGGTGTCT AGTCCTGGCCTCAGCCTCAGAATATTTGTTTCTGAA CATGTTAGTTTTGGGGGTTGGGGATGCTGGTTTGAT TTCCTCCTTTTTGCCTTTTGAGTGTGTGCAATTTATG GTATAGCTGGGAAACGTCAAAGTCAAGAGTTTTGTA GGAAAGTCACGTCACTTAGCCCTGTCTCCTGTGCCG GGTGAGACCTGTGTGTGCACTTGGTGACAATGGCTT TGAGTCTGTCAACTCCAGACTGAGGTCAGCCTTACA CACCCATAGTTCCCAAAGCTGAAAACAGGCCTGCCT CCAACGGTACCTGCTAATATCAGGGGAGCCTTTTCA GCTTACAGAGCACCTGTATGTGTTTGTCTTAGTTC AGGCCACCATCTCCACCTTACCAGGCATCTAGAACC TTCTCCACACTTTGCCAACAGGGTTCGTTTGCAGAA TTGAAATCTTAGTTAAGGTTTGTGGAAGTTTGTGTT GTTTTTTTTTTTTTTTACAATTGGCTGTTCCCACCCA CATTCCCTTGAGACATAAATAGAAAAAAAAAAAAAAAA GAGGTTTCATGAGTAAGACAAGACATTTGAGCTGCA TCCACTTGATCCTTGAAAAGGAAATCTAAGAGGTTG TAACTATCACTTTTTCTAGCCTATATAAGGTAGGTCA GTAAGGTAGCAAAAACACATCTGTTGTTTTGCTCCTT CAACTCTTTTTCCTGATTCTTCCTGGGGGGAAACCG

SEQ ID NO:	Descripción de la secuencia promotora	Secuencia del promotor
5		AAAACGGTGAGTAACTGGTGGACACATCAGACCCC AGACTCTTTTCTTCACTGCATGCATTCATATTAGGCT CAGGTGCTTAGACTCCTGTTTTCCGGTGGCTCTGAC ACCTGGAAGGATTTTAATCTCTGGGAGATGGGCTTT TCATCCATCTGCTTCCCACCTTTCAGGACAGGTGCA TGCCTTCTTCCACAGAATGTCTGCAAGCAGCCCCAAA CTGTATCCTTTCCCACGTGGAATTTGCAACATTGCAT
10		CTCTCGGGCTGCTGTAGGAAAATGCCAGTGCATGT GTAACATGGTTTACGGCTGCCTATGCAAATGACTGA TTATGTCAGTATAATTTTTATAAGAAAACAATTGAATC CTTCTTTGGGTCAATTTTTTTTTTCCATTTTTGGCATGT ATTCAAAGAAGGCTCTGAGACAAAAAAGGCTGGG
15		GTGTTTTCCGTATCTGGTTTTAATTTGGATATTCTGT CCCGTCACTTAATACAAAACCATGCTTATCACATTTT AAAAATTCTAGACAGGCCTGGCTCGGTGGCTTGCAT CTGTCATCCCAGCACTTTGTGAGGCCAAGGCAGGC AGATCACCTGAGGTCAGGAGCTCAAGACCAGCCTG
20		GCCAACATGGCAAACCCCGTCTCTACTAAAAACAC AAAAATTAGCCAGGCATGGTAGTGCGCACCTGTAAT CCCAGCTACTGGGAAGGCTTAGGCAGGAGAATCAC TTGAGCCCAGGAGGCGGAGGTTGCGGTGAGCCGA GATCACGCTCTTGCACTCCAGCCTGGGTGACAGAG
25		TGAGACTCCGTCTTAATTTAAAAAATAATCTA GACACACATACAGTTTCAGTGGGCCTGGGAAGATGT GTTTCCCCTGGATGTGCACATTCCTGTTTGTGGCTT ATCGCCTCTCATTTATTCTGTGTGAGTAGGTAGAAAA
30		TGAGCATCACGGACGTGCTCAGTGCTGACGACATT GCAGCAGCGCTCCAGGAATGCCGAGGTAGAGGGG ACGTGAGGCGGGGGTGGGATTTCTCACAGCTTTG CACCTCCAGCGAGTCAACACAAAATCAAAATGTAGG CCAGGCGGCCAGACGCAGTGGCTCACACCTGTAAT
35		CCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCAC

SEQ ID NO:	Descripción de la secuencia promotora	Secuencia del promotor
5		GAGGTCAGGAGTTTCGAGACCAGCCTGGCCAAGATG GTGAAACCCCATCTCTACTAAAAATACAAAAAATTA ACCGGGCGTGGTGGTGGGTGCCTGTAATCCCAGCT ACTCGGGAGGCTGAGGCAGAGAATTGCTTGAACCC GGGAGGCAGAAGTTGCAGTGAGCTGAGATCATGCC ACTGCACTCCAGCCTGGGCA
10 15 20 25 30	3 Secuencia del promotor de OCM humana que contiene regiones de la SEQ ID NO: 2 que se conservan en todas las especies de mamíferos	GTTCCCAAAGCTGAAAACAGGCCTGCCTCCAACGG TACCTGCTAATATCAGGGGAGCCTTTTCAGCTTACA GAGCACCCTGTATGTGTTTGTCTTAGTTCAGGCACC TTACCAGGCATCTAGAACCTTCTCCACACTTTGCCA ACAGGGTTCGTTTGCAGAATTGAAATCTTAGTTAAG GTTTGTGTAAGTTTGTGTTGTTTTTTTTTTTTTTTA CAATTGGCTGTTCCCAACCCACATTCCCTTGAGACAT AAATAGAAAAAAAAAAAAAAAAAGAGGTTTCATGAGTAA GACAAGACATTTGAGCTGCATCCACTTGATCCTTGA AAAGGAAATCTAAGAGGTTGTAACATCACTTTTTCT AGCCTATATAAGGTAGGTCAGTAAGGTAGCAAAAC ACATCTGTTGTTTGTCTCCTTCAACTCTTTTCTGA TTCTTCCTGGGGGGAAACCGAAACGGTGAGTAAC GGTGGACACATCAGACCCAGACTCTTTCTTCACT GCATGCATTCATATTAGGCTCAGGTGCTTAGACTCC TGTTTTCCGGTTTACGGCTGCCTATGCAAATGACTG ATTATGTCAGTATAATTTTTATAAGAAAACAATTGAAT CCTTCTTTGGGTCATTTTTTTTTTCCATTTTGGCATG TATGTGCACATTCCTGTTTGTGGCTTATCGCCTCTCA TTTATTCTGTGTGAGTAGGTAGAAAATGAGCATCAC GGACGTGCTCAGTGCTGACGACATTGCAGCAGCGC TCCAGGAATGCCGAGGTAGAGGGGACGTGAGGCG GGGGTGGGATTTCCTCACAGCTTTGCACCTCCAGC

Los polinucleótidos anteriores pueden incluirse en un vector de ácido nucleico y unirse de forma operativa a un transgén para expresar el transgén específicamente en las OHC. En los vectores

descritos en la presente, el transgén puede codificar una parte de extremo N de una proteína estereocilina. De acuerdo con los métodos descritos en la presente, puede administrarse a un sujeto una composición que contenga uno de los polinucleótidos anteriores (por ejemplo, cualquiera de las secuencias de polinucleótidos enumeradas en la Tabla 3 (por ejemplo, SEQ ID NO: 1-3) o una secuencia de polinucleótidos que tenga al menos un 85 % de identidad de secuencia con esta (por ejemplo, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia con cualquiera de las SEQ ID NO: 1-3)) unida de forma operativa a un polinucleótido que codifica, por ejemplo, una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4 o la SEQ ID NO: 5) para el tratamiento de la pérdida auditiva.

Una vez que el polinucleótido que codifica la estereocilina se ha incorporado al ADN nuclear de una célula de mamífero, la transcripción de este polinucleótido puede inducirse por métodos conocidos en la técnica. Por ejemplo, la expresión puede inducirse exponiendo la célula de mamífero a un reactivo químico externo, como un agente que modula la unión de un factor de transcripción y/o ARN polimerasa al promotor de mamífero y regula así la expresión génica. El reactivo químico puede servir para facilitar la unión de la ARN polimerasa y/o factores de transcripción al promotor de mamífero, por ejemplo, mediante la eliminación de una proteína represora que se ha unido al promotor. De manera alternativa, el reactivo químico puede servir para mejorar la afinidad del promotor de mamífero por la ARN polimerasa y/o los factores de transcripción de manera que la tasa de transcripción del gen localizado corriente abajo del promotor aumenta en presencia del reactivo químico. Los ejemplos de reactivos químicos que potencian la transcripción de polinucleótidos mediante los mecanismos anteriores incluyen tetraciclina y doxiciclina. Estos reactivos están disponibles comercialmente (Life Technologies, Carlsbad, CA) y se pueden administrar a una célula de mamífero para promover la expresión génica de acuerdo con protocolos establecidos.

Otros elementos de la secuencia de ADN que se pueden incluir en los polinucleótidos para su uso en las composiciones y métodos descritos en la presente incluyen secuencias potenciadoras. Los potenciadores representan otra clase de elementos reguladores que inducen un cambio conformacional en el polinucleótido que contiene el gen de interés de manera que el ADN adopta una orientación tridimensional que es favorable para la unión de factores de transcripción y ARN polimerasa en el sitio de inicio de la transcripción. Por lo tanto, los polinucleótidos para usar en las composiciones y métodos descritos en la presente incluyen aquellos que codifican una proteína STRC y además incluyen una secuencia potenciadora de mamífero. En la actualidad se conocen muchas secuencias de potenciadores a partir de los genes de los mamíferos, como por ejemplo los potenciadores de los genes que codifican la globina de los mamíferos, la elastasa, la albúmina,

α -fetoproteína e insulina. Los potenciadores para usar en las composiciones y métodos descritos en la presente también incluyen aquellos que se derivan del material genético de un virus capaz de infectar una célula eucariota. Los ejemplos incluyen el potenciador de SV40 en el lado tardío del origen de la replicación (100-270 pb), el potenciador del promotor temprano del citomegalovirus, el potenciador del poliovirus en el lado tardío del origen de la replicación y los potenciadores de adenovirus. Las secuencias potenciadoras adicionales que inducen la activación de la transcripción de genes eucarióticos incluyen el potenciador de CMV y el potenciador de RSV. Un potenciador puede empalmarse en un vector que contiene un polinucleótido que codifica una proteína de interés, por ejemplo, en una posición 5' o 3' de este gen. En una orientación preferida, el potenciador se coloca en el lado 5' del promotor, que a su vez está localizado en 5' con respecto al polinucleótido que codifica una proteína estereocilina.

Los vectores de ácido nucleico descritos en la presente pueden incluir un elemento regulador postranscripcional de marmota (WPRE, por su sigla en inglés). El WPRE actúa al nivel del ARNm, al promover la exportación nuclear de transcripciones y/o al aumentar la eficiencia de la poliadenilación de la transcripción naciente, lo cual aumenta de este modo la cantidad total de ARNm en la célula. La adición de WPRE a un vector puede resultar en una mejora sustancial en el nivel de expresión del transgén de varios promotores diferentes, tanto in vitro como in vivo.

En algunas modalidades, los vectores de ácido nucleico descritos en la presente incluyen una secuencia indicadora, que puede ser útil para verificar la expresión de la estereocilina, por ejemplo, en células y tejidos (por ejemplo, en OHC). Las secuencias indicadoras que pueden proporcionarse en un transgén incluyen secuencias de ADN que codifican β -lactamasa, β -galactosidasa (LacZ), fosfatasa alcalina, timidina cinasa, proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), luciferasa y otras conocidas en la técnica. Cuando se asocian con elementos reguladores que dirigen su expresión, tal como un promotor OCM, las secuencias indicadoras proporcionan señales detectables por medios convencionales, que incluyen ensayos enzimáticos, radiográficos, colorimétricos, de fluorescencia u otros ensayos espectrográficos, ensayos de clasificación de células activadoras fluorescentes y ensayos inmunológicos, que incluyen el ensayo inmunoabsorbente ligado a enzimas (ELISA), el radioinmunoensayo (RIA) y la inmunohistoquímica. Por ejemplo, cuando la secuencia marcadora es el gen LacZ, la presencia del vector que lleva la señal se detecta mediante ensayos de actividad de β -galactosidasa. Cuando el transgén es una proteína verde fluorescente o luciferasa, el vector que lleva la señal puede medirse visualmente mediante la producción de luz o color en un luminómetro.

Sistemas de expresión de vector dual

Vectores duales superpuestos

Un enfoque para expresar proteínas de gran tamaño en células de mamífero implica utilizar vectores duales superpuestos. Este enfoque se basa en el uso de dos vectores de ácido nucleico, cada uno de los cuales contiene una parte de un polinucleótido que codifica una proteína de interés y tiene una región definida de superposición de secuencia con el otro polinucleótido. La recombinación homóloga puede producirse en la región de superposición y dar lugar a la formación de un único polinucleótido que codifica la proteína de interés de longitud completa (por ejemplo, una proteína estereocilina).

Los vectores duales superpuestos para su uso en los métodos y composiciones descritos en la presente contienen al menos 200 bases de secuencia de superposición (por ejemplo, al menos 200 b, 300 b, 400 b, 500 b, 600 b, 700 b, 800 b, 900 b, 1,0 kilobase (kb), 1,1 kb, 1,2 kb, 1,3 kb, 1,4 kb, 1,5 kb o más de secuencia superpuesta). Los vectores de ácido nucleico están diseñados de tal manera que la región superpuesta se centra en o cerca de una posición dentro del polinucleótido codificador de estereocilina que corresponde aproximadamente a la mitad de la longitud del polinucleótido codificador de estereocilina, con una cantidad igual de superposición a cada lado de la posición central. El centro de la región superpuesta también puede elegirse basándose en el tamaño del promotor y las localizaciones de los elementos de secuencia de interés en el polinucleótido que codifica la estereocilina. En algunas modalidades, el polinucleótido que codifica la estereocilina se divide en dos mitades de longitud aproximadamente igual con cierto grado de superposición (por ejemplo, 50 b, 100 b, 150 b, 200 b, 250 b, 300 b, 350 b, 400 b, 450 b, 500 b, 600 b, 700 b, 800 b, 900 b, 1 kb, 1,1 kb, 1,2 kb, 1,3 kb, 1,4 kb, 1,5 kb, o más), en el que la mitad 5' del polinucleótido codifica una parte de extremo N de la proteína estereocilina y la mitad 3' del polinucleótido codifica una parte de extremo C de la proteína estereocilina. Los vectores de ácidos nucleicos para usar en los métodos y composiciones descritos en la presente también están diseñados de manera que aproximadamente la mitad del polinucleótido que codifica la estereocilina esté contenido dentro de cada vector (por ejemplo, cada vector contiene un polinucleótido que codifica aproximadamente la mitad de la proteína estereocilina).

En algunas modalidades, el primer vector de ácido nucleico codifica una parte de extremo N de la proteína estereocilina. En algunas modalidades, el segundo vector de ácido nucleico codifica una parte de extremo C de la proteína estereocilina. En algunas modalidades, la proteína estereocilina tiene la secuencia de la SEQ ID NO: 4 o al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con esta. En algunas modalidades, la proteína estereocilina tiene la secuencia de la SEQ ID NO: 5 o al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con esta. En

algunas modalidades, el polinucleótido que codifica una proteína estereocilina humana de longitud completa tiene la secuencia de la SEQ ID NO: 6 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 6. En algunas modalidades, el polinucleótido que codifica una proteína estereocilina murina de longitud completa tiene la secuencia de la SEQ ID NO: 7 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 7.

Un sistema de vector dual superpuesto de ejemplo incluye un primer vector de ácido nucleico que contiene un promotor OCM descrito anteriormente (por ejemplo, un promotor OCM que tiene al menos un 85 % de identidad de secuencia (por ejemplo, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia) con cualquiera de las SEQ ID NO: 1-3) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4 o la SEQ ID NO: 5) que incluye 500 b inmediatamente 3' de la posición seleccionada como posición central; y un segundo vector de ácido nucleico que contiene la parte de extremo C del polinucleótido que codifica la proteína estereocilina, que incluye 500 b inmediatamente 5' de la posición seleccionada como la posición central, y una secuencia poli(A) (por ejemplo, una secuencia señal poli(A) de hormona de crecimiento bovino (bGH)). Los vectores de ácido nucleico pueden opcionalmente contener regiones no traducidas (UTR) de STRC. En algunas modalidades, el promotor de OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 1 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 1. En algunas modalidades, el promotor de OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 2 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 2. En algunas modalidades, el promotor de OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 3 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 3.

En algunas modalidades, el primer miembro del sistema de vector dual incluye el promotor OCM de la SEQ ID NO:1 (también representado por los nucleótidos 225-1364 de la SEQ ID NO: 43) unido de forma operativa a los nucleótidos que codifican una parte de extremo N de una proteína estereocilina. En determinadas modalidades, la secuencia de nucleótidos que codifica una parte de extremo N de una proteína estereocilina son los nucleótidos 1375-4574 de la SEQ ID NO: 43. Las

secuencias de nucleótidos que codifican una parte de extremo N de una proteína estereocilina pueden estar optimizadas por codón de forma parcial o completa para la expresión. En modalidades particulares, el primer miembro del sistema de vector dual incluye los nucleótidos 225-4574 de la SEQ ID NO:43 flanqueado en cada uno de los lados 5' y 3' por una repetición terminal invertida. En algunas modalidades, las repeticiones terminales invertidas flanqueantes son cualquier variante de las repeticiones terminales invertidas de AAV2 que pueden encapsidarse mediante un plásmido que porta el gen Rep de AAV2. En determinadas modalidades, la repetición terminal invertida flanqueante 5' tiene una secuencia correspondiente a los nucleótidos 1-130 de la SEQ ID NO:43 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí; y la repetición terminal invertida flanqueante 3' tiene una secuencia correspondiente a los nucleótidos 4662-4791 de la SEQ ID NO:43 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí. Los expertos en la técnica entenderán que, para cualquier par dado de secuencias de repeticiones terminales invertidas en un plásmido de transferencia que se usa para crear el vector viral (típicamente al transfectar células con ese plásmido junto con otros plásmidos que llevan el AAV de genes necesario para la formación de vectores virales) (por ejemplo, la SEQ ID NO: 43), que la secuencia correspondiente en el vector viral puede alterarse debido a que las ITR adoptan una orientación "flip" o "flop" durante la recombinación. Por tanto, la secuencia de la ITR en el plásmido de transferencia no es necesariamente la misma secuencia que se encuentra en el vector viral preparado a partir de este. Sin embargo, en algunas modalidades muy específicas, el primer miembro del sistema de vector dual incluye los nucleótidos 1-4791 de la SEQ ID NO:43.

En algunas modalidades, el segundo miembro del sistema de vector dual incluye nucleótidos que codifican la parte de extremo C de la proteína estereocilina seguido inmediatamente por un codón de terminación. En determinadas modalidades, la secuencia de nucleótidos que codifica los aminoácidos de extremo C de la proteína estereocilina son los nucleótidos 211-3440 de la SEQ ID NO:44. Las secuencias de nucleótidos que codifican la parte de extremo C de una proteína STRC pueden estar optimizadas por codón de forma parcial o completa para la expresión. En algunas modalidades, el segundo miembro del sistema de vector dual incluye una secuencia WPRE correspondiente a los nucleótidos 3452-3999 de la SEQ ID NO:44. En algunas modalidades, el segundo miembro del sistema de vector dual incluye la secuencia poli(A) correspondiente a los nucleótidos 4012-4219 de la SEQ ID NO:44. En modalidades particulares, el segundo miembro del sistema de vector dual incluye los nucleótidos 211-4219 de la SEQ ID NO:44 flanqueado en cada

uno de los lados 5' y 3' por una repetición terminal invertida. En algunas modalidades, las repeticiones terminales invertidas flanqueantes son cualquier variante de las repeticiones terminales invertidas de AAV2 que pueden encapsidarse mediante un plásmido que porta el gen Rep de AAV2. En determinadas modalidades, la repetición terminal invertida flanqueante 5' tiene una secuencia correspondiente a los nucleótidos 1-130 de la SEQ ID NO:44 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí; y la repetición terminal invertida flanqueante 3' tiene una secuencia correspondiente a los nucleótidos 4307-4436 de la SEQ ID NO:44 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí. Los expertos en la técnica entenderán que, para cualquier par dado de secuencias de repeticiones terminales invertidas en un plásmido de transferencia que se usa para crear el vector viral (típicamente al transfectar células con ese plásmido junto con otros plásmidos que llevan el AAV de genes necesario para la formación de vectores virales) (por ejemplo, la SEQ ID NO: 44), que la secuencia correspondiente en el vector viral puede alterarse debido a que las ITR adoptan una orientación "flip" o "flop" durante la recombinación. Por tanto, la secuencia de la ITR en el plásmido de transferencia no es necesariamente la misma secuencia que se encuentra en el vector viral preparado a partir de este. Sin embargo, en algunas modalidades muy específicas, el primer miembro del sistema de vector dual incluye los nucleótidos 1-4436 de la SEQ ID NO:44.

En las Tabla 4 y 5 se proporcionan plásmidos de transferencia que pueden usarse para producir los vectores de ácido nucleico para usar en las composiciones y métodos descritos en la presente. Un plásmido de transferencia (por ejemplo, un plásmido que contiene una secuencia de ADN para ser administrada por un vector de ácido nucleico, por ejemplo para ser administrado por un AAV) se puede co-administrar en células productoras con un plásmido auxiliar (por ejemplo, un plásmido que proporciona las proteínas necesarias para la fabricación de AAV) y un plásmido rep/cap (por ejemplo, un plásmido que proporciona proteínas de la cápside de AAV y proteínas que insertan la secuencia de ADN del plásmido de transferencia en la cubierta de la cápside) para producir un vector de ácido nucleico (por ejemplo, un vector de AAV) para su administración. Los vectores de ácido nucleico (por ejemplo, un vector de ácido nucleico (por ejemplo, un vector de AAV) que contiene un polinucleótido que codifica una parte de extremo N de una proteína de estereocilina y un vector de ácido nucleico (por ejemplo, un vector de AAV) que contiene un polinucleótido que codifica una parte de extremo C de una proteína de estereocilina) se puede combinar (por ejemplo, en una sola formulación) antes de la administración.

Los plásmidos de transferencia que pueden utilizarse para producir vectores de ácido nucleico (por ejemplo, vectores AAV) para la coformulación o coadministración (por ejemplo, la administración simultánea o secuencial) en un sistema de vector dual superpuesto se proporcionan en la Tabla 4 (SEQ ID NO: 43 y SEQ ID NO: 44).

Tabla 4. Plásmidos de transferencia diseñados para producir vectores duales superpuestos

SEQ ID NO.	Descripción	Secuencia de plásmido
43	Plásmido P959	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAG CCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTG ITR 5' en las posiciones de los nucleótidos 1-130 AGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC ACTAGGGGTTCTTGTAGTTAATGATTAACCCGCCATGCTA CTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTTCGC CCTTAAGCTAGCGGCGCGCCACCGGTAGCAGGTTTGTTA Promotor de OCM murino en las posiciones de los nucleótidos 225-1364 CAGAAACCTTAGTTAAGGTTTGTTGAGGGTTTTTTTTCTCT CTCTCTCTCTTAATTGGCTGTCCCAATCCATCCTTCTATAA ATAGAAAAGAGAGACAGGGAGTGTGTGTGGTTTCATTACT AAGGTAAAGACACTTGAGCTACACACACTTGATCCCTGAA CATGAAATCTAAGAGGTTGAACGATCACAGTTTCAGGACT Secuencia codificante de STRC de extremo N en las posiciones de los nucleótidos 1375-4574 ATATAAGGTGGTGAAAGACCATCTGCTTCGTTTTTCTGTTT NGTTCCTACAACCTCTTCCCTCCGCTTGATTTTAACTCTAAAT TGGTGAGTAGCTGGTGGGCTCACCAGACTCCGAGATCCT CTTCTCTGCACGCACTGTATTAGACTTGGCACCCGGGAGG (que incluye la superposición en las posiciones de los nucleótidos 4075-4574 con P724) ATTTTCACCTCTGCTGCATGGGCTAATCTTCCACAAGGGAT CTGTGGTATTGCAATCTCGGGTTGATGCATGACGGTGATG TTGTGTTTATAGCATGGCTAAGGTTTAGCTGCCTATGATGA TTGGTTAGGGAAGGATAATTTTGGCTAGAAGATTGGACTTT AGGGAAAAAAACCCCACTTTTATTTGCTTTTAGAATTTTAA AAGACTGGGCCATGTAGCTCAGGCTGGTTTGGAGTTCATT ITR 3' en las posiciones de los nucleótidos 4662-4791 ATGTAGTCAAGGATGCTCTTGGACTCTTTCAGCATCCTCCTC CTCCTCTTTTTCTCCTCCTCCTTCTTGTCTTCTTCTTGT CTTCCTCTTCCCCTTCTTCCCCCTTCTCTCCCTCTTCCT CCTCTTCCTCCTCCTTCTTGTCTTCTTCCCTCTTCCCCTTCT

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CCTCTCCCCCTTGTCTCCTCTTCCTTCTCCTCCTCCTCCT CTTCTTCTTTCTGAGTACCAAGATTGCAAGTGTGCACACGA TGACCAGCTTGGTCTTTCTTTGTCTTTTTTTTTTAACTTCAA TTTTGGAGTGAATTCAAGAGCAACCATGTAGTCAAGAGGT GGCTGGAGTCTTTTCTGTATCTGGGTTTGGTTTAGTACTCT GCCCCATCACTTAACAGGTCCTTATGGCCACATCTTAAAAA AATTCTAGAGATACACGGTGTCGGTGAGTGGCTGAGAATG TGTGGTCTTCCCATTCTCTGTACCGTGGCTCACATCTTG TTTCTCTGTTCGGCCAGGTAGAAAGGCGGCCGCCATGG CTCTGAGCCTCCAGCCCCAGCTGCTCCTTCTCCTGTGCT CCTGCCGCAGGAAGTGACTTCAGCCCCTACTGGGCCTCA GTCTTTGGATGCTGGTCTCTCCCTTCTGAAGTCATTCGTAG CCACTCTGGACCAAGCTCCTCAGCGTTCCCTCAGCCAGTC ACGGTTCTCTGCGTTCCTGGCCAACATTTCTTCATCCTTCC AGCTTGGGAGGATGGGGGAGGGACCGGTGGGAGAGCCC CCACCTCTCCAGCCCCCTGCACTTCGACTTCATGATTTCC TCGTGACACTGAGAGGTAGCCCAGACTGGGAGCCAATGC TAGGGCTTCTGGGAGATGTGCTGGCACTCCTGGGACAGG AACAGACTCCCCGGGACTTTTTTGGTGCACCAGGCAGGTGT ACTGGGTGGACTTGTAGAGGCATTGTTGGGAGCGTTAGTT CCTGGAGGCCCCCCTGCCCCCACTCGACCCCCATGCACC CGTGATGGCCCTTCTGACTGTGTCCTGGCTGCTGATTGGT TGCCTTCTCTGATGTTGTTATTAGAGGGTACACGCTGGCA GGCCCTGGTGCAGTTGCAGCCCAGTGTGGACCCAACCAA TGCCACAGGTCTTGATGGTAGAGAGCCAGCTCCTCACTTT TTACAGGGTCTGCTGGGCTTGCTTACCCAGCAGGAGAGT TGGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGCTGCGCA CAGTGGGGGGCCCCCCTCTATGCTGCCTTCCAGGAGGGGC TACTGCGAGTCACTCATTCTCTGCAAGATGAGGTCTTTTCT ATTATGGGACAGCCAGAGCCTGATGCCAGTGGGCAGTGC CAGGGAGGCAACCTTCAACAGCTGCTTTTATGGGGCATGC
10		
15		
20		
25		
30		
35		

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GGAACAACCTTTCTTGGGACGCCCCGAGCACTGGGTTTTCT ATCTGGATCACCACCTCCACCCCCTGCTCTCCTGCACTGC CTGAGCAGAGGTGTGCCTCTGCCCAGGGCTTCCCAGCCT GCGGCTCACATCAGCCCTCGACAGCGGCGAGCCATCTCT GTGGAGGCCCTCTGCGAGAACCACTCAGGCCCAGAGCCA CCCTACAGCATCTCCAATTCTCCATCTACTTGCTCTGCCA
10		GCACATCAAGCCTGCCACCCCGCGGCCCCCTCCTACCAC CCCACGGCCTCCTCCTACCACCCACAGCCCCCTCCTAC CACTACACAGCCCATTCTGACACTACACAGCCCCCTCCT GTCACCCCAAGGCCTCCTCCTACCACCCACAACCCCTC CTAGCACAGCTGTCATCTGCCAGACAGCTGTATGGTACGC
15		AGTCTCGTGGGCACCAGGTGCCCGAGGTTGGCTCCAAGC CTGCCATGATCAGTTTCCTGATCAATTTCTGGATATGATCT GCGGCAACCTCTCATTTTCAGCCCTGTCTGGCCCCAGTCG TCCTTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCCACCC
20		CCCACTAGCTGTCCACCAGGCCTGATCCCTGTGCCCTCA CCCCAGAAATATTCTGGGGCTGTTTCCTGGAGAATGAGAC ACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTCTGCA GGCTGTGCCCCCGAGGAACCAGGCTTGGGTTTCAGCATGT
25		GTGTGCGGGGCCCCACCTTGGACGCCACTGATTTTCCACC GTGCCGCGTTGGACCCTGTGGGGAACGCTGCCCAGATGG GGGCAGCTTCCTGCTCATGGTCTGTGCCAATGACACTCTG TATGAAGCCTTGGTTCCCTTCTGGGCTTGGCTAGCAGGCC
30		AATGCAGAATTAGTCGTGGAGGAAATGATACTTGCTTTCTA GAAGGCATGCTGGGCCCCCTTGTTGCCCTCTCTGCCCCCT CTGGGACCATCCCCACTCTGTCTGGCTCCTGGTCCTTTTC TGCTTGGCATGTTATCCCAGTTGCCACGCTGTCAGTCCTC
35		CGTGCCAGCCCTCGCCACCCACGCGCCTACATTACCT CCTGCGCCTACTGACCTTCCTTCTGGGTCCAGGGACTGG GGGTGCCGAGACGCAGGGGATGTTAGGTCAAGCCCTGCT GCTCTCTAGTCTCCCAGACAACTGTTCAATTCTGGGATGCC

SEQ ID NO.	Descripción	Secuencia de plásmido
5		TTCCGCCCAGAGGGCCGGAGAAGTGTACTGAGGACAGTC GGAGAGTACTTGCAGCGGGAAGAGCCAACCCCACCAGGC TTAGACTCCTCCCTCAGCCTCGGCTCTGGTATGAGCAAGA TGGAGCTTCTGTCCTGCTTCAGTCCTGTACTGTGGGATCT ACTCCAGAGAGAGAAGAGCGTTTGGGCCCTGAGGACCCT GGTGAAGGCCTACCTGCGCATGCCTCCAGAAGACCTTCA 10 GCAGCTTGTGCTTTCAGCAGAGATGGAGGCTGCACAGGG CTTCCTGACGCTCATGCTTCGTTCCCTGGGCTAAGCTGAAG GTTCAACCATCCGAGGAGCAGGCCATGGGCCCGCCTGACA GCCTTGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAA CTCTTTATCGACATGTCACCGCTCATCCCCTTCCTGGCTGT 15 CCCTGACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAAC GACAGTGTCTGCTGGCTGCCATCAGGGATCACAGCTCAGGA ATGAAGCCTGAACAGAAGGAGGCCCTGGCAAACGACTG CTGGCCCCTGAGCTGTTTGGAGAAGTGCCTGATTGGCCC CAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCTTCCCCAT 20 CTGCCTCTGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGA TCCAGGCCCTGGAGGATAGCTGGCCAGTAGCAGATCTTG GGCCGGGACACGCCCCGACATGTGCTTCGTAGCCTAGTAA ACCAGAGCATGGAGGATGGGGAGGAGCAGGTGCTCAGG CTTGGGTCCCTCGCCTGTTTCCTGAGTCCTGAGGAGCTAC 25 AGAGTCTGGTGCCCTTGAGTGATCCAATGGGGCCTGTAGA ACAGGGTCTGCTGGAATGTGCGGCCAATGGGACCCTCAG CCCAGAAGGACGGGTGGCATATGAACTTCTGGGAGTGTT GCGTTCATCTGGAGGAACTGTCTTAAGCCCCCGAGAGCTG AGGGTCTGGGCACCTCTCTTTCCCCAGCTGGGCCTCCGC 30 TTCCTGCAGGAGCTCTCAGAGACCCAGCTTAGAGCCATGC TTCCTGCCCTACAGGGAGCCAGTGTACACCCTCGAGTTA AGGGCGAATTCCCGATAAGGATCTTCCTAGAGCATGGCTA CGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAG 35 GAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGC

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACGCCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGA GCGAGCGCGCAGCCTTAATTAACCTAATTCACTGGCCGTC GTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC AACTTAATCGCCTTG CAGCACATCCCCCTTTCGCCAGCTG GCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCA 10 ACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTG TAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCG CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCCGC TCCTTTCGCTTTCTTCCCTTCCTTTCGCCACGTTGCGCG 15 GCTTTCCTCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGG GTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAA CTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCT GATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTT CTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA 20 ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTG CCGATTTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACA AAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAAT TTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA TTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCA TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAA 25 AGGAAGAGTATGAGCCATATTCAACGGGAAACGTGAGG CCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTA TAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACA ATCTATCGCTTGATGGGAAGCCCGATGCGCCAGAGTTGT TTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTAC 30 AGATGAGATGGTCAGACTAACTGGCTGACGGAATTTATG CCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGA TGCATGGTTACTCACCCTGCGATCCCCGAAAAACAGCA TTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATAT 35 TGTTGATGCGCTGGCAGTGTTCTGCGCCGGTTGCATTG

SEQ ID NO.	Descripción	Secuencia de plásmido
5		ATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATT TCGTCTTGCTCAGGCGCAATCACGAATGAATAACGGTTTG GTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGC CTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCA TTCTCACCGGATTTCAGTCGTCACTCATGGTGATTTCTCACT TGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTA 10 TTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGA TCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTTCTCCTT CATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAAT CCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGA GTTTTTCTAACTGTCAGACCAAGTTTACTCATATATACTTTA 15 GATTGATTTAAACTTCATTTTTAATTTAAAAGGATCTAGGT GAAGATCCTTTTTGATAATCTCATGACCAAATCCCTTAAC GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAA GATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGT 20 GGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGA AGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATACTGT TCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAC TCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT ACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACC 25 GGGTTGGAAGTCAAGACGATAGTTACCGGATAAGGCGCAG CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAG CGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGA AAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACA 30 GGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG GTATCTTTATAGTCCTGTGCGGGTTTCGCCACCTCTGACTTG AGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCC TATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGT 35

SEQ ID NO.	Descripción	Secuencia de plásmido
5		TATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAG TGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGC AGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATA CGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAAT GCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCA GTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATT 10GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATG TTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAA ACAGCTATGACCATGATTACGCCAGATTTAATTAAGGCCTT AATTAGG
15	44 Plásmido P724	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAG CCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTG ITR 5' en las posiciones de los nucleótidos 1-130
20		AGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC ACTAGGGGTTCTTGTAGTTAATGATTAACCCGCCATGCTA CTTATCTACGTAGCCATGCTCTAGGAAGATCGGAATTCGC CCTTAAGCTAGCGCCTCGGCTCTGGTATGAGCAAGATGGA Secuencia codificante de STRC de extremo C
25		CCAGAGAGAGAAGAGCGTTTGGGCCCTGAGGACCCTGGTG en las posiciones de nucleótidos 211-3440
30		AAGGCCTACCTGCGCATGCCTCCAGAAGACCTTCAGCAG CTTGTGCTTTCAGCAGAGATGGAGGCTGCACAGGGCTTCC (que incluye la superposición en las posiciones de nucleótidos 211-710 con P959)
35		TGACGCTCATGCTTCGTTCCCTGGGCTAAGCTGAAGGTTCA ACCATCCGAGGAGCAGGCCATGGGCCCGCCTGACAGCCTT GCTGCTCCAGCGGTACCCACGCCTCACCTCCCAACTCTTT ATCGACATGTCACCGCTCATCCCCTTCCTGGCTGTCCCTG ACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAACGACAG TGTCCCTGGCTGCCATCAGGGATCACAGCTCAGGAATGAAG CCTGAACAGAAGGAGGCCCTGGCAAAACGACTGCTGGCC de los nucleótidos 3452-3999
		CCTGAGCTGTTTGGAGAAGTGCCTGATTGGCCCCAGGAG CTGCTGTGGGCAGCCCTGCCTCTGCTTCCCCATCTGCCTC TGGAGAGCTTTCTCCAGCTCAGCCCTCACCAGATCCAGGC poli(A) de bGH en las
		CCTGGAGGATAGCTGGCCAGTAGCAGATCTTGGGCCGGG

SEQ ID NO.	Descripción	Secuencia de plásmido
5	posiciones de los nucleótidos 4012-4219	ACACGCCCCGACATGTGCTTCGTAGCCTAGTAAACCAGAGC ATGGAGGATGGGGAGGAGCAGGTGCTCAGGCTTGGGTCC CTCGCCTGTTTCCTGAGTCCTGAGGAGCTACAGAGTCTGG
10	ITR 3' en las posiciones de los nucleótidos 4307- 4436	TGCCCTTGAGTGATCCAATGGGGCCTGTAGAACAGGGTCT GCTGGAATGTGCGGCCAATGGGACCCTCAGCCCAGAAGG ACGGGTGGCATATGAACTTCTGGGAGTGTTGCGTTCATCT GGAGGAACTGTCTTAAGCCCCGAGAGCTGAGGGTCTGG GCACCTCTCTTTCCCCAGCTGGGCCTCCGCTTCCTGCAGG AGCTCTCAGAGACCCAGCTTAGAGCCATGCTTCCTGCCCT ACAGGGAGCCAGTGTACACCTGCCCAGGCTGTTCTGTT GTTTGGAAGGCTCCTTCCTAAGCATGATCTGTCCCTGGAG 15 GAACTCTGCTCCCTGCACCCTCTCCTGCCAGGTCTCAGCC CCCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTGG TGCTTGTTCTCCTGCCTGGGCCCTGAACTGTCAAGGCTTTCA GCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTCGGGTAA AAGATGGTGTATAAAAATATGGGTGCAGCAGGTGCCGGCTC 20 AGCCGTGTGCATTCTGGGCAGCCCACCACTTGGCCAGA CTGCCTGCTTCCCCTGCTCCCATTAAGCTGCTACAGCTG GACGCTGCAGCTCTTCTGGCAAACCGAAGACTCTATCGGC AGCTGCCTTGGTCTGAGCAACAGGCACAGTTTCTCTGGAA GAAAATGCAAGTGCCTACCAACCTGAGCCTGAGGAATCTG 25 CAGGCTCTGGGCAACTTGGCAGGAGGCATGACCTGCGAG TTTCTGCAGCAGATCAGCTCAATGGTTGACTTTCTTGATGT GGTACACATGCTCTACCAGCTGCCCCACTGGTGTTCGAGAG AGCCTGCGGGCCTGTATCTGGACAGAGCTACAGCGGAGG ATGACAATGCCAGAGCCAGAGCTGACCACCCTAGGGCCA 30 GAACTGAGTGAAGTTGACACAAAGCTACTCCTGGACTTGC CGATCCAGCTGATGGACAGATTGTCCAATGATTCCATTAT GTTGGTGGTGGAGATGGTCCAAGGCGCTCCAGAGCAGCT GCTGGCACTGACCCCACTCCACCAGACAGCCTTGGCAGA GCGAGCACTTAAAAACCTGGCTCCAAAGGAGACCCCAATC 35

SEQ ID NO.	Descripción	Secuencia de plásmido
5		TCCAAAGAAGTGCTGGAGACACTGGGCCCCTTGTTGGAT TCCTGGGAATAGAGAGCACGCGACGGATCCCTTTACCCAT TCTACTGTCTCATCTCAGTCAGCTGCAGGGCTTCTGCCTA GGAGAGACATTTGCCACAGAGCTGGGATGGCTGCTGTTG CAGGAGCCTGTTCTTGAAAACCAGAATTGTGGAGCCAGG ATGAAATAGAGCAAGCTGGACGCCTAGTATTCACTCTGTC
10		TGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGGGC CCAGAGACACTGGAGAGGCTTCTGGGAAAGCATCAAAGC TGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGGAGTCA CAGCTTGCCCAACAAGAAAGCAGCTCTGGTAGCTGGGATTG TGCATCCAGCTGCTGAGGGTCTCCAAGAGCCTGTACCAAA
15		CTGTGCAGACATACGGGGAACCTTCCCAGCGGCCTGGTC TGCGACACAAATCTCAGAGATGGAACCTCTCAGACTTTGAA GACTGCCTGTCACTATTTGCTGGAGATCCAGGACTTGGTC CTGAGGAACTACGGGCAGCCATGGGCAAGGCCAAGCAGT TGTGGGGTCCCCCTCGAGGATTCCGTCCTGAGCAGATCTT
20		GCAGCTGGGCCGTCTCCTGATAGGTCTAGGAGAACGGGA ACTGCAGGAGCTTACCTTGGTGGACTGGGGTGTGCTGAG CAGCCTGGGGCAAATAGATGGCTGGAGTTCCATGCAGCT CCGAGCCGTGGTCTCCAGTTTCCTAAGGCAGAGTGGTCG GCATGTGAGCCACCTGGACTTCATTTATCTGACAGCACTG
25		GGTTACACAGTCTGTGGATTGCGACCAGAGGAGTTACAGC ACATCAGCAGTTGGGAGTTTAGCCAAGCAGCTCTCTTCCT GGGTAGCTTGCACTCTCCCGTGCTCTGAGGAACAGCTGGA AGTTCTGGCCTATCTCCTTGTGTTGCCTGGTGGCTTTGGC CCAGTCAGTAACTGGGGGCCTGAGATCTTCACTGAAATTG
30		GCACAATAGCAGCTGGCATCCCAGACCTGGCTCTTTCAGC ATTACTGCGGGGACAGATCCAAGGCCTGACTCCTCTTGCC ATTTCTGTCATTCTGCTCCCAAGTTTGCAGTGGTCTTCAA CCCCATCCAGTTATCTAGTCTCACCAGGGGTCAGGCCGTA GCTGTTACTCCTGAACAGCTGGCCTATCTGAGTCCTGAGC
35		

SEQ ID NO.	Descripción	Secuencia de plásmido
5		AGCGGCGAGCAGTTGCATGGGCCCCAACACGAAGGGAAGG AGATCCCAGAGCAGCTGGGTGCGAACTCAGCCTGGGGTC TCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGCATTGCC CGTCAGCATTTTTGGCCACCTATTATGATAATAAGCTTGGA TCCAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGAC TGGTATTCTTAAGTATGTTGCTCCTTTTACGCTATGTGGAT 10 ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCT GTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGT GGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACT GGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGG 15 ACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCA TCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGC TGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATC ATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGG ATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCC 20 TCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGG CTCTGCGGCCTCTTCCGCGTCTTCGAGATCTGCCTCGACT GTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGT CCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA 25 GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG ACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATG CTGGGGACTCGAGTTAAGGGCGAATTCCCGATAAGGATCT TCCTAGAGCATGGCTACGTAGATAAGTAGCATGGCGGGTT AATCATTAAGTACAAGGAACCCCTAGTGATGGAGTTGGCC 30 ACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGG CGACCAAAGGTGCGCCGACGCCCCGGGCTTTGCCCGGGC GGCCTCAGTGAGCGAGCGAGCGCGCAGCCTTAATTAACC TAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAA AACCCTGGCGTTACCCAACCTAATCGCCTTGCAGCACATC 35

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCA CCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCG AATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGG GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCA GCGCCCTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCCTT TCTCGCCACGTTGCGCGGCTTCCCCGTCAAGCTCTAAAT CGGGGGCTCCCTTTAGGGTTCGATTTAGTGCTTTACGGC ACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACG TAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCA AACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTG ATTTATAAGGGATTTTGCCGATTTGCGCCTATTGGTTAAAA AATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAA AATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAA TGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATT CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG CTTCAATAATATTGAAAAAGGAAGAGTATGAGCCATATTCA ACGGGAAACGTGAGGCCGCGATTAAATTCCAACATGGAT GCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCG GGCAATCAGGTGCGACAATCTATCGCTTGATGGGAAGCC CGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGC GTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACT GGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTT ATCCGTA CTCTGATGATGCATGGTTACTCACCCTGCGA TCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCC TGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTT CTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTT TAACAGCGATCGCGTATTTGCTCTTGCTCAGGCGCAATCA CGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATG ACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA AATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCA
10		
15		
20		
25		
30		
35		

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAG GGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAA TCGCAGACCGATACCAGGATCTTGCCATCCTATGGAAGT CCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTTC AAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAG TTTTATTTGATGCTCGATGAGTTTTTCTAACTGTCAGACCA 10 AGTTTACTCATATATACTTTAGATTGATTTAAACTTCATTTT TAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA 15 TCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGA GCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGA GCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTT AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATAC 20 CTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATA GTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGG GTTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA CCGAAGTGAATACCTACAGCGTGAGCTATGAGAAAGCGC 25 CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGC TTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCT CGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG 30 CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT CACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAA CCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAA GCGGAAGAGCGCCAATACGCAAACCGCCTCTCCCCGCG 35 CGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTC

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATG TGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTT ATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT AACAATTTACACAGGAAACAGCTATGACCATGATTACGC CAGATTTAATTAAGGCCTTAATTAGG
10		

Vectores duales de transempalme

Un segundo enfoque para expresar proteínas de gran tamaño en células de mamífero implica utilizar vectores duales de transempalme. En este enfoque, se utilizan dos vectores de ácido nucleico que contienen secuencias de ácido nucleico distintas, y el polinucleótido que codifica la parte de extremo N de la proteína de interés y el polinucleótido que codifica la parte de extremo C de la proteína de interés no se superponen. En su lugar, el primer vector de ácido nucleico incluye una secuencia donante de empalme 3' del polinucleótido que codifica la parte de extremo N de la proteína de interés, y el segundo vector de ácido nucleico incluye una secuencia aceptora de empalme 5' del polinucleótido que codifica la parte de extremo C de la proteína de interés. Cuando el primer y segundo ácido nucleico están presentes en la misma célula, sus ITR pueden concatenarse, formando una única estructura de ácido nucleico en la que los ITR concatenados se sitúan entre el donante de empalme y el aceptor de empalme. El transempalme ocurre entonces durante la transcripción, produciendo una molécula de ácido nucleico en la que los polinucleótidos que codifican las partes de extremo N y de extremo C de la proteína de interés son contiguos, formando así la secuencia codificante de longitud completa.

Los vectores duales de transempalme para usar en los métodos y composiciones descritos en la presente también están diseñados de manera que aproximadamente la mitad de la secuencia codificante que codifica la estereocilina esté contenido dentro de cada vector (por ejemplo, cada vector contiene un polinucleótido que codifica aproximadamente la mitad de la proteína estereocilina, tal como se discute anteriormente). La determinación de cómo dividir la secuencia de polinucleótidos entre los dos vectores de ácido nucleico se realiza basándose en el tamaño del promotor y las ubicaciones de los elementos de secuencia de interés en el polinucleótido que codifican la proteína de estereocilina (por ejemplo, exones del gen de STRC). El primer vector del sistema de vector dual de transempalme puede contener una secuencia promotora 5' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina. Los vectores de

ácido nucleico pueden contener opcionalmente UTR de STRC (por ejemplo, tanto las UTR 5' como 3' de STRC, por ejemplo, UTR de longitud completa). Un sistema de vector dual de transempalme de ejemplo para su uso en las composiciones y métodos descritos en la presente incluye un primer vector de ácido nucleico que contiene un promotor OCM (por ejemplo, cualquiera de las SEQ ID NO: 1-3) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de una proteína estereocilina humana, por ejemplo, una parte de extremo N de la SEQ ID NO: 4) y una secuencia donante de empalme 3' de la secuencia de polinucleótidos; y un segundo vector de ácido nucleico que contiene una secuencia aceptora de empalme 5' de un polinucleótido que codifica una parte de extremo C de la proteína estereocilina (por ejemplo, una parte de extremo C de estereocilina humana, por ejemplo, una parte de extremo C de la SEQ ID NO: 4) y una secuencia poli(A). Un sistema de vector dual de transempalme alternativo incluye un primer vector de ácido nucleico que contiene un promotor OCM (por ejemplo, cualquiera de las SEQ ID NO: 1-3) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de la proteína estereocilina (por ejemplo, una parte de extremo N de una proteína estereocilina murina, por ejemplo, una parte de extremo N de la SEQ ID NO: 5) y una secuencia donante de empalme 3' de la secuencia de polinucleótidos; y un segundo vector de ácido nucleico que contiene una secuencia aceptora de empalme 5' de un polinucleótido que codifica una parte de extremo C de la proteína estereocilina (por ejemplo, una parte de extremo C de una proteína estereocilina murina, por ejemplo, una parte de extremo C de la SEQ ID NO: 5) y una secuencia poli(A). En algunas modalidades, el promotor de OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 1 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 1. En algunas modalidades, el promotor OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 2 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 2. En algunas modalidades, el promotor OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 3 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 3. Estos vectores de ácido nucleico también pueden contener UTR 5' y/o 3' de STRC de longitud completa en el primer y segundo vector de ácido nucleico, respectivamente (por ejemplo, el primer vector de ácido nucleico puede contener la UTR 5' humana de STRC en sistemas de vectores duales que codifican estereocilina humana, o la UTR 5' de ratón en sistemas de vectores duales que codifican estereocilina de ratón; y el segundo vector de ácido nucleico puede contener la UTR 3' humana de STRC en sistemas de vectores duales que

codifican estereocilina humana, o la UTR 3' de ratón de STRC en sistemas de vectores duales que codifican estereocilina de ratón). Para acomodar una UTR de STRC, la secuencia codificante de la estereocilina puede dividirse en una posición tal que acomode la longitud de la secuencia promotora y la secuencia que codifica la parte de extremo N de la estereocilina.

- 5 En algunas modalidades, el polinucleótido que codifica una proteína estereocilina humana de longitud completa tiene la secuencia de la SEQ ID NO: 6 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 6. En algunas modalidades, el polinucleótido que codifica una proteína estereocilina murina de longitud
- 10 completa tiene la secuencia de la SEQ ID NO: 7 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 7.

Vectores híbridos duales

- 15 Un tercer enfoque para expresar proteínas grandes en células de mamífero implica el uso de vectores híbridos duales. Este enfoque combina elementos de la estrategia de vector dual superpuesto y de la estrategia de transempalme, en el sentido de que presenta tanto una región superpuesta en la que puede producirse recombinación homóloga como secuencias donantes y
- 20ceptoras de empalme. En los sistemas de vectores híbridos duales, la región superpuesta es una región recombinogénica que está contenida tanto en el primer vector como en el segundo vector de ácido nucleico, en lugar de una parte de la secuencia de polinucleótidos que codifica la proteína de interés – el polinucleótido que codifica la parte de extremo N de la proteína de interés y el polinucleótido que codifica la parte de extremo C de la proteína de interés no se superponen en este enfoque. La región recombinogénica es 3' de la secuencia donante de empalme en el primer
- 25vector de ácido nucleico y 5' de la secuencia aceptora de empalme en la segunda secuencia de ácido nucleico. La primera y segunda secuencia de ácido nucleico pueden unirse para formar una única secuencia basándose en uno de estos dos mecanismos 1) recombinación en la región superpuesta, o 2) concatenación de las ITR. La región o regiones recombinogénicas restantes y/o las ITR concatemerizadas pueden eliminarse mediante empalme, lo que conduce a la formación
- 30de una secuencia de polinucleótidos contigua que codifica la proteína de interés de longitud completa. Las regiones recombinogénicas, las secuencias donantes de empalme y las secuencias aceptoras de empalme que pueden usarse en las composiciones y métodos descritos en la presente incluyen las bien conocidas por un experto en la técnica. Las regiones recombinogénicas de ejemplo incluyen el gen AK del fago F1 y fragmentos del gen de la fosfatasa alcalina (AP) tal
- 35como se describe en las patentes estadounidenses n.º 10.494.645 y 8.236.557, que se incorporan

en la presente mediante referencia. En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

5 CCCCAGGTGCGCGGCGTCGGTGGTGCCGGCGGGGGGCGCCAGGTCGCAGGCGGTGTAGG
 GCTCCAGGCAGGCGGCGAAGGCCATGACGTGCGCTATGAAGGTCTGCTCCTGCACGCCGTG
 AACCAGGTGCGCCTGCGGGCCGCGCGCGAACACCGCCACGTCCTCGCCTGCGTGGGTCTCT
 TCGTCCAGGGGCACTGCTGACTGCTGCCGATACTCGGGGCTCCCGCTCTCGCTCTCGGTAAC
 ATCCGGCCGGGCGCCGTCCTTGAGCACATAGCCTGGACCGTTTCCGTATAGGAGGACCGTGT
 AGGCCTTCCTGTCCCGGGCCTTGCCAGCGGCCAGCCCGATGAAGGAGCTCCCTCGCAGGGG
 10 GTAGCCTCCGAAGGAGAAGACGTGGGAGTGGTCGGCAGTGACGAGGCTCAGCGTGTCTCC
 TCGCTGGTGAGCTGGCCCGCCCTCTCAATGGCGTCGTGGAACATGATCGTCTCAGTCAGTGC
 CCGGTAAGCCCTGCTTTCATGATGACCATGGTCGATGCGACCACCCTCCACGAAGAGGAAGA
 AGCCGCGGGGGTGTCTGCTCAGCAGGCGCAGGGCAGCCTCTGTCTCATCTCCATCAGGGAGGG
 GTCCAGTGTGGAGTCTCGGTGGATCTCGTATTTTCATGTCTCCAGGCTCAAAGAGACCCATGAG
 15 ATGGGTCACAGACGGGTCCAGGGAAGCCTGCATGAGCTCAGTGCGGTTCCACACGTACCGG
 GCACCCTGGCGTTCCCGAGCCATTCCTGCACCAGATTCTTCCCGTCCAGCCTGGTCCCACC
 TTGGCTGTAGTCATCTGGGTACTCAGGGTCTGGGGTTCCCATGCGAAACATGTACTTTTCGGCC
 TCCA (SEQ ID NO: 47).

20 En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

CCCCAGGTGCGCGGCGTCGGTGGTGCCGGCGGGGGGCGCCAGGTCGCAGGCGGTGTAGG
 GCTCCAGGCAGGCGGCGAAGGCCATGACGTGCGCTATGAAGGTCTGCTCCTGCACGCCGTG
 AACCAGGTGCGCCTGCGGGCCGCGCGCGAACACCGCCACGTCCTCGCCTGCGTGGGTCTCT
 25 TCGTCCAGGGGCACTGCTGACTGCTGCCGATACTCGGGGCTCCCGCTCTCGCTCTCGGTAAC
 ATCCGGCCGGGCGCCGTCCTTGAGCACATAGCCTGGACCGTTTCCGTATAGGAGGACCGTGT
 AGGCCTTCCTGTCCCGGGCCTTGCCAGCGGCCAGCCCGATGAAGGAGCTCCCTCGCAGGGG
 GTAGCCTCCGAAGGAGAAGACGTGGGAGTGGTCGGCAGTGACGAGGCTCAGCGTGTCTCC
 TCG CTGGTGA (SEQ ID NO: 48).

30

En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

GCTGGCCCGCCCTCTCAATGGCGTCGTGGAACATGATCGTCTCAGTCAGTGCCCGGTAAGCC
 CTGCTTTCATGATGACCATGGTCGATGCGACCACCCTCCACGAAGAGGAAGAAGCCGCGGGG
 35 GTGTCTGCTCAGCAGGCGCAGGGCAGCCTCTGTCTCATCTCCATCAGGGAGGGGTCCAGTGTG

GAGTCTCGGTGGATCTCGTATTTTCATGTCTCCAGGCTCAAAGAGACCCATGAGATGGGTGACA
 GACGGGTCCAGGGAAGCCTGCATGAGCTCAGTGCGGTTCCACACGTACCGGGCACCCCTGGC
 GTTCGCCGAGCCATTCTGCACCAGATTCTTCCCGTCCAGCCTGGTCCCACCTTGGCTGTAGT
 CATCTGGGTACTCAGGGTCTGGGGTTCCTCATGCGAAACATGTACTTTCGGCCTCCA (SEQ ID
 NO: 49).

En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

CCCCGGGTGCGCGGCGTCCGGTGGTGCCGGCGGGGGGCGCCAGGTGCGAGGCGGTGTAGG
 GCTCCAGGCAGGCGGCGAAGGCCATGACGTGCGCTATGAAGGTCTGCTCCTGCACGCCGTG
 AACCAGGTGCGCCTGCGGGCCGCGCGCGAACACCGCCACGTCTCGCCTGCGTGGGTCTCT
 TCGTCCAGGGGCACTGCTGACTGCTGCCGATACTCGGGGCTCCCGCTCTCGCTCTCGGTAAC
 ATCCGGCCGGGCGCCGTCCTTGAGCACATAGCCTGGACCGTTTC (SEQ ID NO: 50)

En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

CGTATAGGAGGACCGTGTAGGCCTTCCTGTCCCGGGCCTTGCCAGCGGCCAGCCCGATGAA
 GGAGCTCCCTCGCAGGGGGTAGCCTCCGAAGGAGAAGACGTGGGAGTGGTCGGCAGTGACG
 AGGCTCAGCGTGTCTCCTCGCTGGTGAGCTGGCCCCGCCCTCTCAATGGCGTCGTGGAACAT
 GATCGTCTCAGTCAGTGCCCGGTAAGCCCTGCTTTCATGATGACCATGGTCGATGCGACCACC
 CTCCACGAAGAGGAAGAAGCCGCGGGGGTGTCTGCTCAGCAGG (SEQ ID NO: 51).

En algunas modalidades, el fragmento del gen AP tiene la secuencia de:

CGCAGGGCAGCCTCTGTCATCTCCATCAGGGAGGGGTCCAGTGTGGAGTCTCGGTGGATCTC
 GTATTTTCATGTCTCCAGGCTCAAAGAGACCCATGAGATGGGTACAGACGGGTCCAGGGAAG
 CCTGCATGAGCTCAGTGCGGTTCCACACGTACCGGGCACCCCTGGCGTTCCGCCGAGCCATTCC
 TGACACAGATTCTTCCCGTCCAGCCTGGTCCCACCTTGGCTGTAGTCATCTGGGTACTCAGGG
 TCTGGGGTTCCCATGCGAAACATGTACTTTCGGCCTCCA (SEQ ID NO: 52).

Los vectores híbridos duales para usar en los métodos y composiciones descritos en la presente
 también están diseñados de manera que aproximadamente la mitad de la secuencia codificante
 que codifica la estereocilina esté contenido dentro de cada vector (por ejemplo, cada vector
 contiene un polinucleótido que codifica aproximadamente la mitad de la proteína estereocilina). La
 determinación de cómo dividir la secuencia de polinucleótidos entre los dos vectores de ácido

nucleico se realiza basándose en el tamaño del promotor y las ubicaciones de los elementos de secuencia de interés en el polinucleótido que codifican la proteína de estereocilina (por ejemplo, exones del gen de STRC). El primer vector del sistema de vector dual de transempalme puede contener una secuencia promotora 5' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina. Los vectores de ácido nucleico pueden contener opcionalmente UTR de STRC (por ejemplo, UTR 5' y 3' de longitud completa).

Un sistema de vector híbrido dual de ejemplo incluye un primer vector de ácido nucleico que contiene un promotor OCM (por ejemplo, cualquiera de las SEQ ID NO: 1-3) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de una estereocilina humana, por ejemplo, una parte de extremo N de la SEQ ID NO: 4), una secuencia donante de empalme 3' de la secuencia de polinucleótidos, y una región 3' recombinogénica de la secuencia donante de empalme; y un segundo vector de ácido nucleico que contiene una región recombinogénica (por ejemplo, la misma región recombinogénica que está incluida en el primer vector), una secuencia aceptora de empalme 3' de la región recombinogénica, un polinucleótido 3' de la secuencia aceptora de empalme que codifica una parte de extremo C de la proteína estereocilina (por ejemplo, una parte de extremo C de estereocilina humana, por ejemplo, una parte de extremo C de la SEQ ID NO: 4) y una secuencia poli(A). En algunas modalidades, el promotor de OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 1 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 1. En algunas modalidades, el promotor OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 2 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 2. En algunas modalidades, el promotor OCM es un polinucleótido que tiene la secuencia de la SEQ ID NO: 3 o una variante que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 3. El primer y segundo vector de ácido nucleico también pueden contener las UTR 5' y/o 3' de STRC de longitud completa, respectivamente (por ejemplo, el UTR 5' de STRC humana se puede incluir en el primer vector de ácido nucleico, y el UTR 3' de STRC humana se puede incluir en el segundo vector de ácido nucleico). Otro sistema de vector híbrido dual de ejemplo que incluye un promotor de OCM incluye un primer vector de ácido nucleico que contiene un promotor OCM (por ejemplo, cualquiera de las SEQ ID NO: 1-3) unido de forma operativa al polinucleótido que codifica una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de una estereocilina murina, por ejemplo, una parte de extremo N de la SEQ ID NO: 5),

una secuencia donante de empalme 3' de la secuencia de polinucleótidos, y una región 3' recombinogénica de la secuencia donante de empalme; y un segundo vector de ácido nucleico que contiene una región recombinogénica (por ejemplo, la misma región recombinogénica que está incluida en el primer vector), una secuencia aceptora de empalme 3' de la región recombinogénica, un polinucleótido 3' de la secuencia aceptora de empalme que codifica una parte de extremo C de la proteína estereocilina (por ejemplo, una parte de extremo C de estereocilina murina, por ejemplo, una parte de extremo C de la SEQ ID NO: 5) y una secuencia poli(A). El primer y segundo vector de ácido nucleico también pueden contener las UTR 5' y/o 3' de STRC de longitud completa, respectivamente (por ejemplo, el UTR 5' de STRC de ratón se puede incluir en el primer vector de ácido nucleico, y el UTR 3' de STRC de ratón se puede incluir en el segundo vector de ácido nucleico). Para acomodar una UTR de STRC, la secuencia codificante de la estereocilina puede dividirse en una posición diferente a la que tendría en un sistema de vector híbrido dual que no incluya una UTR de STRC.

En algunas modalidades, el polinucleótido que codifica una proteína estereocilina humana de longitud completa tiene la secuencia de la SEQ ID NO: 6 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 6. En algunas modalidades, el polinucleótido que codifica una proteína estereocilina murina de longitud completa tiene la secuencia de la SEQ ID NO: 7 o es una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 7.

Los vectores híbridos duales utilizados en los métodos y composiciones descritos en la presente pueden incluir opcionalmente una secuencia señal de degradación en el primer y en el segundo vector de ácido nucleico. La secuencia señal de degradación puede incluirse para prevenir o reducir la expresión de partes de la proteína estereocilina a partir de polinucleótidos que no se pudieron recombinar y/o fueron sometidos a empalme. La secuencia señal de degradación se sitúa a 3' de la región recombinogénica en el primer vector de ácido nucleico, y se sitúa entre la región recombinogénica y el aceptor de empalme en el segundo vector de ácido nucleico. Las secuencias señal de degradación adecuadas que pueden utilizarse en las composiciones y métodos descritos en la presente son conocidas en la técnica y se describen, por ejemplo, en la publicación de solicitud internacional n.º WO 2016/139321, que se incorpora a la presente mediante referencia.

En algunas modalidades, el primer miembro del sistema de vector dual incluye el promotor OCM de la SEQ ID NO:1 (también representado por los nucleótidos 225-1364 de la SEQ ID NO: 45) unido de forma operativa a los nucleótidos que codifican una parte de extremo N de una proteína estereocilina. En determinadas modalidades, la secuencia de nucleótidos que codifica una parte de

extremo N de una proteína estereocilina son los nucleótidos 1378-4077 de la SEQ ID NO: 45. Las secuencias de nucleótidos que codifican una parte de extremo N de una proteína estereocilina pueden estar optimizadas por codón de forma parcial o completa para la expresión. En algunas modalidades, el primer miembro del sistema de vector dual incluye la secuencia del donante de empalme correspondiente a los nucleótidos 4078-4161 de la SEQ ID NO:45. En algunas modalidades, el primer miembro del sistema de vector dual incluye la secuencia de cabeza AP correspondiente a los nucleótidos 4168-4454 de la SEQ ID NO:45. En modalidades particulares, el primer miembro del sistema de vector dual incluye los nucleótidos 225-4454 de la SEQ ID NO:45 flanqueado en cada uno de los lados 5' y 3' por una repetición terminal invertida. En algunas modalidades, las repeticiones terminales invertidas flanqueantes son cualquier variante de las repeticiones terminales invertidas de AAV2 que pueden encapsidarse mediante un plásmido que porta el gen Rep de AAV2. En determinadas modalidades, la repetición terminal invertida flanqueante 5' tiene una secuencia correspondiente a los nucleótidos 1-130 de la SEQ ID NO:45 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí; y la repetición terminal invertida flanqueante 3' tiene una secuencia correspondiente a los nucleótidos 4548-4677 de la SEQ ID NO:45 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí. Los expertos en la técnica entenderán que, para cualquier par dado de secuencias de repeticiones terminales invertidas en un plásmido de transferencia que se usa para crear el vector viral (típicamente al transfectar células con ese plásmido junto con otros plásmidos que llevan el AAV de genes necesario para la formación de vectores virales) (por ejemplo, la SEQ ID NO: 45), que la secuencia correspondiente en el vector viral puede alterarse debido a que las ITR adoptan una orientación "flip" o "flop" durante la recombinación. Por tanto, la secuencia de la ITR en el plásmido de transferencia no es necesariamente la misma secuencia que se encuentra en el vector viral preparado a partir de este. Sin embargo, en algunas modalidades muy específicas, el primer miembro del sistema de vector dual incluye los nucleótidos 1-4677 de la SEQ ID NO:45.

En algunas modalidades, el segundo miembro del sistema de vector dual incluye nucleótidos que codifican la parte de extremo C de la proteína estereocilina seguido inmediatamente por un codón de terminación. En determinadas modalidades, la secuencia de nucleótidos que codifica la parte de extremo C de la proteína estereocilina son los nucleótidos 615-3344 de la SEQ ID NO:46. Las secuencias de nucleótidos que codifican la parte de extremo C de una proteína estereocilina pueden estar optimizadas por codón de forma parcial o completa para la expresión. En algunas

modalidades, el segundo miembro del sistema de vector dual incluye la secuencia aceptora de empalme correspondiente a los nucleótidos 566-614 de la SEQ ID NO:46. En algunas modalidades, el segundo miembro del sistema de vector dual incluye la secuencia de cabeza AP correspondiente a los nucleótidos 257-543 de la SEQ ID NO:46. En algunas modalidades, el

5 segundo miembro del sistema de vector dual incluye la secuencia poli(A) correspondiente a los nucleótidos 3376-3597 de la SEQ ID NO:46. En modalidades particulares, el segundo miembro del sistema de vector dual incluye los nucleótidos 257-3597 de la SEQ ID NO:46 flanqueado en cada uno de los lados 5' y 3' por una repetición terminal invertida. En algunas modalidades, las repeticiones terminales invertidas flanqueantes son cualquier variante de las repeticiones

10 terminales invertidas de AAV2 que pueden encapsidarse mediante un plásmido que porta el gen Rep de AAV2. En determinadas modalidades, la repetición terminal invertida flanqueante 5' tiene una secuencia correspondiente a los nucleótidos 12-141 de la SEQ ID NO:46 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de

15 secuencia del 99 %) allí; y la repetición terminal invertida flanqueante 3' tiene una secuencia correspondiente a los nucleótidos 3685-3814 de la SEQ ID NO:46 o una secuencia que tiene al menos un 80 % de identidad de secuencia (al menos 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o identidad de secuencia del 99 %) allí. Los expertos en la técnica entenderán que, para cualquier par dado de

20 secuencias de repeticiones terminales invertidas en un plásmido de transferencia que se usa para crear el vector viral (típicamente al transfectar células con ese plásmido junto con otros plásmidos que llevan el AAV de genes necesario para la formación de vectores virales) (por ejemplo, la SEQ ID NO: 46), que la secuencia correspondiente en el vector viral puede alterarse debido a que las ITR adoptan una orientación "flip" o "flop" durante la recombinación. Por tanto, la secuencia de la

25 ITR en el plásmido de transferencia no es necesariamente la misma secuencia que se encuentra en el vector viral preparado a partir de este. Sin embargo, en algunas modalidades muy específicas, el primer miembro del sistema de vector dual incluye los nucleótidos 12-3814 de la SEQ ID NO:46.

Los plásmidos de transferencia que pueden utilizarse para producir vectores de ácido nucleico (por

30 ejemplo, vectores AAV) para la coformulación o coadministración (por ejemplo, la administración simultánea o secuencial) en un sistema de vector híbrido dual se proporcionan en la Tabla 5 (SEQ ID NO: 45 y SEQ ID NO: 46).

Tabla 5. Plásmidos de transferencia diseñados para producir vectores híbridos duales

SEQ ID NO.	Descripción	Secuencia de plásmido
5	Plásmido P960	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAA AGCCCCGGGCGTCGGGCGACCTTTGGTTCGCCCCGGCCTC
10	ITR 5' en las posiciones de los nucleótidos 1-130	AGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAAC TCCATCACTAGGGGTTCTTGTAGTTAATGATTAACCCG CCATGCTACTTATCTACGTAGCCATGCTCTAGGAAGATC GGAATTCGCCCTTAAGCTAGCGGCGCGCCACCGGTAGC
15	Promotor de OCM murino en las posiciones de los nucleótidos 225- 1364	AGGTTTGTACAGAAACCTTAGTTAAGGTTTGTGAGGG TTTTTTTTCTCTCTCTCTCTTAATTGGCTGTCCCAATC CATCCTTCTATAAATAGAAAAGAGAGACAGGGAGTGTGT GTGGTTTCATTACTAAGGTAAAGACACTTGAGCTACACA CACTTGATCCCTGAACATGAAATCTAAGAGGTTGAACGA TCACAGTTTCAGGACTATATAAGGTGGTGAAAGACCATC
20	Secuencia de codificación STRC de extremo N en las posiciones de los nucleótidos 1378- 4077	TGCTTCGTTTTTCTGTTTGTTCCTACAACCTTTCCCTCC GCTTGATTTTAACTCTAAATTGGTGAGTAGCTGGTGGGC TCACCAGACTCCGAGATCCTCTTCTCTGCACGCACTGTA TTAGACTTGGCACCCGGGAGGATTTTCACCTCTGCTGC ATGGGCTAATCTTCCACAAGGGATCTGTGGTATTGCAAT CTCGGGTTGATGCATGACGGTGATGTTGTGTTTATAGCA TGGCTAAGGTTTAGCTGCCTATGATGATTGGTTAGGGAA
25	Donante de empalme AP en las posiciones 4078-4161	GGATAATTTTGTCTAGAAGATTGGACTTTAGGGAAAAAA AACCCCACTTTTATTTGCTTTTAGAATTTTAAAAGACTGG GCCATGTAGCTCAGGCTGGTTTGGAGTTCATTATGTAGT CAAGGATGCTCTTGGACTCTTAGCATCCTCCTCCTCCT
30	Secuencia de cabeza AP en las posiciones de los nucleótidos 4168-4454	CTTTTTCCTCCTCCTCCTTCTTGTTCTTCTTCTTGTTCTTC CTCTTCCCCTTCTCTTCCCCCTTCTCTCCCTCTTCCTCCT CTTCCTCCTCCTTCTTGTTCTTCTTCTCCTCTTCCCCTTCTC CTCTCCCCCTTGTCTCCTCCTCTCCTTCTCCTCCTCCTCC TCTTCTTCTTTCTGAGTACCAAGATTGCAAGTGTGCACA
35	ITR 3' en las posiciones de los nucleótidos 4548-	CGATGACCAGCTTGGTCTTTCTTTGTCTTTTTTTTTTAAC TTCAATTTTGGAGTGAATTCAAGAGCAACCATGTAGTCA AGAGGTGGCTGGAGTCTTTTCTGTATCTGGGTTTGGTTT

SEQ ID NO.	Descripción	Secuencia de plásmido
5	4677	AGTACTCTGCCCCATCACTTAACAGGTCCTTATGGCCAC ATCTTAAAAAAATTCTAGAGATACACGGTGTGCGGTGAGT GGCTGAGAATGTGTGGTCTTCCCATTCTCTGTCAACGT GGCTCACATCTTGTTCCTCTGTTGCGCCAGGTAGAAAG GCGGCCGCCACCATGGCTCTGAGCCTCCAGCCCCAGC TGCTCCTTCTCCTGTGCTCCTGCCGCAGGAAGTGA 10 TCAGCCCCTACTGGGCCTCAGTCTTTGGATGCTGGTCT CTCCCTTCTGAAGTCATTCTGTAGCCACTCTGGACCAAGC TCCTCAGCGTTCCCTCAGCCAGTCACGGTTCTCTGCGTT CCTGGCCAACATTTCTTCATCCTTCCAGCTTGGGAGGAT GGGGGAGGGACCGGTGGGAGAGCCCCCACCTCTCCAG 15 CCCCCTGCACTTCGACTTCATGATTTCTCGTGACACTG AGAGGTAGCCCAGACTGGGAGCCAATGCTAGGGCTTCT GGGAGATGTGCTGGCACTCCTGGGACAGGAACAGACTC CCCGGGACTTTTTTGGTGCACCAGGCAGGTGTACTGGGT GGACTTGTAGAGGCATTGTTGGGAGCGTTAGTTCCTGG 20 AGGCCCCCCTGCCCCCACTCGACCCCCATGCACCCGT GATGGCCCTTCTGACTGTGTCCTGGCTGCTGATTGGTT GCCTTCTCTGATGTTGTTATTAGAGGGTACACGCTGGCA GGCCCTGGTGCAGTTGCAGCCCAGTGTGGACCCAACCA ATGCCACAGGTCTTGATGGTAGAGAGCCAGCTCCTCAC 25 TTTTACAGGGTCTGCTGGGCTTGCTTACCCCAGCAGG AGAGTTGGGCTCTGAGGAGGCTCTTTGGGGTGGTCTGC TGCGCACAGTGGGGGCCCCCCTCTATGCTGCCTTCCAG GAGGGGCTACTGCGAGTCACTCATTCTCTGCAAGATGA GGTCTTTTCTATTATGGGACAGCCAGAGCCTGATGCCA 30 GTGGGCAGTGCCAGGGAGGCAACCTTCAACAGCTGCTT TTATGGGGCATGCGGAACAACCTTTCTTGGGACGCCCG AGCACTGGGTTTTCTATCTGGATCACCACCTCCACCCCC TGCTCTCCTGCACTGCCTGAGCAGAGGTGTGCCTCTGC 35 CCAGGGCTTCCCAGCCTGCGGCTCACATCAGCCCTCGA

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CAGCGGCGAGCCATCTCTGTGGAGGCCCTCTGCGAGA ACCACTCAGGCCCAGAGCCACCCTACAGCATCTCCAAC TTCTCCATCTACTTGCTCTGCCAGCACATCAAGCCTGCC ACCCCGCGGGCCCCCTCCTACCACCCACGGCCTCCTCC TACCACCCACAGCCCCCTCCTACCACTACACAGCCCA TTCCTGACACTACACAGCCCCCTCCTGTCACCCCAAGG
10		CCTCCTCCTACCACCCACAACCCCTCCTAGCACAGC TGTCATCTGCCAGACAGCTGTATGGTACGCAGTCTCGT GGGCACCAGGTGCCCCGAGGTTGGCTCCAAGCCTGCCA TGATCAGTTTCCTGATCAATTTCTGGATATGATCTGCGG CAACCTCTCATTTTCAGCCCTGTCTGGCCCCAGTCGTCC
15		TTTGGTAAAGCAGCTCTGTGCTGGCTTGCTCCACCCC CCACTAGCTGTCCACCAGGCCTGATCCCTGTGCCCTC ACCCAGAAATATTCTGGGGCTGTTTCCTGGAGAATGA GACACTGTGGGCTGAACGGTTGTGTGTGGAGGACAGTC
20		TGCAGGCTGTGCCCCGAGGAACCAGGCTTGGGTTCA GCATGTGTGTGCGGGCCCCACCTTGGACGCCACTGATT TTCCACCGTGCCGCGTTGGACCCTGTGGGGAACGCTGC CCAGATGGGGGCAGCTTCCTGCTCATGGTCTGTGCCAA
25		TGACACTCTGTATGAAGCCTTGGTTCCCTTCTGGGCTTG GCTAGCAGGCCAATGCAGAATTAGTCGTGGAGGAAATG ATACTTGCTTTCTAGAAGGCATGCTGGGCCCCTTGTTGC CCTCTCTGCCCCCTCTGGGACCATCCCCACTCTGTCTG
30		GCTCCTGGTCCTTTTCTGCTTGGCATGTTATCCCAGTTG CCACGCTGTCAGTCCTCCGTGCCAGCCCTCGCCCACCC CACGCGCCTACATTACCTCCTGCGCCTACTGACCTTCCT TCTGGGTCCAGGGACTGGGGGTGCCGAGACGCAGGGG
35		ATGTTAGGTCAAGCCCTGCTGCTCTCTAGTCTCCCAGAC AACTGTTCAATTCTGGGATGCCTTCCGCCCAGAGGGCCG GAGAAGTGTACTGAGGACAGTCGGAGAGTACTTGCAGC GGGAAGAGCCAACCCACCAGGCTTAGACTCCTCCCTC

SEQ ID NO.	Descripción	Secuencia de plásmido
5		AGCCTCGGCTCTGGTATGAGCAAGATGGAGCTTCTGTC CTGCTTCAGTCCTGTACTGTGGGATCTACTCCAGAGAGA GAAGAGCGTTTGGGCCCTGAGGACCCTGGTGAAGGCC TACCTGCGCATGCCTCCAGAAGACCTTCAGCAGCTTGT GCTTTCAGCAGAGATGGAGGCTGCACAGGGCTTCCTGA CGCTCATGCTTCGTTCCCTGGGCTAAGCTGAAGGTTCAA
10		CCATCCGAGGAGCAGGCCATGGGCCGCTGACAGCCT TGCTGCTCCAGCGGTACCCACGCCTCACCTCCCAACTC TTTATCGACATGTCACCGCTCATCCCCTTCCTGGCTGTC CCTGACCTCATGCGCTTCCCACCGTCCCTTTTGGCCAA CGACAGTGTCTGGCTGCCATCAGGGATCACAGCTCAG
15		GAATGAAGCCTGAACAGAAGGAGGCCCTGGCAAAACGA CTGCTGGCCCCTGAGCTGTTTGGAGAAGTGCCTGATTG GCCCCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAG GAGACCAATAGAACTGGGCTTGTCGAGACAGAGAAGA
20		CTCTTGCGTTTCTGAGCTAGCCCCCGGGTGCGCGGCGT CGGTGGTGCCGGCGGGGGGCGCCAGGTGCGAGGCGG TGTAGGGCTCCAGGCAGGCGGCGAAGGCCATGACGTG CGCTATGAAGGTCTGCTCCTGCACGCCGTGAACCAGGT
25		GCGCCTGCGGGCCGCGCGCGAACACCGCCACGTCCTC GCCTGCGTGGGTCTCTTCGTCCAGGGGCACTGCTGACT GCTGCCGATACTCGGGGCTCCCGCTCTCGCTCTCGGTA ACATCCGGCCGGGCGCCGTCCTTGAGCACATAGCCTG
30		GACCGTTTCGTGACCTCGAGTTAAGGGCGAATTCCCG ATAAGGATCTTCCTAGAGCATGGCTACGTAGATAAGTAG CATGGCGGGTTAATCATTAACACAAGGAACCCCTAGTG ATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCT
35		CACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCG GGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGC GCAGCCTTAATTAACCTAATTCAGTGGCCGTCGTTTTAC AACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTT

SEQ ID NO.	Descripción	Secuencia de plásmido
5		AATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCG TAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAAC AGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGT AGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGC GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC CGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTT 10 CGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCC CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACC CCAAAAAACTTGATTAGGGTGATGGTTACGTAGTGGG CCATCGCCCTGATAGACGGTTTTTCGCCCTTGACGTTG 15 GAGTCCACGTTCTTTAATAGTGGA CTCTTGTTCCAACT GGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATT TATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAA ATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAA AATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGA AATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATAC 20 ATTCAAATATGTATCCGCTCATGAGACAATAACCCTGAT AAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGCC ATATTCAACGGGAAACGTGAGGCCGCGATTAAATTCCA ACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCG ATAATGTGCGGCAATCAGGTGCGACAATCTATCGCTTGT 25 ATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACAT GGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGAT GGTCAGACTAACTGGCTGACGGAATTTATGCCTCTTCC GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATG 30 GTTACTCACCCTGCGATCCCCGGAAAAACAGCATTCCA GGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTT GATGCGCTGGCAGTGTTCTGCGCCGGTTGCATTGAT TCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTT CGTCTTGCTCAGGCGCAATCACGAATGAATAACGGTTTG 35 GTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTG

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTT GCCATTCTCACC GGATT CAGTCGTC ACTCATGGTGATTT CTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATA GGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCG ATACCAGGATCTTGCCATCCTATGGA ACTGCCTCGGTGA GTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATAT 10 GGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATT TGATGCTCGATGAGTTTTTCTAACTGTCAGACCAAGTTT ACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAA TTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCA 15 TGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA AAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATC AAGAGCTACCAACTCTTTTTCCGAAGGTA ACTGGCTTCA 20 GCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGC CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCT GCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCG 25 GCTGAACGGGGGGTTCTGTGCACACAGCCCAGCTTGGA GCGAACGACCTACACCGAACTGAGATACCTACAGCGTG AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG GAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG 30 GTATCTTTATAGTCCTGTCTGGGTTTCGCCACCTCTGACT TGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGGCGGA GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGG TTCCTGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTC CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCG 35 CCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACG

SEQ ID NO.	Descripción	Secuencia de plásmido
5		ACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG
10		AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGG
		CCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGA
		CTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGA
		GTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTA
		TGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGA
		TAACAATTTACACAGGAAACAGCTATGACCATGATTAC
		GCCAGATTTAATTAAGGCCTTAATTAGG

SEQ ID NO.	Descripción	Secuencia de plásmido
5	46 Plásmido P726	CCTTAATTAGGCTGCGCGCTCGCTCGCTCACTGAGGCC GCCCCGGGCAAAGCCCCGGGCGTCCGGGCGACCTTTGGTC
10	ITR 5' en las posiciones de los nucleótidos 12-141	GCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG AGTGGCCAACCTCCATCACTAGGGGTTCTTGTAGTTAAT GATTAACCCGCCATGCTACTTATCTACGTAGCCATGCTC TAGGAAGATCGGAATTCGCCCTTAAGCTAGCGGCGCGC
15	Secuencia de cabeza AP en las posiciones de los nucleótidos 257-543	CCAATTGGCTTCGAATTCTAGCGGCCGCCCCGGGTGC GCGGCGTCGGTGGTGCCGGCGGGGGGCGCCAGGTCCG CAGGCGGTGTAGGGCTCCAGGCAGGCGGCGAAGGCCA TGACGTGCGCTATGAAGGTCTGCTCCTGCACGCCGTGA ACCAGGTGCGCCTGCGGGCCGCGCGCGAACACCGCCA
20	Aceptor de empalme AP en las posiciones 566-614	CGTCCTCGCCTGCGTGGGTCTCTTCGTCCAGGGGCACT GCTGACTGCTGCCGATACTCGGGGCTCCCGCTCTCGCT CTCGGTAACATCCGGCCGGGCGCCGTCTTGAGCACAT AGCCTGGACCGTTTTCTTAAGCGACGCATGCTCGCGAT
25	Secuencia de codificación STRC de extremo C en las posiciones de los nucleótidos 615-3344	AGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTC TCTCCACAGGAGCTGCTGTGGGCAGCCCTGCCTCTGCT TCCCCATCTGCCTCTGGAGAGCTTTCTCCAGCTCAGCC CTCACCAGATCCAGGCCCTGGAGGATAGCTGGCCAGTA GCAGATCTTGGGCCGGGACACGCCCCGACATGTGCTTCG TAGCCTAGTAAACCAGAGCATGGAGGATGGGGAGGAGC
30	secuencia de poliA de bGH en las posiciones de los nucleótidos 3376- 3597	AGGTGCTCAGGCTTGGGTCCCTCGCCTGTTTCCTGAGT CCTGAGGAGCTACAGAGTCTGGTGCCCTTGAGTGATCC AATGGGGCCTGTAGAACAGGGTCTGCTGGAATGTGCGG CCAATGGGACCCTCAGCCCAGAAGGACGGGTGGCATAT GAACTTCTGGGAGTGTTGCGTTCATCTGGAGGAACTGT CTTAAGCCCCCGAGAGCTGAGGGTCTGGGCACCTCTCT
35	ITR 3' en las posiciones de los nucleótidos 3685- 3814	TTCCCCAGCTGGGCCTCCGCTTCCTGCAGGAGCTCTCA GAGACCCAGCTTAGAGCCATGCTTCCTGCCCTACAGGG AGCCAGTGTCACACCTGCCCAGGCTGTTCTGTTGTTTG GAAGGCTCCTTCCTAAGCATGATCTGTCCCTGGAGGAA

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CTCTGCTCCCTGCACCCTCTCCTGCCAGGTCTCAGCCC CCAGACACTCCAGGCCATCCCTAAGAGAGTTCTGGTTG GTGCTTGTTCTGCTGGGCCCTGAACTGTCAAGGCTT TCAGCTTGCCAGATTGCAGCTCTGCTGCAGACCTTTTCG GGTAAAAGATGGTGTAAAAATATGGGTGCAGCAGGTG 10 CCGGCTCAGCCGTGTGCATTCTGGGCAGCCCACCACT TGCCAGACTGCCTGCTTCCCCTGCTCCCATTAAGCT GCTACAGCTGGACGCTGCAGCTCTTCTGGCAAACCGAA GACTCTATCGGCAGCTGCCTTGGTCTGAGCAACAGGCA CAGTTTCTCTGGAAGAAAATGCAAGTGCCTACCAACCTG 15 AGCCTGAGGAATCTGCAGGCTCTGGGCAACTTGGCAGG AGGCATGACCTGCGAGTTTCTGCAGCAGATCAGCTCAA TGGTTGACTTTCTTGATGTGGTACACATGCTCTACCAGC TGCCCACTGGTGTTTCGAGAGAGCCTGCGGGCCTGTATC TGGACAGAGCTACAGCGGAGGATGACAATGCCAGAGCC AGAGCTGACCACCCTAGGGCCAGAACTGAGTGAACCTTG 20 ACACAAAGCTACTCCTGGACTTGCCGATCCAGCTGATG GACAGATTGTCCAATGATTCCATTATGTTGGTGGTGGAG ATGGTCCAAGGCGCTCCAGAGCAGCTGCTGGCACTGAC CCCACTCCACCAGACAGCCTTGGCAGAGCGAGCACTTA AAAACCTGGCTCCAAAGGAGACCCCAATCTCCAAAGAA 25 GTGCTGGAGACACTGGGCCCCTTGGTTGGATTCTGCG AATAGAGAGCACGCGACGGATCCCTTTACCCATTCTACT GTCTCATCTCAGTCAGCTGCAGGGCTTCTGCCTAGGAG AGACATTTGCCACAGAGCTGGGATGGCTGCTGTTGCAG GAGCCTGTTCTTGGAACCAGAAATTGTGGAGCCAGGA 30 TGAATAGAGCAAGCTGGACGCCTAGTATTCACTCTGTC TGCTGAGGCTATTTCTCGATCCCCAGGGAGGCTTTGG GCCCAGAGACACTGGAGAGGCTTCTGGGAAAGCATCAA AGCTGGGAGCAGAGCAGAGTGGGCCATCTGTGTGGGG AGTCACAGCTTGCCCAAGAAAGCAGCTCTGGTAGCT 35

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GGGATTGTGCATCCAGCTGCTGAGGGTCTCCAAGAGCC TGTACCAAACCTGTGCAGACATACGGGGAACCTTCCCAG CGGCCTGGTCTGCGACACAAATCTCAGAGATGGAATC TCAGACTTTGAAGACTGCCTGTCACTATTTGCTGGAGAT CCAGGACTTGGTCCTGAGGAACTACGGGCAGCCATGG GCAAGGCCAAGCAGTTGTGGGGTCCCCCTCGAGGATTC
10		CGTCCTGAGCAGATCTTGCAGCTGGGCCGTCTCCTGAT AGGTCTAGGAGAACGGGAACTGCAGGAGCTTACCTTGG TGGACTGGGGTGTGCTGAGCAGCCTGGGGCAAATAGAT GGCTGGAGTTCCATGCAGCTCCGAGCCGTGGTCTCCAG TTTCCTAAGGCAGAGTGGTCGGCATGTGAGCCACCTGG
15		ACTTCATTTATCTGACAGCACTGGGTTACACAGTCTGTG GATTGCGACCAGAGGAGTTACAGCACATCAGCAGTTGG GAGTTTAGCCAAGCAGCTCTCTTCCTGGGTAGCTTGCAT CTCCCGTGCTCTGAGGAACAGCTGGAAGTTCTGGCCTA
20		TCTCCTTGTGTTGCCTGGTGGCTTTGGCCCAGTCAGTAA CTGGGGGCCTGAGATCTTCACTGAAATTGGCACAATAG CAGCTGGCATCCCAGACCTGGCTCTTTCAGCATTACTG CGGGGACAGATCCAAGGCCTGACTCCTCTTGCCATTTCT
25		TGTCATTCTGCTCCCAAGTTTGCAGTGGTCTTCAACCC CATCCAGTTATCTAGTCTCACCAGGGGTCAGGCCGTAG CTGTTACTCCTGAACAGCTGGCCTATCTGAGTCCTGAGC AGCGGCGAGCAGTTGCATGGGCCCAACACGAAGGGAA
30		GGAGATCCCAGAGCAGCTGGGTCGAAACTCAGCCTGG GGTCTCTACGACTGGTTCCAAGCCTCCTGGGCCCTGGC ATTGCCCCTCAGCATTTTTGGCCACCTATTATGAGCGGC CGCGGTACCAAGGGCGGATCCTGCATAGAGCTCGCTGA
35		TCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTT GTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGG TGCCACTCCCAGTGCCTTTCCTAATAAAATGAGGAAAT TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGG

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTG GGAAGACAATAGCAGGCATCTCGAGTTAAGGGCGAATT CCCGATAAGGATCTTCCTAGAGCATGGCTACGTAGATAA GTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCT AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCT CGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACG 10 CCCGGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGA GCGCGCAGCCTTAATTAACCTAATTCACTGGCCGTCGTT TTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCA ACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTG GCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCC 15 AACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCC CTGTAGCGGCGCATTAAAGCGCGGGCGGTGTGGTGGTT ACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC GCCCCGCTCCTTTCGCTTTCTTCCCTTCCTTCTCGCCAC GTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGC 20 TCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCG ACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTG GGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACG TTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT 25 GATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTA AAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTA ACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCG GGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA AATACATTCAAATATGTATCCGCTCATGAGACAATAACC 30 CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAT GAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAA ATTCCAACATGGATGCTGATTTATATGGGTATAAATGGG CTCGCGATAATGTGCGGCAATCAGGTGCGACAATCTAT CGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCT

SEQ ID NO.	Descripción	Secuencia de plásmido
5		GAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAG ATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGC CTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATG ATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACA GCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAA AATATTGTTGATGCGCTGGCAGTGTTCTGCGCCGGTT
10		GCATTGATTCTGTTTGTAAATTGTCCTTTTAACAGCGAT CGCGTATTTGCTCTTGCTCAGGCGCAATCACGAATGAAT AACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGT AATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCAT AACTTTTGCCATTCTCACCGGATTCAGTCGTCATCAT
15		GGTGATTTCTCACTTGATAACCTTATTTTGGACGAGGGG AAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATC GCAGACCGATACCAGGATCTTGCCATCCTATGGAACCTG CCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTT TCAAAAATATGGTATTGATAATCCTGATATGAATAAATTG
20		CAGTTTCATTTGATGCTCGATGAGTTTTTCTAACTGTCAG ACCAAGTTTACTCATATATACTTTAGATTGATTTAAACTT CATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTG ATAATCTCATGACCAAATCCCTTAACGTGAGTTTTCGTT
25		CCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAC
30		TGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCT AGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGT AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC AGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCG
35		GGTTGGAATCAAGACGATAGTTACCGGATAAGGCGCAG CGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCA GCTTGGAGCGAACGACCTACACCGAACTGAGATACCTA

SEQ ID NO.	Descripción	Secuencia de plásmido
5		CAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG GAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTC GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA ACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGG GGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT 10 TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATG TTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGT ATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAG CCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAA GCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCG 15 CGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGG TTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAAT TAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTT ACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGT GAGCGGATAACAATTTACACAGGAAACAGCTATGACCA 20 TGATTACGCCAGATTTAATTAAGG
25		
30		

Otros pares de ejemplos de superposición, transempalme y vectores híbridos duales se describen en la Tabla 6 a continuación.

Tabla 6. Pares representativos de superposición, transempalme y vectores híbridos para usar en los métodos y composiciones descritos en la presente

Número de par de vectores	Tipo de vector	Par de vectores
1	Superposición	El primer vector de ácido nucleico contiene: un promotor OCM (por ejemplo, SEQ ID NO: 1) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina humana (una parte de extremo N de la SEQ ID NO: 4), en el que el polinucleótido que codifica la parte de extremo N de la proteína estereocilina humana incluye los 500 pb 3' de la posición seleccionada como punto central de la región superpuesta de STRC
		El segundo vector de ácido nucleico contiene: un polinucleótido que codifica una parte de extremo C de la proteína estereocilina humana y una secuencia poli(A) en la que el polinucleótido que codifica la parte de extremo C de la proteína estereocilina humana incluye los 500 pb 5' de la posición seleccionada como punto central de la región superpuesta de STRC
2	Transempalme	El primer vector de ácido nucleico contiene: un promotor OCM (por ejemplo, SEQ ID NO: 1) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina humana (una parte de extremo N de la SEQ ID NO: 4) y una secuencia donante de empalme 3' del polinucleótido
		El segundo vector de ácido nucleico contiene: una secuencia aceptora de empalme 5' de un polinucleótido que codifica una parte de extremo C de la proteína estereocilina humana y una secuencia poli(A)
3	Híbrida	El primer vector de ácido nucleico contiene: un promotor OCM (por ejemplo, SEQ ID NO: 1) unido de forma operativa a un polinucleótido que codifica una parte de extremo N de una proteína estereocilina humana (por

Número de par de vectores	Tipo de vector	Par de vectores
5		ejemplo, una parte de extremo N de la SEQ ID NO: 4), una secuencia donante de empalme 3' del polinucleótido y una región recombinogénica 3' de la secuencia donante de empalme
10		El segundo vector de ácido nucleico contiene: una región recombinogénica, una secuencia aceptora de empalme 3' de la región recombinogénica, un polinucleótido que codifica una parte de extremo C de la proteína estereocilina humana 3' de la secuencia aceptora de empalme y una secuencia poli(A)
15		

Sistemas de expresión de inteínas

Otro enfoque de terapia génica para expresar proteínas grandes en células de mamífero implica el uso de inteínas. Una inteína, también conocida como "intrón de proteína", es una parte de una proteína que típicamente tiene 100-900 residuos de aminoácidos y es capaz de autoextirparse y ligar los residuos de extremo N y C de los fragmentos proteicos que la flanquean ("exteínas"). Las inteínas pueden dividirse en tres clases diferentes, que incluyen la maxi-inteína, la mini-inteína y la inteína dividida. Las maxi-inteínas se refieren a las regiones de empalme de extremo N y C de una proteína interrumpida por un dominio de endonucleasa homing (HEG). Los "HEG" se refieren a una clase de endonucleasas codificadas como genes independientes dentro de intrones, como fusiones de proteínas con otras proteínas o como inteínas autoempalmables. Los HEG generalmente hidrolizan muy pocas y selectas regiones de ADN. Una vez que una HEG hidroliza un fragmento de ADN, el gen que codifica la HEG suele incorporarse al sitio de escisión, aumentando así su frecuencia alélica. Las mini-inteínas se refieren a los dominios de empalme de extremo N y C que carecen el dominio de HEG. Las "inteínas divididas" se refieren a las inteínas que se transcriben y traducen como dos polipéptidos separados que se unen con una exteína. Las inteínas de alanina son otra clase de inteínas que tienen una unión de empalme de una alanina en lugar de una cisteína o serina.

El dominio de empalme de las inteínas contiene dos subdominios, a saber, los dominios de empalme de extremo N y C, que contienen motivos conservados con residuos conservados que

median la actividad de empalme. El dominio de empalme de extremo N contiene motivos estructurales A, N2, B y N4, mientras que el dominio de empalme de extremo C contiene los motivos F y G. El motivo A contiene Cys/Ser o Thr como residuos conservados; el motivo B incluye residuos His y Thr; el motivo F contiene residuos Asp y His; los motivos G llevan dos residuos conservados, que incluyen un His penúltimo y un extremo Asn. Los motivos C, D, E y H están generalmente relacionados con el dominio HEG de las maxi-inteínas.

El empalme de las inteínas se encuadra en tres estrategias distintas: 1) empalme de inteínas de clase 1 (o clásico/canónico) que implica (a) un desplazamiento de acil (N-S/N-O) que transforma el enlace de péptidos de una unión de empalme de extremo N en un enlace tio(éster), (b) reacción de transesterificación que forma un intermediario ramificado, (c) la ciclización del Asn, que elimina el intermediario ramificado al escindir la unión de empalme de extremo C, y (d) un segundo desplazamiento de acil (S-N/O-N) que liga los segmentos de exteína flanqueantes mediante la formación de enlaces amida; 2) las inteínas de clase 2 (también conocidas como alanina-inteínas) omiten la etapa (a) de la reacción de empalme clásica; y 3) el mecanismo de clase 3, que implica la formación de dos intermedios ramificados.

Entre los diversos sistemas de inteína descritos anteriormente, se ha demostrado que el enfoque de transempalme de inteína dividida supera con éxito las limitaciones de tamaño de los vectores de terapia génica tradicionales (por ejemplo, AAV: límite de tamaño máximo de ~5 kb). Por ejemplo, Subramanyam et ál. (PNAS 110:15461-6 (2013)) han empleado el sistema de inteína dividida para reconstituir la subunidad $\alpha 1C$ del canal de calcio de tipo L en cardiomiocitos a partir de dos mitades separadas. De manera similar, Truong et ál. (Nucleic Acids Res. 43:6450-8 (2015)) han demostrado la reconstitución exitosa de dos mitades de la proteína Cas9 utilizando un sistema de inteína dividida. Por consiguiente, la presente divulgación proporciona sistemas de transempalme de inteína dividida para el empaquetamiento y la entrega de una secuencia codificante de estereocilina que está unida de forma operativa a un promotor OCM. Este método permite que dos polinucleótidos separados, cada uno de los cuales contiene aproximadamente la mitad del gen STRC e incluye una secuencia de polinucleótidos que codifica un fragmento de N-inteína o un fragmento de C-inteína, se expresen a partir de dos vectores de expresión separados (por ejemplo, cualquiera de los vectores de ácido nucleico divulgados en la presente) y se reconstituyan postraduccionalmente para producir una proteína estereocilina de longitud completa. Tales sistemas pueden incorporarse en vectores de expresión de ácido nucleico divulgados en la presente, tales como, por ejemplo, vectores rAAV.

En un ejemplo, la presente divulgación proporciona un sistema de inteína dividido en dos vectores que contiene: a) un primer vector de ácido nucleico que contiene un polinucleótido que incluye una secuencia que codifica una parte de extremo N de una proteína estereocilina humana (por

ejemplo, una parte de extremo N de la SEQ ID NO: 4), en la que la secuencia que codifica una parte de extremo N de una proteína estereocilina incluye en su extremo 3' una secuencia de polinucleótidos que codifica una N-inteína; b) un segundo vector que contiene un polinucleótido que incluye una secuencia que codifica una parte de extremo C de una proteína estereocilina humana (por ejemplo, una parte de extremo C de la SEQ ID NO: 4), en la que la secuencia que codifica una parte de extremo C de una proteína estereocilina incluye en su extremo 5' una secuencia de polinucleótidos que codifica una C-inteína.

En otro ejemplo, la presente divulgación proporciona un sistema de inteína dividido en dos vectores que contiene: a) un primer vector que contiene un polinucleótido que incluye una secuencia que codifica una parte de extremo N de una proteína estereocilina murina (por ejemplo, una parte de extremo N de la SEQ ID NO: 5), en la que la secuencia que codifica una parte de extremo N de una proteína estereocilina incluye en su extremo 3' una secuencia de polinucleótidos que codifica una N-inteína; b) un segundo vector que contiene un polinucleótido que incluye una secuencia que codifica una parte de extremo C de una proteína estereocilina murina (por ejemplo, una parte de extremo C de la SEQ ID NO: 5), en la que la secuencia que codifica una parte de extremo C de una proteína estereocilina incluye en su extremo 5' una secuencia de ácido nucleico que codifica una C-inteína.

En algunas modalidades, tanto el primer vector como el segundo vector incluyen además una secuencia promotora, tal como una secuencia promotora de OCM (por ejemplo, una secuencia promotora OCM de cualquiera de las SEQ ID NO: 1-3) unida de forma operativa al extremo 5' de un polinucleótido que codifica la primera proteína de fusión (una parte de extremo N de una proteína estereocilina fusionada a una N-inteína) y/o al extremo 5' del polinucleótido que codifica la segunda proteína de fusión (una parte de extremo C de una proteína estereocilina fusionada a una C-inteína). En algunas modalidades, el promotor OCM tiene la secuencia de la SEQ ID NO: 1 o una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más) de identidad de secuencia con la SEQ ID NO: 1. En algunas modalidades, el promotor OCM tiene la secuencia de la SEQ ID NO: 2 o una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 2. En algunas modalidades, el promotor OCM tiene la secuencia de la SEQ ID NO: 3 o una variante de esta que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 3.

En algunas modalidades, la N-inteína y la C-inteína derivan de la misma inteína o gen de inteína dividido. De manera alternativa, las secuencias de la N-inteína y la C-inteína derivan de dos genes

de inteína diferentes que pueden realizar el transempalme de proteína para reconstituir una proteína estereocilina de longitud completa. En algunas modalidades, el mismo gen procede del mismo organismo o de organismos diferentes. Las inteínas de división comúnmente utilizadas derivan del gen *DnaE* de diversos organismos. En algunas modalidades, el polinucleótido que codifica una proteína estereocilina se divide en dos partes, cada una de las cuales corresponde aproximadamente a la mitad de la secuencia de codificación total del gen de longitud completa, a saber, una parte de extremo N y una parte de extremo C. El polinucleótido que codifica la parte de extremo N de la estereocilina se fusiona en marco en su extremo 3' con el polinucleótido que codifica la N-inteína, mientras que el polinucleótido que codifica la parte de extremo C de la estereocilina se fusiona en marco en su extremo 5' con el polinucleótido que codifica la C-inteína. En algunas modalidades, el primer vector y el segundo vector, cuando se introducen en una célula (por ejemplo, una célula de un sujeto, como un sujeto con pérdida auditiva neurosensorial, por ejemplo, DFNB16) producen una primera proteína de fusión y una segunda proteína de fusión. En algunas modalidades, la primera proteína de fusión contiene la parte de extremo N de la proteína estereocilina fusionada en su extremo C con la N-inteína. En algunas modalidades, la segunda proteína de fusión contiene la parte de extremo C de la proteína estereocilina fusionada en su extremo N con la C-inteína. En algunas modalidades, la N-inteína de la primera proteína de fusión y la C-inteína de la segunda proteína de fusión se unen selectivamente para producir una tercera proteína de fusión que contiene del extremo N al extremo C: una parte de extremo N de la proteína estereocilina, una N-inteína unida en su extremo C a la C-inteína, y la parte de extremo C de la proteína estereocilina. En algunas modalidades, la N-inteína unida a la C-inteína es capaz de llevar a cabo una reacción de transempalme que extirpa la N-inteína y la C-inteína y liga el extremo C de la parte de extremo N y el extremo N de la parte de extremo C de la proteína estereocilina.

El sistema de inteína dividida descrito en la presente puede incluir inteínas divididas que están codificadas por un gen que posteriormente se manipula mediante métodos rutinarios para codificar dos fragmentos de inteína separados (por ejemplo, una inteína dividida). En algunas modalidades, las inteínas divididas están codificadas por dos genes separados.

Las inteínas divididas de las composiciones y métodos divulgados pueden derivar del gen *DnaE* (por ejemplo, ADN polimerasa III subunidad alfa) de cianobacterias, tales como, por ejemplo, *Nostoc punctiforme* (Npu), *Synechocystis* sp. PCC6803 (Ssp), *Fischerella* sp. PCC9605 (Fsp), *Scytonema tolypothrichoides* (Sto), bacteria Cyanobacteria SW_9_47_5, *Nodularia spumigena* (Nsp), *Nostoc flagelliforme* (Nfl), *Crocospaera watsonii* (Cwa) WH8502, *Chroococcidiopsis cubana* (Ccu) CCALA043, *Trichodesmium erythraeum* (Ter), *Rhodothermus marinus* (Rma), *Saccharomyces cerevisiae* (Sce), *Saccharomyces castellii* (Sca), *Saccharomyces unisporus* (Sun), *Zygosaccharomyces bisporus* (Zbi), *Torulaspora pretoriensis* (Tpr), *Mycobacteria tuberculosis*

(Mtu), *Mycobacterium leprae* (Mle), *Mycobacterium smegmatis* (Msm), *Pyrococcus abyssi* (Pab), *Pyrococcus horikoshii* (Pho), *Coxiella burnetti* (Cbu), *Coxiella neoformans* (Cne), *Coxiella gattii* (Cga), *Histoplasma capsulatum* (Hca) y cloroplasto *Porphyra purpurea* (Ppu), entre otras. En algunas modalidades, la inteína dividida deriva de estudios de alineación de secuencias múltiples de *DnaE* para identificar un diseño de consenso (por ejemplo, Cfa) para diseñar una inteína dividida con la estabilidad y actividad deseables (por ejemplo, las inteínas de empalme son inteínas Cfa). Otros sistemas de inteína dividida adecuados para usar con las composiciones y métodos descritos en la presente incluyen aquellos descritos en las publicaciones de solicitudes de patentes internacionales n.º WO 2017/132580, WO 2020/079034, WO 2018/071868, WO 2020/249723, WO 2021/099607, WO 2021/040703, WO 2013/045632, WO 2020/146627, y WO 2021/047558, y las patentes estadounidenses n.º US 10,066.027, US 10.526.401 y US 8.394.604, cada una de las cuales se incorpora a la presente mediante referencia en lo que se refiere a los sistemas de inteína dividida.

En algunas modalidades, el primer vector y el segundo vector incluyen además una repetición de terminal invertida de 5' (ITR) en su extremo 5' y una ITR 3' en su extremo 3'. En algunas modalidades, la ITR 5' y la ITR 3' son ITR de AAV. En algunas modalidades, las ITR de AAV son ITR de AAV2.

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye: a) un primer vector que contiene de 5' a 3': i) opcionalmente, un ITR 5' (por ejemplo, ITR 5' de AAV2); ii) un polinucleótido que contiene un promotor OCM (por ejemplo, un promotor OCM de cualquiera de las SEQ ID NO: 1-3); iii) un polinucleótido que codifica una parte de extremo N de una proteína estereocilina (por ejemplo, una parte de extremo N de la proteína estereocilina de la SEQ ID NO: 4 o la SEQ ID NO: 5); iv) un polinucleótido que codifica una N-inteína; (v) opcionalmente, una secuencia poli(A); y (vi) opcionalmente, una ITR 3' (por ejemplo, ITR 3' de AAV2); y b) un segundo vector que contenga de 5' a 3': i) opcionalmente, un 5' ITR (por ejemplo, ITR 5' de AAV2); ii) un polinucleótido que contenga un promotor OCM (por ejemplo, un promotor OCM de cualquiera de las SEQ ID NO: 1-3); iii) un polinucleótido que codifica una C-inteína; iv) un polinucleótido que codifica una parte de extremo C de la proteína estereocilina (por ejemplo, una parte de extremo C de la proteína STRC de la SEQ ID NO: 4 o la SEQ ID NO: 5); (v) opcionalmente, una secuencia poli(A); y (vi) opcionalmente, una ITR 3' (por ejemplo, ITR 3' de AAV2).

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 8 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con

la SEQ ID NO: 8, como se muestra debajo.

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRA
TKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 8)

5

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 9 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 9, como se muestra debajo.

10

VKIISRKSLGTQNVYDIGVEKDHNFLKNGLVASN (SEQ ID NO: 9)

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 10 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 10, como se muestra debajo.

15

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRA
TKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 10)

20

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 11 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 11, como se muestra debajo.

25

MVKIISRKSLGTQNVYDIGVEKDHNFLKNGLVASN (SEQ ID NO: 11)

30

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 12 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 12, como se muestra debajo.

35

VKIISRKSLGTQNVYDIGVGEPHNFLKNGLVASN (SEQ ID NO: 12)

En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector
 5 que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de
 aminoácidos de la SEQ ID NO: 8 o la SEQ ID NO: 10 (por ejemplo, posicionado 3' de un
 polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo
 vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia
 de aminoácidos de la SEQ ID NO: 9, SEQ ID NO: 11, o SEQ ID NO: 12 (por ejemplo, posicionado
 10 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina). En
 algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que
 incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de
 aminoácidos de la SEQ ID NO: 8 y un segundo vector que incluye un polinucleótido que codifica un
 polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 9. En algunas
 15 modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un
 polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la
 SEQ ID NO: 8 y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-
 inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 11. En algunas modalidades, el
 sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido
 20 que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 8
 y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene
 una secuencia de aminoácidos de la SEQ ID NO: 12. En algunas modalidades, el sistema de
 inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica
 un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 10 y un
 25 segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una
 secuencia de aminoácidos de la SEQ ID NO: 9. En algunas modalidades, el sistema de inteína
 dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un
 péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 10 y un segundo
 vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia
 30 de aminoácidos de la SEQ ID NO: 11. En algunas modalidades, el sistema de inteína dividido en
 dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-
 inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 10 y un segundo vector que
 incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de
 aminoácidos de la SEQ ID NO: 12.

35 En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye

un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 13 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 13, como se muestra debajo.

5

CLSYETEILTVEYGLLPKIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRA
TKDHFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 13)

10

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 14 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 14, como se muestra debajo.

15

MIKIATRKYLGKQNVYDIGVERDHNFKNGFIASN (SEQ ID NO: 14)

20

En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 13 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 14 (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina).

25

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 15 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 15, como se muestra debajo.

30

CLSYDTEILTVEYGFLPKIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRA
TKDHFMTTDDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 15)

35

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 16 o que tiene al menos un 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90

%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con la SEQ ID NO: 16, como se muestra debajo.

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 16)

5

En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 15 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 16 (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina).

10

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CFSGDTLVALTD (SEQ ID NO: 17). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CLAGDTLITLA (SEQ ID NO: 18). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CLQNGTRLLR (SEQ ID NO: 19). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CLTGDSQVLTR (SEQ ID NO: 20). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CLTYETEIMTV (SEQ ID NO: 21). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de CLSGNTKVRFRY (SEQ ID NO: 22). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) identidad de secuencia con cualquiera de las SEQ ID NO: 17-22.

15

20

25

30

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de C-inteína que tiene una secuencia de aminoácidos de GV FVHN (SEQ ID NO: 23). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de C-inteína que tiene

35

una secuencia de aminoácidos de GLLVHN (SEQ ID NO: 24). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de C-inteína que tiene una secuencia de aminoácidos de GLIASN (SEQ ID NO: 25). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido C-inteína que tiene una secuencia de aminoácidos de GLVVHN (SEQ ID NO: 26). En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye un polinucleótido que codifica un péptido de C-inteína que tiene una secuencia de aminoácidos que tiene al menos 85 % (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) identidad de secuencia con cualquiera de las SEQ ID NO: 23-26.

En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 17 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 23 (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina). En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 20 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 24 (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina). En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 21 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 25 (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina). En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 22 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido C-inteína que tiene una secuencia de aminoácidos de la SEQ ID NO: 26 (por ejemplo, posicionado 5' de un polinucleótido que codifica

una parte de extremo C de una proteína estereocilina).

En algunas modalidades, el sistema de inteína dividido en dos vectores de la divulgación incluye de forma colectiva uno o más polinucleótidos que codifican un par de N-inteína y C-inteína descrito en la Tabla 7 o uno o más polinucleótidos que codifican un par de N-inteína y C-inteína que tienen al menos un 85 % de identidad de secuencia (por ejemplo, al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más) de identidad de secuencia con un par de N-inteína y C-inteína descrito en la Tabla 7, como se muestra debajo. En algunas modalidades, el sistema de inteína dividido en dos vectores incluye un primer vector que incluye un polinucleótido que codifica un péptido de N-inteína que tiene una secuencia de aminoácidos enumerada en la Tabla 7 (por ejemplo, posicionado 3' de un polinucleótido que codifica una parte de extremo N de una proteína estereocilina) y un segundo vector que incluye un polinucleótido que codifica un polipéptido de C-inteína que tiene una secuencia de aminoácidos enumerada en la misma fila de la Tabla 7 que la secuencia de aminoácidos de N-inteína (por ejemplo, posicionado 5' de un polinucleótido que codifica una parte de extremo C de una proteína estereocilina).

Tabla 7. Pares de secuencia de inteína dividida representativos

Inteína	Secuencia de aminoácidos del N-inteína	SEQ ID NO	Secuencia de aminoácidos del C-inteína	SEQ ID NO
<i>Npu-DnaE</i>	CLSYETEILTVEYGLLPICKIVE KRIECTVYSVDNNGNIYTQPVA QWHDRGEQEVFEYCLEDGSLI RATKDHKFMTVDGQMLPIDEIF EREGLMRVDNLPN	27	IKIATRKYLGKQNVYDIGVE RDHNFALKNGFIASN	28
<i>Rma-DnaB</i>	CLAGDTLITLADGRRVPIRELV SQQNFSVWALNPQTYRLERA RVSRAFCTGIKPVYRLTTRLGR SIRATANHRFLTPQGWKRVDE LQPGDYALPRRIPTASTPTL	29	AAACPELRQLAQSDVYWD PIVSIEPDGVEEVFDLTVPG PHNFVANDIIAHN	30
<i>mNpu-DnaE</i>	CLSYDTEILTVEYGILPICKIVEK RIECTVYSVDNNGNIYTQPVAQ WHDRGEQEVFEYCLEDGSLIR ATKDHKFMTVDGQMMPIDEIF EREGLMRVDNLPN	31	VKVIGRRSLGVQRIFDIGLP QYHNFLLANGAIAAN	32
<i>Cfa</i>	CLSYDTEILTVEYGFLPICKIVE	33	VKIISRKSLGTQNVYDIGVE	34

	Inteína	Secuencia de aminoácidos del N-inteína	SEQ ID NO	Secuencia de aminoácidos del C-inteína	SEQ ID NO
5		ERIECTVYTVDKNGFVYTQPIA QWHNRGEQEVFEYCLEDGSII RATKDHKFMTTDGQMLPIDEIF ERGLDLKQVDGLP		KDHNFLLKNGLVASN	
10	<i>DnaE</i>	CLSYDTEILTVEYGFLPIGKIVE ERIECTVYTVDKNGFVYTQPIA QWHNRGEQEVFEYCLEDGSII RATKDHKFMTTDGQMLPIDEIF ERGLDLKQVDGLP	35	KRTADGSEFESPKKKRKV KIISRKSLGTQNVYDIGVEK DHNFLLKNGLVASN	36
15	<i>Ssp DnaE</i>	CLSFGTEILTVEYGFLPIGKIVS EEINCSVYSVDPEGRVYTQAIA QWHDRGEQEVLEYELEDGSVI RATSDHRFLTDDYQLLAIEEIFA RQLDLLTLENIKQTEEALDNHR LPFPLLDAGTIK	37	VKVIGRRSLGVQRIFDIGLP QDHNFLLANGAIAANC	38
20	<i>DnaE</i>	CLSYDTEILTVEYGFLPIGKIVE ERIECTVYTVDKNGFVYTQPIA QWHNRGEQEVFEYCLEDGSII RATKDHKFMTTDGQMLPIDEIF ERGLDLQVD	39	GLPVKIISRKSLGTQNVYDI GVEKDHNFLLKNGLVASN	40

25 La N-inteína Npu de la SEQ ID NO: 27 puede estar codificada por un polinucleótido que tiene la secuencia de ADN de la SEQ ID NO: 41, como se muestra a continuación.

30 TGCCTGAGCTACGAGACCGAGATCCTGACCGTGGAGTACGGCCTGCTGCCCATCGGCAAGAT
CGTGGAGAAGAGAATCGAGTGCACCGTGTACAGCGTGGACAACAACGGCAACATCTACACCC
AGCCCGTGGCCCAGTGGCACGACAGAGGCGAGCAGGAGGTGTTTCGAGTACTGCCTGGAGGA
CGGCAGCCTGATCAGAGCCACCAAGGACCACAAGTTCATGACCGTGGACGGCCAGATGCTGC
CCATCGACGAGATCTTCGAGAGAGAGCTGGACCTGATGAGAGTGGACAACCTGCCCAAC
(SEQ ID NO: 41)

35 La C-inteína Npu de la SEQ ID NO: 28 puede estar codificada por un polinucleótido que tiene la

secuencia de ADN de la SEQ ID NO: 42, como se muestra a continuación.

ATCAAGATCGCCACAAGAAAGTACCTGGGCAAGCAGAACGTGTACGACATCGGCGTGGAGAG
AGACCACAACCTTCGCCCTGAAGAACGGCTTCATCGCCAGCAAT (SEQ ID NO: 42)

5

Una inteína dividida de la divulgación (es decir, la N-inteína y la C-inteína) puede incluir un aminoácido nucleófilo en o cerca de su extremo N o C que es capaz de realizar la reacción de transempalme. En algunas modalidades, el aminoácido nucleófilo se selecciona de serina, treonina, cisteína o alanina.

10

En algunas modalidades, el primer vector y/o el segundo vector incluyen además una o más secuencias reguladoras adicionales, tales como, por ejemplo, un elemento regulador postranscripcional del virus de la hepatitis de Woodchuck (WPRE), una secuencia potenciadora, una secuencia poli(A), una secuencia terminadora o una señal de degradación, entre otras.

15

En algunas modalidades, el sistema de inteína dividida descrito en la presente incluye una inteína dependiente de ligando, que realiza el empalme de proteínas al entrar en contacto con un ligando (por ejemplo, moléculas pequeñas tales como 4-hidroxitamoxifeno, péptidos, proteínas, polinucleótidos, aminoácidos, nucleótidos, etc.). Varias inteínas dependientes de ligando se describen en US 2014/0065711, cuya divulgación se incorpora mediante referencia en la presente en lo que se refiere a inteínas dependientes de ligando.

20

La presente divulgación proporciona vectores que contienen una o más señales de degradación dentro del polipéptido o polipéptidos de inteína (por ejemplo, N-inteína o C-inteína) que median la degradación de proteínas por el sistema ubiquitina-proteasoma y/o las vías autofagia-lisosoma.

25

Tales secuencias pueden incorporarse a los sistemas de vectores de la divulgación para evitar o reducir la acumulación de proteínas inteína extirpadas dentro de las células objetivo. Entre las señales de degradación de ejemplo se incluyen los N-degrones y los C-degrones, que son secuencias de péptidos que contienen motivos con residuos de lisina capaces de ser poliubiquitilados y posteriormente dirigidos a la degradación. En algunas modalidades, los degrones son señales de degradación localizadas dentro de una secuencia de proteínas (por

30

ejemplo, una secuencia de inteína) que no está en el extremo N ni en el extremo C de la secuencia de proteínas. En algunas modalidades, la proteína N-inteína incluye uno o más (por ejemplo, 2, 3, 4, 5 o más) degrones. En algunas modalidades, la proteína C-inteína incluye uno o más (por ejemplo, 2, 3, 4, 5 o más) degrones. En algunas modalidades, el degron es un degron CL1, que es un péptido desestabilizador de extremo C que comparte similitud estructural con proteínas mal

35

plegadas y es reconocido por el sistema de ubiquitinación. En algunas modalidades, el degron es un degron PB29, SMN, CIITA u ODC. Tales señales de degradación se describen en WO

2016/13932, que se incorpora mediante referencia en la presente en lo que se refiere a señales de degradación. Otro ejemplo de señal de degradación incluye el degron derivado de la dihidrofolato reductasa de *E. coli* (ecDHFR), tal como se describe en WO 2020/079034 (incorporado por referencia en la presente). Las señales de degradación adicionales incluyen dominios de degradación FKBP12 (Banaszynski et ál., *Cell* 126:995-1004, 2006), dominios de degradación PEST (Rechsteiner y Rogers, *Trends Biochem Sci.* 21:267-271, 1996), señales de ubiquitinación de etiquetas UbR (Chassin et ál., *Nat Commun.* 10:2013, 2019), y mutaciones desestabilizadas de ELRBD humano (Miyazaki et ál., *J. Am. Chem. Soc.*, 134:3942-3945, 2012).

10 Vectores para la expresión de estereocilina

Además de lograr altas tasas de transcripción y traducción, se puede lograr la expresión estable de un gen exógeno en una célula de mamífero mediante la integración del polinucleótido que contiene el gen en el genoma nuclear de la célula de mamífero. Se ha desarrollado una variedad de vectores para el suministro e integración de polinucleótidos que codifican estereocilinas en el ADN nuclear de una célula de mamífero. Ejemplos de vectores de expresión se describen en, por ejemplo, Gellissen, *Production of Recombinant Proteins: Novel Microbial and Eukaryotic Expression Systems* (John Wiley & Hijos, Marblehead, MA, 2006). Los vectores de expresión para usar en las composiciones y métodos descritos en la presente contienen un promotor de OCM (por ejemplo, un polinucleótido que tiene al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, de identidad de secuencia) con respecto a cualquiera de las secuencias promotoras enumeradas en la Tabla 3 (por ejemplo, cualquiera de las SEQ ID NO: 1-3)) unida de forma operativa a una secuencia de polinucleótidos que codifica una parte de una proteína estereocilina (por ejemplo, una parte de la SEQ ID NO: 4 o la SEQ ID NO: 5), así como, por ejemplo, elementos de secuencia adicionales utilizados para la expresión de estos agentes y/o la integración de estas secuencias de polinucleótidos en el genoma de una célula de mamífero. Los vectores que pueden contener un promotor específico de OCM unido de forma operativa a un transgén que codifica una parte de una proteína de estereocilina incluye plásmidos (por ejemplo, moléculas de ADN circular que pueden replicarse de forma autónoma dentro de una célula), cósmidos (por ejemplo, vectores pWE o sCos), cromosomas artificiales (por ejemplo, un cromosoma artificial humano (HAC), un cromosoma artificial de levadura (YAC), un cromosoma artificial bacteriano (BAC) o un cromosoma artificial derivado de P1 (PAC)) y vectores virales. Determinados vectores que pueden usarse para la expresión de estereocilina incluyen plásmidos que contienen secuencias reguladoras, tales como regiones potenciadoras, que dirigen la transcripción génica. Otros vectores útiles para la expresión de una proteína de estereocilina contienen secuencias de polinucleótidos que mejoran la

velocidad de traducción de este gen o mejoran la estabilidad o exportación nuclear del ARNm que resulta de la transcripción del gen. Estos elementos de secuencia incluyen, por ejemplo, regiones no traducidas 5' y 3', un sitio de entrada ribosómico interno (IRES) y un sitio de señal de poliadenilación para dirigir la transcripción eficaz del gen transportado en el vector de expresión.

Los vectores de expresión adecuados para usar con las composiciones y métodos descritos en la presente también pueden contener un polinucleótido que codifica un marcador para la selección de células que contienen dicho vector. Los ejemplos de un marcador adecuado incluyen genes que codifican resistencia a antibióticos, tales como ampicilina, cloranfenicol, kanamicina o nourseotricina.

Vectores virales para el suministro de polinucleótidos

Los genomas virales proporcionan una rica fuente de vectores que pueden utilizarse para el suministro eficiente de STRC en el genoma de una célula objetivo (por ejemplo, una célula de mamífero, tal como una célula humana). Los genomas virales son vectores particularmente útiles para la administración de genes porque los polinucleótidos contenidos en dichos genomas se incorporan típicamente en el genoma nuclear de una célula de mamífero mediante transducción generalizada o especializada. Estos procesos se producen como parte del ciclo de replicación viral natural y no requieren proteínas o reactivos agregados para inducir la integración de genes. Los ejemplos de vectores virales incluyen un retrovirus (por ejemplo, Vector viral de la familia Retroviridae), adenovirus (por ejemplo, Ad5, Ad26, Ad34, Ad35 y Ad48), parvovirus (por ejemplo, Virus adenoasociados), coronavirus, virus de ARN de hebra negativa como ortomixovirus (por ejemplo, virus de la influenza), rabdovirus (por ejemplo, virus de la rabia y estomatitis vesicular), paramixovirus (por ejemplo, sarampión y Sendai), virus de ARN de cadena positiva, como picornavirus y alfavirus, y virus de ADN de doble cadena, incluidos adenovirus, herpesvirus (por ejemplo, Virus del herpes simple tipos 1 y 2, virus de Epstein-Barr, citomegalovirus) y poxvirus (por ejemplo, Vaccinia, vaccinia modificada de Ankara (MVA), viruela aviar y viruela del canario). Otros virus incluyen el virus Norwalk, togavirus, flavivirus, reovirus, papovavirus, hepadnavirus, virus del papiloma humano, virus de la espuma humana y virus de la hepatitis, por ejemplo. Los ejemplos de retrovirus incluyen: leucosis-sarcoma aviar, virus aviares de tipo C, virus de mamíferos de tipo C, virus de tipo B, virus de tipo D, oncoretrovirus, grupo HTLV-BLV, lentivirus, alfaretrovirus, gammaretrovirus, spumavirus (Coffin, J. M., Retroviridae: The viruses and their replication, Virology, Tercera Edición (Lippincott-Raven, Filadelfia, 1996)). Otros ejemplos incluyen virus de la leucemia murina, virus del sarcoma murino, virus del tumor mamario de ratón, virus de la leucemia bovina, virus de la leucemia felina, virus del sarcoma felino, virus de la leucemia aviar, virus de la leucemia de células T humanas, virus endógeno del babuino, virus de la leucemia del mono

Gibbon, Mason Pfizer virus del mono, virus de la inmunodeficiencia de los simios, virus del sarcoma de los simios, virus del sarcoma de Rous y lentivirus. Se describen otros ejemplos de vectores, por ejemplo, la patente de EE.UU. nº 5.801.030, cuya descripción se incorpora aquí como referencia en lo que se refiere a vectores virales para uso en terapia génica.

5

Vectores AAV para el suministro de polinucleótidos

En algunas modalidades, los polinucleótidos de las composiciones y métodos descritos en la presente se incorporan en vectores y/o viriones de rAAV para facilitar su introducción en una célula. Los vectores de rAAV útiles en las composiciones y métodos descritos en la presente son construcciones de polinucleótidos recombinantes que incluyen (1) un promotor de OCM descrito en la presente (por ejemplo, un polinucleótido que tiene al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, de identidad de secuencia) con respecto a cualquiera de las secuencias de promotores enumeradas en la Tabla 3 (por ejemplo, cualquiera de las SEQ ID NO: 1-3)), (2) una secuencia heteróloga que se va a expresar (por ejemplo, un polinucleótido que codifica una parte de extremo N o parte de extremo C de una proteína de estereocilina), y (3) secuencias virales que facilitan la integración y expresión de los genes heterólogos. Las secuencias virales pueden incluir aquellas secuencias de AAV que se requieren en cis para la replicación y empaquetamiento (por ejemplo, ITR funcionales) del ADN en un virión. En aplicaciones típicas, el transgén codifica una forma de tipo salvaje de una proteína (por ejemplo, estereocilina) que está mutada en sujetos con formas de pérdida auditiva hereditaria que puede ser útil para mejorar la audición en sujetos que tienen mutaciones que han sido asociadas con pérdida auditiva o sordera (por ejemplo, DFNB16). Tales vectores de rAAV también pueden contener genes indicadores o marcadores. Los vectores rAAV útiles tienen uno o más de los genes de tipo salvaje del AAV suprimidos en su totalidad o en parte, pero retienen las secuencias ITR flanqueantes funcionales. Los ITR de AAV pueden ser de cualquier serotipo adecuado para una aplicación particular. Para usar en los métodos y composiciones descritos en la presente, los ITR pueden ser ITR de AAV2. Los métodos para usar vectores de rAAV se describen, por ejemplo, en Tal et ál., J. Biomed. Sci. 7:279 (2000) y Monahan y Samulski, Gene Delivery 7:24 (2000), las descripciones de cada una de las cuales se incorporan en la presente mediante referencia ya que se refieren a los vectores de AAV para el suministro de genes.

Los polinucleótidos y vectores descritos en la presente (por ejemplo, un promotor OCM unido de forma operativa a un polinucleótido que codifica una parte de extremo N, o en algunas modalidades, una parte de extremo C de la proteína de estereocilina) pueden incorporarse en un virión rAAV para facilitar la introducción del polinucleótido o vector en una célula. Las proteínas de

35

la cápside de AAV componen la parte exterior de ácido no nucleico del virión y están codificadas por el gen cap de AAV. El gen cap codifica tres proteínas de la cubierta viral, VP1, VP2 y VP3, que son necesarias para el ensamblaje del virión. La construcción de viriones de rAAV se ha descrito, por ejemplo, en US 5.173.414; US 5.139.941; US 5.863.541; US 5.869.305; US 6.057.152; y US 6.376.237; así como en Rabinowitz et ál., J. Virol. 76:791 (2002) y Bowles et ál., J. Virol. 77:423 (2003), las descripciones de cada una de las cuales se incorporan a la presente mediante referencia ya que se refieren a los vectores AAV para el suministro de genes.

Los viriones de rAAV útiles junto con las composiciones y métodos descritos en la presente incluyen los derivados de una variedad de serotipos de AAV, que incluyen AAV 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, rh10, rh39, rh43, rh74, Anc80, Anc80L65, DJ/8, DJ/9, 7m8, PHP.B, PHP.eb y PHP.S. Para dirigirse a las células ciliadas, pueden ser especialmente útiles AAV1, AAV2, AAV2quad(Y-F), AAV6, AAV8, AAV9, Anc80, Anc80L65, DJ/9, 7m8 y PHP.B. Los serotipos desarrollados para la transducción de la retina también pueden usarse en los métodos y composiciones descritos en la presente. La construcción y el uso de vectores de AAV y proteínas de AAV de diferentes serotipos se describen, por ejemplo, en Chao et ál., Mol. Ther. 2:619 (2000); Davidson et ál., Proc. Natl. Acad. Sci. USA 97:3428 (2000); Xiao et ál., J. Virol. 72:2224 (1998); Halbert et ál., J. Virol. 74:1524 (2000); Halbert et ál., J. Virol. 75:6615 (2001); y Auricchio et ál., Hum. Molec. Genet. 10:3075 (2001), las descripciones de cada una de las cuales se incorporan a la presente mediante referencia ya que se refieren a los vectores de AAV para el suministro de genes.

También son útiles junto con las composiciones y métodos descritos en la presente los vectores de rAAV pseudotipados. Los vectores pseudotipados incluyen vectores AAV de un serotipo dado (por ejemplo, AAV9) pseudotipados con un gen de la cápside derivado de un serotipo distinto del serotipo dado (por ejemplo, AAV1, AAV2, AAV2quad (Y-F), AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, etc.). Las técnicas que implican la construcción y el uso de viriones rAAV pseudotipados son conocidas en la técnica y se describen, por ejemplo, en Duan et ál., J. Virol. 75:7662 (2001); Halbert et ál., J. Virol. 74:1524 (2000); Zolotukhin et ál., Methods, 28:158 (2002); y Auricchio et ál., Hum. Molec. Genet. 10:3075 (2001).

Los viriones de AAV que tienen mutaciones dentro de la cápside del virión se pueden usar para infectar tipos de células particulares de manera más eficaz que los viriones de la cápside no mutados. Por ejemplo, los mutantes de AAV adecuados pueden tener mutaciones de inserción de ligandos para facilitar el direccionamiento de AAV a tipos de células específicos. La construcción y caracterización de mutantes de cápside de AAV que incluyen mutantes de inserción, mutantes de cribado de alanina y mutantes de etiqueta de epítipo se describen en Wu et ál., J. Virol. 74:8635 (2000). Otros viriones de rAAV que se pueden usar en los métodos descritos en la presente incluyen aquellos híbridos de cápside que se generan por reproducción molecular de virus así

como por transposición de exones. Ver, por ejemplo, Soong et ál., Nat. Genet., 25:436 (2000) y Kolman y Stemmer, Nat. Biotechnol. 19:423 (2001).

Composiciones farmacéuticas

5 Los vectores de ácido nucleico descritos en la presente pueden incorporarse en un vehículo para la administración a un paciente, tal como un paciente humano que padece pérdida auditiva neurosensorial o neuropatía auditiva. Las composiciones farmacéuticas que contienen vectores, tales como vectores virales, que contienen un polinucleótido que codifica una parte de una proteína estereocilina pueden prepararse utilizando métodos conocidos en la técnica. Por ejemplo, 10 tales composiciones se pueden preparar usando, por ejemplo, portadores, excipientes o estabilizadores fisiológicamente aceptables (Remington: The Science and Practice of Pharmacology 22^a edición, Allen, L. Ed. (2013); incorporado a la presente mediante referencia), y en una forma deseada, por ejemplo, en forma de formulaciones liofilizadas o soluciones acuosas. Las mezclas de los vectores de ácido nucleico (por ejemplo, vectores virales) descritos en la 15 presente se pueden preparar en agua mezcladas de manera adecuada con uno o más excipientes, portadores o diluyentes. También pueden prepararse dispersiones en glicerol, polietilenglicoles líquidos y mezclas de estos, así como en aceites. En condiciones habituales de almacenamiento y uso, estas preparaciones pueden contener un conservante para evitar el desarrollo de microorganismos. Las formas farmacéuticas adecuadas para uso inyectable incluyen soluciones o 20 dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de soluciones o dispersiones inyectables estériles (descrito en US 5.466.468, cuya divulgación se incorpora a la presente mediante referencia). En cualquier caso, la formulación puede ser estéril y fluida en la medida en que exista la posibilidad de colocarla en una jeringa fácilmente. Las formulaciones pueden ser estables en las condiciones de fabricación y almacenamiento y pueden conservarse frente a la acción contaminante de microorganismos, tales como bacterias y hongos. El portador 25 puede ser un medio de dispersión o disolvente que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido, y similares), mezclas adecuadas de estos, y/o aceites vegetales. La fluidez adecuada puede mantenerse, por ejemplo, mediante el uso de un recubrimiento, tal como lecitina, mediante el mantenimiento del tamaño de partícula necesario en el caso de la dispersión y mediante el uso de tensioactivos. La prevención de la acción de los 30 microorganismos puede ser provocada por diversos agentes antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido sórbico, timerosal, y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares o cloruro de sodio. La absorción prolongada de las composiciones inyectables se puede provocar mediante el uso en las 35 composiciones de agentes que retrasen la absorción, por ejemplo, monoestearato de aluminio y

gelatina.

Por ejemplo, una solución que contiene una composición farmacéutica descrita en la presente se puede amortiguar de manera adecuada, si es necesario, y el diluyente líquido primero puede volverse isotónico con suficiente solución salina o glucosa. Estas soluciones acuosas particulares son especialmente adecuadas para la administración por vía intravenosa, intramuscular, subcutánea e intraperitoneal. Con relación a esto, los expertos en la técnica conocerán un medio acuoso estéril que puede emplearse en virtud de la presente descripción. Por ejemplo, se puede disolver una dosificación en 1 ml de solución de NaCl isotónica y agregarla a 1000 ml de fluido de hipodermóclisis o inyectarla en el punto de infusión sugerido. Necesariamente se producirá alguna variación en la dosificación dependiendo del estado del sujeto que se esté tratando. Para la administración local en el oído interno, la composición se puede formular para que contenga una solución de perilinfa sintética. Una solución de perilinfa sintética de ejemplo perilinfa incluye NaCl 20-200 mM, KCl 1-5 mM, CaCl_2 0,1-10 mM, glucosa 1-10 mM, y HEPES 2-50 mM, con un pH entre alrededor de 6 y 9 y una osmolalidad de alrededor de 300 mOsm/kg. La persona responsable de la administración, en cualquier caso, determinará la dosis adecuada para el sujeto individual. Además, para la administración a seres humanos, las preparaciones deberán cumplir con los estándares de esterilidad, pirogenicidad, seguridad general, y pureza como lo requieren las normas de la Oficina de Biología de la FDA.

Métodos de tratamiento

Las composiciones descritas en la presente se pueden administrar a un sujeto que tiene o corre riesgo de desarrollar pérdida auditiva neurosensorial mediante una variedad de vías, tales como la administración local en el oído interno (por ejemplo, la administración en la perilinfa o endolinfa, tal como hacia o a través de la ventana oval, la ventana redonda o el canal semicircular (por ejemplo, el canal horizontal), o mediante inyección transtimpánica o intratimpánica, por ejemplo, administración a una OHC), administración intravenosa, parenteral, intradérmica, transdérmica, intramuscular, intranasal, subcutánea, percutánea, intratraqueal, intraperitoneal, intraarterial, intravascular, por inhalación, perfusión, lavado y administración oral. La vía de administración más adecuada en cualquier caso dependerá de la composición particular administrada, el paciente, los métodos de formulación farmacéutica, los métodos de administración (por ejemplo, el tiempo de administración y la vía de administración), la edad del paciente, el peso corporal, el sexo y la gravedad de la enfermedad que se está tratando, la dieta del paciente y la tasa de excreción del paciente. Las composiciones pueden administrarse una vez o más de una vez (por ejemplo, una vez al año, dos veces al año, tres veces al año, bimensualmente, mensualmente). En algunas modalidades, el primer y segundo vector de ácido nucleico se administran simultáneamente (por

ejemplo, en una composición). En algunas modalidades, el primer y segundo vector de ácido nucleico se administran secuencialmente (por ejemplo, el segundo vector de ácido nucleico se administra inmediatamente después del primer vector de ácido nucleico, o 5 minutos, 10 minutos, 15 minutos, 20 minutos, 25 minutos, 30 minutos, 45 minutos, 1 hora, 2 horas, 8 horas, 12 horas, 1 día, 2 días, 7 días, dos semanas, 1 mes o más después del primer vector de ácido nucleico). El primer y segundo vector de ácido nucleico pueden tener el mismo cápside o diferentes cápsides (por ejemplo, cápsides de AAV).

Los sujetos que pueden ser tratados tal como se describe en la presente son sujetos que tienen o están en riesgo de desarrollar pérdida auditiva neurosensorial. Las composiciones y métodos descritos en la presente se pueden usar para tratar sujetos que tienen una mutación en STRC (por ejemplo, una mutación que reduce la función o expresión de STRC, o una mutación STRC asociada con pérdida auditiva neurosensorial, tales como sujetos que tienen DFNB16), sujetos que tienen antecedentes familiares de pérdida o sordera auditiva neurosensorial autosómica recesiva (por ejemplo, antecedentes familiares de pérdida auditiva relacionada con STRC), o sujetos cuyo estado mutacional de STRC y/o nivel de actividad de STRC se desconoce. Los métodos descritos en la presente pueden incluir una etapa de examinación de un sujeto para detectar una mutación en STRC antes del tratamiento o la administración de las composiciones descritas en la presente. Se puede examinar un sujeto por una mutación de STRC mediante el uso de métodos estándar conocidos por los expertos en la técnica (por ejemplo, pruebas genéticas). Los métodos descritos en la presente también pueden incluir una etapa de evaluación de la audición en un sujeto antes del tratamiento o la administración de las composiciones descritas en la presente. La audición se puede evaluar mediante pruebas estándar, tal como audiometría, respuesta auditiva del tronco encefálico (ABR), electrococcleografía (ECOG) y emisiones otoacústicas. Las composiciones y métodos descritos en la presente también se pueden administrar como tratamiento preventivo a pacientes con riesgo de desarrollar pérdida auditiva o neuropatía auditiva, por ejemplo, pacientes que tienen antecedentes familiares de pérdida auditiva hereditaria o pacientes portadores de una mutación STRC que aún no presentan pérdida o deterioro auditivo.

El tratamiento puede incluir la administración de una composición que contenga los vectores de ácido nucleico (por ejemplo, vectores virales de AAV) descritos en la presente en varias dosis unitarias. Cada dosis unitaria contendrá normalmente una cantidad predeterminada de la composición terapéutica. La cantidad a administrar y la vía particular de administración y formulación están dentro del conocimiento de los expertos en la técnica. No es necesario administrar una dosis unitaria con una inyección única, pero puede comprender una infusión continua durante un período de tiempo establecido. La dosificación se puede realizar usando una bomba de jeringa para controlar la velocidad de infusión con el fin de minimizar el daño al oído

interno (por ejemplo, la cóclea). En casos en los que los vectores de ácidos nucleicos son vectores de AAV (por ejemplo, vectores de AAV1, AAV2, AAV2quad(Y-F), AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, rh10, rh39, rh43, rh74, Anc80, Anc80L65, DJ/8, DJ/9, 7m8, PHP.B, PHP.eb, o PHP.S), los vectores virales se pueden administrar al paciente a una dosis de, por ejemplo, de alrededor de 1×10^9 genomas de vector (VG)/ml a alrededor de 1×10^{16} VG/ml (por ejemplo, 1×10^9 VG/ml, 2×10^9 VG/ml, 3×10^9 VG/ml, 4×10^9 VG/ml, 5×10^9 VG/ml, 6×10^9 VG/ml, 7×10^9 VG/ml, 8×10^9 VG/ml, 9×10^9 VG/ml, 1×10^{10} VG/ml, 2×10^{10} VG/ml, 3×10^{10} VG/ml, 4×10^{10} VG/ml, 5×10^{10} VG/ml, 6×10^{10} VG/ml, 7×10^{10} VG/ml, 8×10^{10} VG/ml, 9×10^{10} VG/ml, 1×10^{11} VG/ml, 2×10^{11} VG/ml, 3×10^{11} VG/ml, 4×10^{11} VG/ml, 5×10^{11} VG/ml, 6×10^{11} VG/ml, 7×10^{11} VG/ml, 8×10^{11} VG/ml, 9×10^{11} VG/ml, 1×10^{12} VG/ml, 2×10^{12} VG/ml, 3×10^{12} VG/ml, 4×10^{12} VG/ml, 5×10^{12} VG/ml, 6×10^{12} VG/ml, 7×10^{12} VG/ml, 8×10^{12} VG/ml, 9×10^{12} VG/ml, 1×10^{13} VG/ml, 2×10^{13} VG/ml, 3×10^{13} VG/ml, 4×10^{13} VG/ml, 5×10^{13} VG/ml, 6×10^{13} VG/ml, 7×10^{13} VG/ml, 8×10^{13} VG/ml, 9×10^{13} VG/ml, 1×10^{14} VG/ml, 2×10^{14} VG/ml, 3×10^{14} VG/ml, 4×10^{14} VG/ml, 5×10^{14} VG/ml, 6×10^{14} VG/ml, 7×10^{14} VG/ml, 8×10^{14} VG/ml, 9×10^{14} VG/ml, 1×10^{15} VG/ml, 2×10^{15} VG/ml, 3×10^{15} VG/ml, 4×10^{15} VG/ml, 5×10^{15} VG/ml, 6×10^{15} VG/ml, 7×10^{15} VG/ml, 8×10^{15} VG/ml, 9×10^{15} VG/ml, o 1×10^{16} VG/ml) en un volumen de 1 μ l a 200 μ l (por ejemplo, 1, 2, 3, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, o 200 μ l). Los vectores de AAV se pueden administrar a un sujeto a una dosis de alrededor de 1×10^7 VG/oído a alrededor de 2×10^{15} VG/oído (por ejemplo, 1×10^7 VG/oído, 2×10^7 VG/oído, 3×10^7 VG/oído, 4×10^7 VG/oído, 5×10^7 VG/oído, 6×10^7 VG/oído, 7×10^7 VG/oído, 8×10^7 VG/oído, 9×10^7 VG/oído, 1×10^8 VG/oído, 2×10^8 VG/oído, 3×10^8 VG/oído, 4×10^8 VG/oído, 5×10^8 VG/oído, 6×10^8 VG/oído, 7×10^8 VG/oído, 8×10^8 VG/oído, 9×10^8 VG/oído, 1×10^9 VG/oído, 2×10^9 VG/oído, 3×10^9 VG/oído, 4×10^9 VG/oído, 5×10^9 VG/oído, 6×10^9 VG/oído, 7×10^9 VG/oído, 8×10^9 VG/oído, 9×10^9 VG/oído, 1×10^{10} VG/oído, 2×10^{10} VG/oído, 3×10^{10} VG/oído, 4×10^{10} VG/oído, 5×10^{10} VG/oído, 6×10^{10} VG/oído, 7×10^{10} VG/oído, 8×10^{10} VG/oído, 9×10^{10} VG/oído, 1×10^{11} VG/oído, 2×10^{11} VG/oído, 3×10^{11} VG/oído, 4×10^{11} VG/oído, 5×10^{11} VG/oído, 6×10^{11} VG/oído, 7×10^{11} VG/oído, 8×10^{11} VG/oído, 9×10^{11} VG/oído, 1×10^{12} VG/oído, 2×10^{12} VG/oído, 3×10^{12} VG/oído, 4×10^{12} VG/oído, 5×10^{12} VG/oído, 6×10^{12} VG/oído, 7×10^{12} VG/oído, 8×10^{12} VG/oído, 9×10^{12} VG/oído, 1×10^{13} VG/oído, 2×10^{13} VG/oído, 3×10^{13} VG/oído, 4×10^{13} VG/oído, 5×10^{13} VG/oído, 6×10^{13} VG/oído, 7×10^{13} VG/oído, 8×10^{13} VG/oído, 9×10^{13} VG/oído, 1×10^{14} VG/oído, 2×10^{14} VG/oído, 3×10^{14} VG/oído, 4×10^{14} VG/oído, 5×10^{14} VG/oído, 6×10^{14} VG/oído, 7×10^{14} VG/oído, 8×10^{14} VG/oído, 9×10^{14} VG/oído, 1×10^{15} VG/oído, o 2×10^{15} VG/oído).

Las composiciones descritas en la presente se administran en una cantidad suficiente para mejorar la audición, aumentar la expresión de una proteína estereocilina (por ejemplo, una proteína

estereocilina WT, tal como una proteína estereocilina que tiene la secuencia de la SEQ ID NO: 4 o la SEQ ID NO: 5, por ejemplo, expresión en una célula ciliada coclear, por ejemplo, una célula ciliada externa), aumentar la función de la estereocilina, mejorar la estructura de la OHC, mejorar la función de la OHC, prevenir o reducir el daño o la muerte de la OHC, mejorar la fijación del haz ciliado de la OHC a la membrana tectorial, o aumentar o mejorar la supervivencia de la OHC. La audición se puede evaluar mediante pruebas de audición estándar (por ejemplo, audiometría, ABR, electrococcleografía (ECOG) y emisiones otoacústicas) y se puede mejorar mediante 5 % o más (por ejemplo, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 125 %, 150 %, 200 % o más) en comparación con las mediciones de audición obtenidas antes del tratamiento. En algunas modalidades, las composiciones se administran en una cantidad suficiente para mejorar la capacidad del sujeto para comprender el habla. Las composiciones descritas en la presente también se pueden administrar en una cantidad suficiente para ralentizar o prevenir el desarrollo o el avance de la pérdida auditiva neurosensorial (por ejemplo, en sujetos que portan una mutación genética en el gen STRC que está asociado con la pérdida auditiva o que tienen antecedentes familiares de pérdida auditiva (por ejemplo, pérdida auditiva autosómica recesiva) pero no presentan discapacidad auditiva o en sujetos que presentan una pérdida auditiva leve a moderada) La expresión de estereocilina puede evaluarse mediante el uso de inmunohistoquímica, análisis de transferencia Western, PCR cuantitativa en tiempo real u otros métodos conocidos en la técnica para la detección de proteínas o ARNm, y puede aumentarse mediante 5 % o más (por ejemplo, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 125 %, 150 %, 200 % o más) en comparación con la expresión de estereocilina antes de la administración de las composiciones descritas en la presente. La función de OHC o la función de la proteína de estereocilina codificada por los vectores de ácido nucleico administrados al sujeto se pueden evaluar de forma indirecta basándose en las pruebas de audición, y puede aumentarse mediante 5 % o más (por ejemplo, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 125 %, 150 %, 200 % o más) en comparación con la función de OHC o la función de la proteína antes de la administración de las composiciones descritas en la presente. Estos efectos pueden ocurrir, por ejemplo, dentro de 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas, 10 semanas, 15 semanas, 20 semanas, 25 semanas o más, tras la administración de las composiciones descritas en la presente. El paciente puede ser evaluado 1 mes, 2 meses, 3 meses, 4 meses, 5 meses, 6 meses o más después de la administración de la composición dependiendo de la dosis y vía de administración usada para el tratamiento. Dependiendo del resultado de la evaluación, el paciente puede recibir tratamientos adicionales.

Kits

Las composiciones descritas en la presente se pueden proporcionar en un kit para usar en el tratamiento de un sujeto con pérdida auditiva neurosensorial, tal como pérdida auditiva neurosensorial asociada con una mutación en el gen de STRC. Las composiciones pueden incluir los polinucleótidos descritos en la presente (por ejemplo, un polinucleótido que tenga al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia) con cualquiera de las secuencias promotoras de OCM enumeradas en la Tabla 3 (por ejemplo, cualquiera de las SEQ ID NO: 1-3) unidas de forma operativa a un polinucleótido que codifica una parte de extremo N de la proteína estereocilina, y un polinucleótido que codifica una parte de extremo C de la proteína estereocilina) y sistemas de vectores de ácido nucleico (por ejemplo, sistemas de dos vectores descritos en la presente) que contienen tales polinucleótidos. Los vectores de ácido nucleico se pueden empaquetar en una cápside de virus AAV (por ejemplo, AAV1, AAV2, AAV2quad(Y-F), AAV6, AAV8, AAV9, Anc80, Anc80L65, DJ/9, 7m8 o PHP.B). El kit puede incluir además un prospecto que instruya a un usuario del kit, tal como un médico, que realice los métodos descritos en la presente. El kit puede incluir opcionalmente una jeringa u otro dispositivo para administrar la composición.

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los expertos en la técnica una descripción de cómo se pueden usar, fabricar y evaluar las composiciones y métodos que se describen en la presente, y se pretende que sean puramente a modo ilustrativo de la divulgación y no se pretende que limiten el alcance de lo que los inventores consideran su divulgación.

Ejemplo 1. La secuencia promotora OCM induce la expresión del transgén en las OHC de la cóclea murina in vivo

Para determinar la eficacia del promotor OCM construido (SEQ ID NO: 1) para inducir la expresión del transgén en las OHC in vivo, se transdujo la cóclea de ratón con un vector AAV que expresaba GFP bajo el control del promotor del citomegalovirus (CMV), o con un vector AAV que expresaba GFP bajo el control del promotor OCM. Específicamente, se infundió el virus AAV-OCM-GFP a través del canal semicircular posterior a ratones CBA/CaJ de dos días de edad a una dosis de 7.7×10^9 genomas vectoriales por oído. Los ratones se recuperaron de la cirugía y fueron eutanasiados y perfundidos con formalina amortiguada normal al 10 % 19 días después. El hueso temporal del oído interno se extrajo y descalcificó en EDTA al 8 % durante tres días. La cóclea se diseccionó del hueso temporal descalcificado, se inmunotizó con el anticuerpo Myosin 7a (Myo7a)

para etiquetar todas las células ciliadas y se montó en un portaobjetos para obtener imágenes confocales. Se muestra la fluorescencia de GFP nativa. Mediante el uso de un promotor ubicuo, el AAV-CMV-GFP indujo la expresión de GFP en muchos tipos celulares de la cóclea, incluidas las células ciliadas internas, las células ciliadas externas, las neuronas ganglionares espirales, las células mesenquimales y la glía (Figura 1A). Utilizando el promotor específico de células ciliadas externas, AAV-OCM (SEQ ID NO: 1)-GFP indujo la expresión de GFP exclusivamente en células ciliadas externas (Figura 1B).

Ejemplo 2. La expresión de GFP impulsada por el promotor OCM se enriquece en las células ciliadas externas del órgano de Corti de primates no humanos

Para comprobar la especificidad del promotor OCM de la SEQ ID NO: 1 en primates no humanos in vivo, se inyectó en los oídos de primates no humanos (*Macaca fascicularis*) un vector AAV que incluía H2B-eGFP dirigida al núcleo y unida de forma operativa al promotor OCM de la SEQ ID NO: 1. A los primates no humanos adultos se les inyectaron 40 μ l de vector ($3,41 \times 10^{13}$ vg/ml) para el vector AAV que expresa H2B-eGFP bajo el control del promotor OCM de la SEQ ID NO: 1 a través de la membrana de ventana redonda, una fenestración en el canal semicircular lateral que permitió el eflujo de perilinfa durante el procedimiento.

Después de cuatro semanas de vida, los animales fueron sacrificados y fijados en NBF al 10 % mediante perfusión cardíaca, sus huesos temporales fueron cosechados y mantenidos en NBF al 10 % durante 4-10 días adicionales. Los oídos se descalcificaron en ácido fórmico (inmunocal) durante 6 días, se embebieron en parafina y se seccionaron en cortes de 5 μ m.

Las secciones se etiquetaron con un anticuerpo contra la GFP y se tiñeron con un anticuerpo secundario conjugado con fosfatasa alcalina; se desarrolló una tinción cromática roja por la reacción del colorante rojo rápido con la fosfatasa alcalina del anticuerpo secundario. Las secciones se contrateñían con Hematoxilina en azul para visualizar todos los núcleos y se visualizaban en una cámara en color con un aumento de 20x y se convertían a escala de grises (Figuras 2A-2B, imágenes superiores (paneles A y B, respectivamente) - La Figura 2A es una micrografía de una sección de parafina de un primer animal y la Figura 2B es una micrografía de una sola sección de parafina de un segundo animal). Para visualizar la señal roja de la tinción cromática anti-GFP, se extrajo todo el color azul (Hematoxilina) de las imágenes en color utilizando el software de procesamiento de imágenes GIMP utilizando la herramienta de selección de color, antes de convertir a escala de grises. Solo los núcleos con señal roja nuclear H2B-GFP permanecieron visibles (Figuras 2A-2B, imágenes inferiores (paneles A' y B', respectivamente)). Las barras de escala representan 100 μ m. Las células ciliadas internas (IHC) y las células ciliadas externas (OHC) están resaltadas a modo de orientación.

Ejemplo 3. Un sistema de vector dual superpuesto Anc80-CMV-mStrc rescató la audición en ratones deficientes en estereocilina

Se utilizó la tecnología CRISPR-Cas9 para generar ratones deficientes en estereocilina en la cepa de fondo CBA/CaJ al crear un desplazamiento de marco en la posición del par de bases 232 de STRC. Los animales de tipo salvaje de la cepa de fondo CBA/CaJ mostraron una clara tinción de anticuerpos de estereocilina en las puntas de los estereocilios de las células ciliadas externas (OHC) (Figura 3A, panel inferior), mientras que los animales Strc-KO de 232 pb carecían de señal para el anticuerpo (Figura 3B, panel inferior). La STRC murina se encapsuló en vectores Anc80 duales, donde el primer vector llevaba un promotor CMV y los nucleótidos 1-3200 del ADNc de la STRC murina y el segundo llevaba los aminoácidos 2201-5430, creando una superposición de 1000 pb entre las dos mitades del ADNc de longitud completa. Tras la administración de ambos vectores a una concentración de $1E10$ vg/oído en la cóclea de ratones Strc-KO de 232 pb postnatales tempranos a través de su canal semicircular posterior, pudo observarse la expresión de proteína estereocilina de-novo en las puntas de la OHC y en el cuerpo de las células ciliadas internas del órgano de Corti en ratones Strc-KO de 232 pb tratados (Figura 3C, panel inferior). Se evaluaron las emisiones otoacústicas de productos de distorsión (DPOAE) como lectura directa de la función de la OHC y se midió la respuesta auditiva del tronco encefálico (ABR) como medida de una vía auditiva ascendente intacta cuatro semanas después del tratamiento con Anc80-CMV-mStrc (superposición). Los oídos contralaterales no tratados de los animales Strc-KO de 232 pb ("oídos no tratados") mostraron DPOAE casi ausentes y umbrales ABR muy elevados que indican la pérdida de la función OHC (Figuras 4A-4B, círculos abiertos), mientras que los animales Strc-KO de 232 pb tratados ("hom tratados") mostraron una recuperación de los umbrales auditivos (Figuras 4A-4B, círculos rellenos). El mejor respondedor (Figuras 4A-4B, cuadrados negros) de los animales Strc-KO de 232 pb tratados mostró umbrales auditivos cercanos al tipo salvaje ("CBA/CaJ") (Figuras 4A-4B, triángulos). Se observó que una fracción elevada de las OHC de ratones Strc-KO de 232 pb que expresaban estereocilina tras el tratamiento con AAV-Anc80-CMV-mStrc favorecía la recuperación de la audición (Figura 4C).

Ejemplo 4. Un sistema de inteína dividida en dos vectores reconstituyó la estereocilina de longitud completa in vitro

Para generar los plásmidos experimentales, el ADN que codifica los aminoácidos 1-746 de la estereocilina ("N-Strc") se fusionó genéticamente con el ADN que codifica el fragmento de N-inteína Npu (SEQ ID NO: 41, que codifica la N-inteína Npu de la SEQ ID NO: 27) y se clonó en un plásmido que contiene el promotor CMV constitutivamente activo para generar CMV.N-Strc-N-Npu.

El ADN que codifica los aminoácidos 747-1809 de la estereocilina ("C-Strc") se fusionó genéticamente aguas abajo del ADN que codifica el fragmento de C-inteína de la Npu (SEQ ID NO: 42, que codifica la C-inteína de la Npu de la SEQ ID NO: 28) y se clonó en un plásmido que contenía el promotor CMV para producir CMV.C-Npu-C-Strc. Como control, también se clonó la

5 secuencia codificante completa de la estereocilina ("FL-Strc") en un plásmido CMV para generar CMV.FL-Strc. Se utilizó CMV.GFP como control negativo.

Las células HEK293T se transfectaron con plásmidos de control o con una combinación de plásmidos N-Strc y C-Strc utilizando el kit Lipofectamine 3000 (Life Technologies) y se incubaron en condiciones estándar de cultivo celular durante tres días. Los cultivos celulares se enjuagaron

10 con PBS y las células se lisaron para extraer proteínas. Las concentraciones de lisado de proteínas se midieron utilizando el ensayo BCA, y se cargó una masa constante de proteínas para análisis de transferencia Western mediante el uso de anticuerpos contra beta actina y estereocilina. Las medidas de densitometría de las intensidades de las bandas de proteínas se utilizaron para determinar la cantidad relativa de estereocilina de longitud completa de la muestra.

15 Como se muestra en las Figuras 5A-5B, los diseños de inteína probados produjeron una banda de estereocilina de longitud completa.

Ejemplo 5. Sistemas de vectores duales AAV para la expresión de estereocilina impulsada por el promotor OCM

20 Se diseñó un sistema de vector dual AAV que utiliza una superposición de secuencias codificantes de estereocilina para la recombinación homóloga utilizando dos plásmidos. El primer plásmido contenía el promotor OCM murino de la SEQ ID NO: 1 unido de forma operativa a una parte de extremo N de STRC murino codificada por los primeros 3200 nucleótidos de la secuencia codificante y flanqueada por las secuencias ITR 5' y 3' (plásmido P959; SEQ ID NO: 43; Figura

25 6A). El segundo plásmido (P724; SEQ ID NO: 44; Figura 6B) contenía, en orden de 5' a 3', una superposición de 500 nucleótidos de la secuencia STRC codificada por el primer plásmido, seguido de 2730 nucleótidos que codificaban la parte de extremo C restante de la STRC murina, un elemento regulador postranscripcional de Woodchuck (WPRE) y una secuencia de poliadenilación de la hormona de crecimiento bovina (bGH). Estos elementos del segundo

30 plásmido también estaban flanqueados por las secuencias ITR 5' y 3'.

Los vectores virales AAV se sintetizan al transfectar células HEK293T con uno de estos plásmidos junto con un plásmido que contiene rep/cap y un plásmido ayudante adenoviral utilizando protocolos estándar. Los plásmidos se empaquetan en el vector del serotipo AAV8 utilizando métodos estándar y se obtienen de un proveedor comercial. El medio de cultivo celular y las

35 células se recolectan posteriormente para extraer y purificar el AAV. El AAV de las células se libera

de las células mediante tres ciclos de congelación y descongelación, y se recoge el medio de cultivo celular para obtener AAV secretado. El AAV del medio de cultivo celular se concentra al agregar PEG8000 a la solución, se incuba a 4 ° C y se centrifuga para recoger las partículas de AAV. Todos los AAV se pasan por centrifugación en gradiente de densidad de iodixanol para purificar las partículas de AAV, y el amortiguador se cambia a PBS con pluronic F68 al 0,01 % pasando el AAV purificado y el amortiguador sobre una columna de centrifugación con un corte de peso molecular de 100 kDa. Los vectores AAV resultantes de cada uno de los dos plásmidos se utilizan en combinación mediante su administración en los oídos de ratones (por ejemplo, administración local en el oído interno).

Un diseño alternativo de un sistema de vector dual AAV utilizó un primer plásmido que contenía, en orden 5' a 3', una secuencia ITR 5', el promotor OCM murino de la SEQ ID NO: 1 unido de forma operativa a 2700 nucleótidos que codifican una parte de extremo N de estereocilina murina, un donante de empalme AP, una secuencia de cabeza AP y una secuencia ITR 3' (P960; SEQ ID NO: 45; Figura 7A). El segundo plásmido de este sistema contenía, en orden 5' a 3', una secuencia ITR 5', una secuencia de cabeza AP homóloga a la del primer plásmido, un aceptor de empalme AP, los 2730 nucleótidos que codifican la parte de extremo C restante de la estereocilina murina, una secuencia bGH poli-A y una ITR 3' (P726; SEQ ID NO: 46; Figura 7B). Los vectores virales AAV que incorporan partes de estos dos plásmidos se sintetizan como se ha descrito anteriormente.

Ejemplo 6. Administración de una composición que contiene un sistema de dos vectores que contiene un promotor OCM unido de forma operativa a una secuencia codificante de estereocilina a un sujeto con pérdida auditiva neurosensorial.

De acuerdo con los métodos descritos en la presente, un médico experto en la técnica puede tratar a un paciente, tal como un paciente humano, con pérdida auditiva neurosensorial (por ejemplo, pérdida auditiva neurosensorial asociada con una mutación en STRC, tal como DFNB16) para mejorar o restaurar la audición. Con este fin, un médico experto en la materia puede administrar al paciente humano una composición que contiene un sistema de expresión de ácido nucleico de dos vectores, tal como un sistema que utiliza dos vectores AAV (por ejemplo, vectores AAV1, AAV2, AAV2quad(Y-F), AAV6, AAV9, Anc80, Anc80L65, DJ/9, 7m8 o PHP.B) que incluyen colectivamente un promotor OCM (por ejemplo, un polinucleótido que tenga al menos un 85 % de identidad de secuencia (por ejemplo, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, o más, identidad de secuencia) con la secuencia de cualquiera de las SEQ ID NO: 1-3) unido de forma operativa a un transgén STRC.

El sistema de dos vectores puede ser un sistema de vector dual superpuesto que contenga un

primer y un segundo vector AAV. El sistema de vector dual superpuesto puede incluir un primer vector AAV que incluye el promotor OCM unido de forma operativa a un polinucleótido que codifica una parte de extremo N de la proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4) y un segundo vector AAV que incluye un polinucleótido que codifica una parte de extremo C de la proteína estereocilina, donde el extremo 3' de la secuencia codificante de estereocilina en el primer vector se superpone con el extremo 5' de la secuencia codificante de estereocilina en el segundo vector. En otro ejemplo, el sistema de dos vectores puede ser un sistema de vector dual superpuesto que contenga un primer y un segundo vector AAV. El sistema de vector dual de transempalme puede incluir un primer vector AAV que incluya el promotor OCM unido de forma operativa a un polinucleótido que codifique una parte de extremo N de la proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4) y una secuencia señal donante de empalme 3' del polinucleótido y un segundo vector AAV que incluye una secuencia señal aceptora de empalme 5' de un polinucleótido que codifica una parte de extremo C de la proteína estereocilina. En otro ejemplo, el sistema de dos vectores puede ser un sistema de vector híbrido dual que contiene un primer y un segundo vector AAV. El sistema de vector híbrido dual puede incluir un primer vector AAV que incluye el promotor OCM unido de forma operativa a un polinucleótido que codifique una parte de extremo N de la proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4), una secuencia señal donante de empalme 3' del polinucleótido, y una primera región recombinogénica 3' de la secuencia señal donante de empalme, y un segundo vector AAV que incluye una segunda región recombinogénica, una secuencia señal aceptora de empalme 3' de la región recombinogénica, y un polinucleótido que codifica una parte de extremo C de la proteína estereocilina 3' de la secuencia señal aceptora de empalme. En aún otro ejemplo, el sistema de dos vectores puede ser un sistema de transempalme de inteína dividido que incluye un primer vector AAV y un segundo vector AAV. El sistema de dos vectores de transempalme de inteína dividida puede incluir un primer vector AAV que incluye el promotor OCM unido de forma operativa a un polinucleótido que codifica una parte de extremo N de la proteína estereocilina (por ejemplo, una parte de extremo N de la SEQ ID NO: 4) y un polinucleótido que codifica una inteína de extremo N (N-inteína) 3' de esta, y un segundo vector AAV que incluye el promotor OCM unido de forma operativa a un polinucleótido que codifica una inteína de extremo C (C-inteína) y un polinucleótido que codifica una parte de extremo C de la proteína estereocilina 3' de esta. Los sistemas de dos vectores mencionados pueden incluir además secuencias reguladoras tales como, por ejemplo, potenciadores, secuencias poli(a) y regiones no traducidas (UTR; por ejemplo, 5' UTR y 3' UTR).

La composición que contiene los vectores AAV puede administrarse al paciente, por ejemplo, mediante administración local en el oído interno (por ejemplo, inyección en la perilinfa o a través de

la membrana de la ventana redonda), para tratar la pérdida auditiva neurosensorial.

Después de la administración de la composición a un paciente, un médico experto en la técnica puede controlar la expresión de la proteína terapéutica codificada por el transgén y la mejora del paciente en respuesta a la terapia, mediante una variedad de métodos. Por ejemplo, un médico puede controlar la audición del paciente al realizar pruebas estándar, tales como audiometría, ABR, electrococcleografía (ECOG) y emisiones otoacústicas después de la administración de la composición. Un hallazgo de que el paciente exhibe una audición mejorada en una o más de las pruebas después de la administración de la composición en comparación con los resultados de las pruebas de audición antes de la administración de la composición indica que el paciente está respondiendo favorablemente al tratamiento. Las dosis posteriores pueden determinarse y administrarse según sea necesario.

Otras modalidades

Será evidente para los expertos en la técnica que se pueden realizar diversas modificaciones y variaciones de la divulgación descrita sin apartarse del espíritu o el alcance de la descripción. Aunque la divulgación se ha descrito en relación con modalidades específicas, debe entenderse que la divulgación reivindicada no debe limitarse indebidamente a tales modalidades específicas. De hecho, varias modificaciones de los modos descritos para llevar a cabo la divulgación que son obvias para los expertos en la técnica pretenden estar dentro del alcance de la divulgación. Otras modalidades se encuentran en las reivindicaciones.