



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 112020007765-8 A2



(22) Data do Depósito: 17/10/2018

(43) Data da Publicação Nacional: 20/10/2020

(54) **Título:** COMPOSIÇÕES DE VÍRUS ADENO-ASSOCIADOS PARA RESTAURAR A FUNÇÃO DO GENE HBB E MÉTODOS DE USO DAS MESMAS

(51) **Int. Cl.:** A61K 48/00; C12N 7/04; C12N 15/864; A61K 35/76; C12N 15/90.

(30) **Prioridade Unionista:** 18/10/2017 US 62/574,163; 24/01/2018 US 62/621,102.

(71) **Depositante(es):** CITY OF HOPE; HOMOLOGY MEDICINES, INC..

(72) **Inventor(es):** SASWATI CHATTERJEE; KAMEHAMEHA K. WONG; MARWA BENHAJSALAH; LAURA JANE SMITH; ALBERT BARNES SEYMOUR; JASON BOKE WRIGHT; JAMES ANTHONY MCSWIGGEN; SERENA NICOLE DOLLIVE; THIA BABOVAL ST. MARTIN; JAIME PROUT.

(86) **Pedido PCT:** PCT US2018056271 de 17/10/2018

(87) **Publicação PCT:** WO 2019/079437 de 25/04/2019

(85) **Data da Fase Nacional:** 17/04/2020

(57) **Resumo:** São providas aqui composições de vírus adeno-associados (AAV) para corrigir uma mutação em um gene de beta globina (HBB) e métodos de usar os mesmos para corrigir uma mutação do gene HBB em uma célula. Também são providos sistemas de empacotamento para produzir composições de vírus adenoassociados.

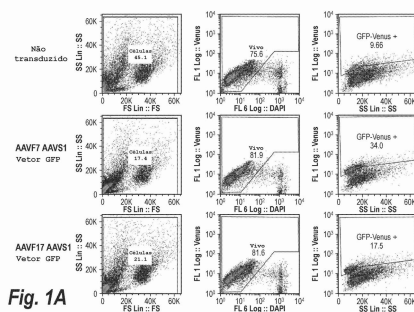


Fig. 1A

COMPOSIÇÕES DE VÍRUS ADENO-ASSOCIADOS PARA RESTAURAR A**FUNÇÃO DO GENE HBB E MÉTODOS DE USO DAS MESMAS****REFERÊNCIA CRUZADA DA PEDIDOS RELACIONADOS**

[0001]Este pedido reivindica o benefício dos Pedidos Provisórios US Nos.: 62/574,163, depositado em 18 de outubro de 2017; e 62/621.102, depositado em 24 de janeiro de 2018, cada sendo aqui incorporado por referência em sua totalidade.

LISTAGEM DE SEQUÊNCIAS

[0002]O presente pedido contém uma Listagem de Sequências que foi apresentada eletronicamente em formato ASCII, criada em 12 de outubro de 2018, denominado 606107_HMT-023PC_Sequence_Listing_ST25.txt, e tem tamanho de 200.630. A Listagem de Sequências é aqui incorporada por referência em sua totalidade.

DECLARAÇÃO DE INTERESSE DO GOVERNO

[0003]Esta invenção foi financiada pelo Governo Federal sob a bolsa n° P30CA033572, concedida pelo National Institutes of Health. O governo tem determinados direitos sobre a invenção.

FUNDAMENTO

[0004]Hemoglobinopatias compreendem uma família de distúrbios genéticos em que a produção, estrutura e/ou função de uma proteína de hemoglobina é anormal. Proteína de hemoglobina representa aproximadamente 97% do peso seco de eritrócitos e aumenta a capacidade de transporte de oxigênio do sangue em cerca de setenta vezes. A proteína de hemoglobina adulta predominante é composta por duas subunidades de alfa globina (HBA), duas subunidades de beta globina (HBB) e um grupo heme associado a cada subunidade.

Defeitos genéticos em HBB no cromossomo 11 podem causar certas hemoglobinopatias, como doença de células falciformes (SCD) e beta-talassemia.

[0005]Doença de células falciformes, também chamada anemia falciforme, é uma doença autossômica recessiva afetando aproximadamente 100.000 americanos. Ela é prevalente entre afro-americanos, mas também está presente em outros grupos étnicos. Na África Ocidental e Central, 1-2% de todos os bebês nascem com SCD. SCD é causada por mutação homozigótica em nucleotídeo 20 na sequência de codificação de HBB. Essa mutação substitui o aminoácido glutamato carregado negativamente (codificado por GAG) por um resíduo hidrofóbico neutro, valina (codificada por GUG) no sexto aminoácido da beta globina madura. Hemoglobina contendo uma cadeia beta globina tendo a mutação SCD tende a se agregar em polímeros de vários filamentos, que distorcem o formato de eritrócitos, tornando essas células frágeis e conformadas como crescentes ou foices. Esses eritrócitos anormais estão mais dispostos para hemólise e entregam menos oxigênio aos tecidos e órgãos. Além disso, a agregação de hemoglobina torna os eritrócitos rígidos e facilmente retidos em pequenos vasos sanguíneos, diminuindo, assim, o fluxo sanguíneo e causando oclusão vascular. Como um resultado, pacientes com SCD sofrem de anemia e episódios de dor chamados "crise", e dano aos órgãos durante a crise é a causa principal da mortalidade e morbidade associada com SCD. Em particular, infarto (isto é, necrose de tecido devido ao suprimento insuficiente de sangue) de osso, baço, rim e pulmões é particularmente comum. Em contraste, pessoas que são heterozigotas para a

mutação de células falciformes são amplamente assintomáticas.

[0006]Beta-talassemia afeta cerca de 1 em cada 100.000 indivíduos em todo o mundo e cerca de 1 em 10.000 pessoas na União Europeia. Beta-talassemia é um grupo de distúrbios causados por várias mutações em HBB que reduzem expressão de beta globina. Até o momento, 884 mutações diferentes, incluindo substituição, inserções e deleções, foram identificadas em beta-talassemia (banco de dados HbVar). Essas mutações estão localizadas em todo o *locus* genômico de HBB. Entre as substituições, inserções e pequenas deleções, variantes patogênicas foram encontradas a montante de 5' UTR e 3' UTR. Além destas 884 mutações conhecidas, variantes adicionais podem ser patogênicas apenas em conjunto com variação específica na sequência de HBB. As mutações de beta-talassemia podem afetar transcrição de genes, processamento de RNA, modificação pós-transcricional, tradução de mRNA, etc. Beta-talassemia é altamente variável em gravidade, com algumas mutações HBB levando à perda completa da produção de beta globina e outras mutações HBB levando a apenas uma redução na quantidade de beta globina. Pacientes com beta talassemia severa (isto é, talassemia major), que com frequência têm mutação(s) de HBB em ambos os alelos, sofrem de anemia, retardo de crescimento e desenvolvimento anormal de órgãos. Pacientes com beta-talassemia leve a moderada (isto é, talassemia menor ou talassemia intermediária) manifestam sintomas menos graves.

[0007]Hemoglobinopatia pode ser controlada por transfusão de sangue e cuidados de suporte, com SCD e beta-

talassemia major requerendo transfusões crônicas. No entanto, transfusões repetidas resultam em sobrecarga de ferro e requerem terapia de quelação de ferro para reduzir a incidência de complicações. Morbidez e mortalidade da SCD e beta-talassemia também podem ser atenuadas por hidroxiureia, o único fármaco aprovado pela FDA até o momento para SCD. No entanto, este tratamento não é amplamente usado devido à sua baixa taxa de prescrição e baixa adesão.

[0008] A fim de curar SCD ou a beta-talassemia, pacientes precisam receber células-tronco hematopoiéticas carregando, pelo menos, uma cópia da beta-hemoglobina funcional que pode ser adequadamente expressa. Uma abordagem consiste em obter células-tronco hematopoiéticas de tipo selvagem a partir de um doador alogênico através de transplante de medula óssea. No entanto, a disponibilidade de doadores compatíveis é um fator limitante importante, e o transplante de medula óssea é frequentemente associado com sérias complicações que levam a uma taxa de mortalidade de 5-10%. Mais recentemente, abordagens de terapia com genes foram empregadas para introduzir uma beta globina expressando polinucleotídeo ex vivo em células-tronco hematopoiéticas mutantes isoladas do paciente.

[0009] Até agora, todas as experiências clínicas de terapia com genes HBB envolveram o uso de vetores retrovirais, como vetores lentivirais. No entanto, a terapia com genes baseada em retrovírus levanta várias preocupações de segurança e eficácia. Por exemplo, como a inserção de vetores retrovirais no genoma humano não é

alvo-marcada, existe o risco de o vetor romper um gene supressor de tumor ou ativar um oncogene, causando, assim, uma malignidade. De fato, em um ensaio clínico para tratar imunodeficiência combinada severa ligada a X (SCID) pela transdução de precursores de medula óssea CD34⁺ com um vetor gama-retroviral, quatro em cada dez pacientes desenvolveram leucemia (Hacein-Bey-Abina *et al.*, *J Clin Invest.* (2008) 118(9):3132-42). Além disso, devido a essas preocupações de segurança, a terapia genética lentiviral só pode ser realizada *ex vivo*. Esse uso *ex vivo* reduz a eficácia da terapia porque o número de células-tronco hematopoiéticas que podem ser extraídas de um indivíduo para terapia *ex vivo* é apenas uma fração pequena daquelas presentes no indivíduo, e atualmente não há método confiável em uso clínico para expandir células-tronco hematopoiéticas *ex vivo*.

[0010] Também foi especulado que tecnologias de edição de genes baseadas em nucleases, como meganucleases, nucleases de dedo de zinco (ZFNs), nucleases efetoras do tipo ativador de transcrição (TALENs) e tecnologia de repetições palindrômicas curtas agrupadas e regularmente interespaçadas (CRISPR) podem ser usadas para corrigir defeito no gene HBB em pacientes com SCD e beta-talassemia. No entanto, cada uma dessas tecnologias levanta questões de segurança devido ao potencial para mutação de sítios fora do alvo no genoma humano, similar em sequência ao sítio alvo pretendido.

[0011] Conseqüentemente, existe uma necessidade na técnica de composições e métodos de terapia gênica aperfeiçoados que possam restaurar de modo eficaz e seguro

a função de gene HBB em pacientes com SCD e beta-talassemia.

SUMÁRIO

[0012]São providas aqui composições de vírus adeno-associados (AAV) para corrigir uma mutação em um gene HBB e métodos de usar as mesmas para corrigir uma mutação do gene HBB em uma célula. Também são providos sistemas de empacotamento para produzir composições de vírus adeno-associado.

[0013]As composições de AAV e métodos divulgados aqui são particularmente vantajosos em que eles permitem uma correção altamente eficaz de mutações em um gene HBB *in vivo*, sem a necessidade para clivagem de DNA genômico usando uma nuclease exógena (por exemplo, uma meganuclease, uma nuclease 'dedo-de-zinco', uma nuclease tipo ativador transcripcional (TALEN), ou uma nuclease RNA-guiada tal como uma Cas9).

[0014]Consequentemente, em um aspecto, a presente divulgação provê um vírus adeno-associado defeituoso na replicação (AAV) compreendendo (a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F, e (b) um genoma de correção compreendendo (i) um elemento de edição para editar um *locus* alvo em um gene alvo, (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo, e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo.

[0015]Em outro aspecto, a presente divulgação provê um

método para corrigir uma mutação em um gene de beta globina (HBB) em uma célula, o método compreendendo transduzir a célula com um vírus adeno-associado defeituoso na replicação (AAV) compreendendo (a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F, e (b) um genoma de correção compreendendo: (i) um elemento de edição para editar um *locus* alvo em um gene alvo; (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo, em que a célula é transduzida sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[0016]Em certas modalidades, a célula é uma célula-tronco pluripotente. Em certas modalidades, a célula é uma célula-tronco hematopoiética. Em certas modalidades, a célula é uma célula-tronco hematopoiética CD34⁺. Em certas modalidades, a célula é de um indivíduo mamífero e o AAV é administrado ao indivíduo em uma quantidade eficaz para transduzir a célula no indivíduo.

[0017]Em outro aspecto, a presente divulgação provê um método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método compreendendo (a) transduzir uma célula progenitora de eritrócitos do indivíduo *ex vivo* com um AAV defeituoso na replicação compreendendo um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e um genoma de

correção compreendendo: (i) um elemento de edição para editar um *locus* alvo em um gene alvo; (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo; e (b) administrar a célula transduzida para o indivíduo, em que a célula é transduzida sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[0018]Em certas modalidades, a célula progenitora de eritrócitos é uma célula-tronco pluripotente. Em certas modalidades, a célula progenitora de eritrócitos é uma célula-tronco hematopoiética. Em certas modalidades, a célula progenitora de eritrócitos é célula-tronco hematopoiética CD34⁺.

[0019]Em outro aspecto, a presente divulgação provê um método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método compreendendo administrar ao indivíduo uma quantidade eficaz de um AAV defeituoso na replicação compreendendo: (a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F, e (b) um genoma de correção compreendendo (i) um elemento de edição para editar um *locus* alvo em um gene alvo; (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo

homologia para uma segunda região genômica 3' para o *locus* alvo, em que a célula é transduzida sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[0020]Em certas modalidades, a doença ou distúrbio é talassemia ou doença de célula falciforme. Em certas modalidades, o indivíduo é um indivíduo humano.

[0021]As seguintes modalidades se aplicam a cada um dos aspectos acima.

[0022]Em certas modalidades, o gene alvo é o gene HBB. Em certas modalidades, o *locus* alvo está em uma mutação de ponto de nucleotídeo ou deleção no gene HBB. Em certas modalidades, a mutação de ponto de nucleotídeo ou deleção no gene HBB é selecionada dentre o grupo consistindo em G em posição -87, G em posição -31, A em posição -30, G em posição -29, G em posição -28, T em posição -10, C em posição 1, A em posição 1, G em posição 2, deleção de C e T em posições 17 e 18, A em posição 19, deleção de A em posição 20, T em posição 20, deleção de A e A em posições 25 e 26, adição de G após posição 26, A em posição 47, A em posição 48, deleção de C em posição 51, A em posição 52, G em posição 58, G em posição 59, A em posição 79, T em posição 82, adição de C após posição 84, T em posição 93, A em posição 93, C em posição 97, C em posição 98, G em posição 202, G em posição 208, C em posição 222, deleção de T em posição 241 ou 242, deleção de T e T e C e T em posições 254 a 257, T em posição 260, deleção de C em posição 264 ou 265, adição de A após posição 343, deleção de G e T em posições 399 e 400, T em posição 401, adição de A após posição 417, A em posição 446, T em posição 1099, A

em posição 1293, T em posição 1344. Em certas modalidades, o elemento de edição compreende uma porção do gene HBB de tipo selvagem que corresponde à mutação.

[0023]Em certas modalidades, o elemento de edição compreende as regiões de codificação de um ou mais éxons de um gene HBB. Em certas modalidades, o elemento de edição consiste nas regiões de codificação de um ou mais éxons de um gene HBB.

[0024]Em certas modalidades, o elemento de edição compreende uma porção de um gene HBB compreendendo a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3. Em certas modalidades, as regiões de codificação foram alteradas silenciosamente para serem menos que 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idênticas aos éxons correspondentes do gene HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende pelo menos uma das sequências de nucleotídeos selecionadas dentre o grupo consistindo em SEQ ID NOs: 43-46 e 105-107.

[0025]Em certas modalidades, o *locus* alvo é AAVS1.

[0026]Em certas modalidades, o elemento de edição compreende uma sequência de codificação do gene HBB ou uma porção da mesma. Em certas modalidades, o elemento de edição compreende uma sequência de nucleotídeo codificando SEQ ID NO: 48. Em certas modalidades, a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste em nucleotídeos 4 a 444 de SEQ ID NO: 27. Em certas modalidades, a sequência de nucleotídeo codificando SEQ ID NO: 48 é alterada silenciosamente para ser menos que 70%,

75%, 80%, 85%, ou 90% idêntica a nucleotídeos 4 a 444 de SEQ ID NO: 27. Em certas modalidades, a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste na sequência de SEQ ID NO: 47 ou 100. Em certas modalidades, o elemento de edição compreende uma sequência de codificação inserida em um fragmento *stuffer* do gene HBB.

[0027]Em certas modalidades, o *locus* alvo é a ligação internucleotídeos entre nucleotídeo 3 e nucleotídeo 4 do gene alvo, pelo que a integração do elemento de edição no *locus* alvo resulta no *locus* alvo compreendendo uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* partindo com o códon de partida do gene alvo. Em certas modalidades, o elemento de edição compreende uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* consistindo em 5' a 3' um códon de partida e a sequência de nucleotídeo codificando SEQ ID NO: 48, ou uma porção da sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*. Em certas modalidades, o *locus* alvo está em um íntron do gene alvo, e em que o elemento de edição compreende 5' a 3' um sítio aceitador de emenda (*splicing*), um elemento de salto ribossômico, e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*. Em certas modalidades, o *locus* alvo está em íntron 1 do gene HBB. Em certas modalidades, o *locus* alvo está adjacientemente 3' para um nucleotídeo codificante do gene alvo, e em que o elemento de edição compreende 5' a 3' um elemento de salto ribossômico e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*. Em

certas modalidades, o *locus* alvo é o códon de parada de um gene alvo de tipo selvagem (por exemplo, gene HBB) ou os nucleotídeos correspondentes de um gene alvo mutante (por exemplo, gene HBB). Em certas modalidades, o *locus* alvo é a ligação internucleotídeos adjacientemente 5' para o códon de parada de um gene alvo de tipo selvagem (por exemplo, gene HBB) ou a ligação internucleotídeos correspondente de um gene alvo mutante (por exemplo, gene HBB).

[0028]Em certas modalidades, a sequência de nucleotídeo de braço de homologia 5' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à primeira região genômica. Em certas modalidades, a sequência de nucleotídeo de braço de homologia 3' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à segunda região genômica. Em certas modalidades, a primeira região genômica está localizada em uma primeira janela de edição, e a segunda região genômica está localizada em uma segunda janela de edição. Em certas modalidades, a primeira e segunda janelas de edição são diferentes. Em certas modalidades, a primeira e segunda janelas de edição são iguais. Em certas modalidades, a primeira janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103. Em certas modalidades, a segunda janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103. Em certas modalidades, a primeira região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101. Em certas modalidades, a segunda região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102. Em certas modalidades, o braço de homologia 5' consiste na sequência de nucleotídeo

especificada em SEQ ID NO: 101. Em certas modalidades, o braço de homologia 3' consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102.

[0029]Em certas modalidades, o elemento de edição compreende adicionalmente uma sequência 3' de poliadenilação exógena para a sequência de nucleotídeo codificando SEQ ID NO: 48. Em certas modalidades, o elemento de edição compreende adicionalmente um sítio de endonuclease de restrição não presente no gene alvo. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28.

[0030]Em certas modalidades, cada uma das sequências de nucleotídeos de braços de homologia 5' e 3' tem, independentemente, um comprimento de cerca de 100 a cerca de 2000 nucleotídeos.

[0031]Em certas modalidades, o genoma de correção compreende adicionalmente uma sequência de nucleotídeo 5' de repetição terminal invertida 5' (5' ITR) da sequência de nucleotídeo de braço de homologia 5', e uma sequência de nucleotídeo 3' de repetição terminal invertida 3' (3' ITR) da sequência de nucleotídeo de braço de homologia 3'. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:18, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:19. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:20, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:21. Em certas

modalidades, o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42 e 104.

[0032]Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO:2, opcionalmente em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

[0033]Em certas modalidades, (a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G; (b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; (c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; (d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou (e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

[0034]Em certas modalidades, a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17.

[0035]Em certas modalidades, a proteína de capsídeo de

AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO:2, opcionalmente em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o

aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

[0036]Em certas modalidades, (a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G; (b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; (c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; (d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou (e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

[0037]Em certas modalidades, a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17.

[0038]Em certas modalidades, a proteína de capsídeo de

AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO:2, opcionalmente em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 68 de SEQ ID NO: 2 é V; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo

correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

[0039]Em certas modalidades, (a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; (b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é Y; (c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; (d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; (e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G; (f) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; (g) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; (h) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou (i) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

[0040]Em certas modalidades, a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17.

[0041]Em certas modalidades, a eficiência de integração do elemento de edição no *locus* alvo é pelo menos 1% quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão. Em certas modalidades, a frequência alélica de integração do elemento de edição no *locus* alvo é pelo menos 0,5%, quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão.

[0042]Em outro aspecto, a presente divulgação provê uma composição farmacêutica compreendendo um AAV como divulgado aqui.

[0043]Em outro aspecto, a presente divulgação provê um sistema de empacotamento para preparação recombinante de um AAV, em que o sistema de empacotamento compreende (a) uma sequência de nucleotídeo Rep codificando uma ou mais proteínas Rep de AAV, (b) uma sequência de nucleotídeo Cap codificando uma ou mais proteínas de capsídeo de AAV Clade F como descrito aqui, e (c) um genoma de correção como divulgado aqui, em que o sistema de empacotamento é operativo em uma célula para encerrar o genoma de correção no capsídeo para formar o AAV.

[0044]Em certas modalidades, o sistema de empacotamento compreende um primeiro vetor compreendendo a sequência de nucleotídeo Rep e a sequência de nucleotídeo Cap, e um segundo vetor compreendendo o genoma de correção. Em certas modalidades, a sequência de nucleotídeo Rep codifica uma proteína Rep de AAV2. Em certas modalidades, a proteína Rep de AAV2 é 78/68 ou Rep 68/52. Em certas modalidades, a proteína Rep de AAV2 compreende uma sequência de aminoácido tendo uma identidade percentual mínima de sequência para a sequência de aminoácido Rep de AAV2 de SEQ ID NO:22, em que a identidade percentual mínima de sequência é pelo menos 70% através do comprimento da sequência de aminoácido codificando a proteína Rep de AAV2.

[0045]Em certas modalidades, o sistema de empacotamento compreende adicionalmente um terceiro vetor, em que o terceiro vetor é um vetor de vírus auxiliar. Em certas modalidades, o vetor de vírus auxiliar é um terceiro vetor

independente. Em certas modalidades, o vetor de vírus auxiliar é integrante com o primeiro vetor. Em certas modalidades, o vetor de vírus auxiliar é integrante com o segundo vetor. Em certas modalidades, o terceiro vetor compreende genes codificando proteínas de vírus auxiliar.

[0046]Em certas modalidades, o vírus auxiliar é selecionado dentre o grupo consistindo em adenovírus, vírus de herpes, vírus de vaccinia, e citomegalovírus (CMV). Em certas modalidades, o vírus auxiliar é adenovírus. Em certas modalidades, o genoma do adenovírus compreende um ou mais genes de RNA de adenovírus selecionados dentre o grupo consistindo em E1, E2, E4 e VA. Em certas modalidades, o vírus auxiliar é vírus herpes simplex (HSV). Em certas modalidades, o genoma de HSV compreende um ou mais de genes de HSV selecionados dentre o grupo consistindo em UL5/8/52, ICPO, ICP4, ICP22 e UL30/UL42.

[0047]Em certas modalidades, o primeiro vetor e o terceiro vetor estão contidos dentro de um primeiro plasmídeo transfectante. Em certas modalidades, os nucleotídeos do segundo vetor e do terceiro vetor estão contidos dentro de um segundo plasmídeo transfectante. Em certas modalidades, os nucleotídeos do primeiro vetor e do terceiro vetor são clonados em um vírus auxiliar recombinante. Em certas modalidades, os nucleotídeos do segundo vetor e do terceiro vetor são clonados em um vírus auxiliar recombinante.

[0048]Em outro aspecto, a presente divulgação provê um método para preparação recombinante de um AAV, o método compreendendo introduzir um sistema de empacotamento como descrito aqui em uma célula sob condições operativas para

encerrar o genoma de correção no capsídeo para formar o AAV.

BREVE DESCRIÇÃO DOS DESENHOS

[0049]As **Figuras 1A e 1B** são gráficos mostrando resultados de citometria de fluxo de integração de AAVS1-FP no genoma das células GM16265, em que o vetor AAVS1-FP foi empacotado em AAVHSC7, AAVHSC15, e AAVHSC17.

[0050]A **Figura 1C** é um gráfico mostrando a porcentagem de alelos tendo integração da sequência de codificação FP em células-tronco hematopoiéticas primárias humanas CD34⁺ (HSCs) transduzida com o vetor AAVS1-FP empacotado em capsídeo AAVHSC17.

[0051]A **Figura 2** mostra o mapa de plasmídeo do vetor de correção HBB hHBB-hL-014 contendo um ligante de 12 bp. Nesta figura, as regiões pretas de HBB denotam sequências de nucleotídeos nos éxons codificando uma proteína HBB humana, e as linhas tracejadas entre as regiões pretas denotam íntrons entre os éxons.

[0052]A **Figura 3A** é uma imagem de eletroforese de DNA mostrando tamanho específico de edição de DNA amplificado a partir do DNA genômico de células GM16265 transduzidas com o vetor hHBB-hL-014 empacotado em capsídeos AAVHSC15 e AAVHSC17.

[0053]A **Figura 3B** é uma imagem de eletroforese de DNA mostrando tamanho específico de edição de DNA amplificado a partir do DNA genômico de LCLs GM16265, GM16266, e GM16267 transduzidas com o vetor hHBB-hL-014 empacotado em capsídeo AAVHSC17.

[0054]As **Figuras 4A, 4B, 4C, e 4D** são mapas de vetores mostrando os elementos genéticos entre os dois ITRs de AAV

de vetores de correção de HBB hHBB-hL-001, hHBB-hLW-013, hHBB-hL-011, e hHBB-hLW-012, respectivamente. Nestas figuras, as regiões pretas de HBB denotam sequências de nucleotídeos nos éxons codificando uma proteína HBB humana, e as linhas tracejadas entre as regiões pretas denotam íntrons entre os éxons.

[0055]As **Figuras 5A, 5B, e 5C** são imagens de eletroforese de DNA mostrando tamanho específico de edição de DNA amplificado a partir do DNA genômico de HSCs CD34⁺ humanas primárias transduzidas com os vetores indicados nas figuras empacotadas em capsídeo AAVHSC17.

[0056]A **Figura 6** é um gráfico mostrando a fração de células CD34⁺ editadas através de amostras, como indicado.

[0057]As **Figuras 7A, 7B, 7C, 7D, e 7E** são mapas de vetores mostrando os elementos genéticos entre os dois ITRs de AAV de vetores de correção de HBB hHBB-hA-009, hHBB-hAW-002, hHBB-h1-010, hHBB-h1W-008, e hHBB-hE3C-001, respectivamente. Em Figuras 7A-D, as regiões pretas marcadas como "região de codificação de HBB" ou "região de codificação de HBB (66%)" denotam sequências de nucleotídeos codificando uma proteína HBB humana, ou a partir do códon de partida para o códon de parada (Figuras 7C e 7D), ou a partir do segundo códon para o códon de parada (Figuras 7 e 7B).

[0058]A **Figura 8** é uma imagem de eletroforese de DNA mostrando tamanho específico de edição (1,874 bp) e tamanho não específico (1,180 bp) de DNA amplificado a partir de DNA genômico de células LCL RKO e GM16265 transduzidas com os vetores indicados nas figuras empacotadas em capsídeo AAVHSC7.

[0059]A **Figura 9** é um gráfico mostrando a porcentagem de alelos tendo integração da sequência de codificação FP em células de sangue, medula óssea ("BM"), e de baço de camundongos NSG xenoenxertados com HSCs humanas após administração do vetor AAVS1-FP empacotado em capsídeos AAVHSC7 e AAVHSC17.

DESCRIÇÃO DETALHADA

[0060]A presente divulgação provê composições de vírus adeno-associados (AAV) para corrigir uma mutação em um gene HBB e métodos de usar as mesmas para corrigir uma mutação do gene HBB em uma célula. Também são providos sistemas de empacotamento para produzir composições de vírus adeno-associado.

I. Definições

[0061]Como usado aqui, o termo "vírus adeno-associado defeituoso na replicação" refere-se a um AAV compreendendo um genoma faltando genes Rep e Cap.

[0062]Como usado aqui, o termo "gene HBB" refere-se a um gene de beta globina humano mutante ou de tipo selvagem, incluindo, mas não limitado a, as regiões de codificação, éxons, introns, 5' UTR, 3' UTR, e regiões regulatórias transcricionais do gene HBB.

[0063]Como usado aqui, o termo "corrigir uma mutação em um gene HBB" refere-se à inserção, deleção, ou substituição de um ou mais nucleotídeos em um *locus* alvo em um gene alvo (por exemplo, um gene HBB mutante) para criar um *locus* que é capaz de expressar uma proteína HBB de tipo selvagem ou um equivalente funcional da mesma. Em certas modalidades, "corrigir uma mutação em um gene HBB" envolve reverter uma mutação em um gene HBB de volta à sequência de tipo

selvagem. Em certas modalidades, "corrigir uma mutação em um gene HBB" envolve inserir uma sequência de nucleotídeo codificando pelo menos uma porção de uma proteína beta globina de tipo selvagem ou um equivalente funcional da mesma no gene alvo (por exemplo, o gene HBB mutante), de modo que uma proteína beta globina de tipo selvagem ou um equivalente funcional da mesma é expressado do *locus* do gene alvo (por exemplo, o *locus* do gene HBB mutante), opcionalmente sob o controle de um promotor do gene alvo endógeno (por exemplo, promotor de gene HBB). Como usado aqui, um "equivalente funcional" refere-se a um produto de um gene ou fragmento do mesmo que pode funcionar como uma beta globina de tipo selvagem. Em certas modalidades, um equivalente funcional de HBB pode incluir outros genes ou pseudogenes de globina, tal como epsilon globina (HBE), delta globina (HBD), gama globina 1 (HBG1), gama globina 2 (HBG2), e pseudogene HBB HBBP. Em certas modalidades, um equivalente funcional de HBB pode ser uma proteína beta globina modificada, em que a modificação confere pelo menos uma característica não encontrada em beta globina tipo selvagem, por exemplo, a capacidade para inibir a agregação de beta globina carregando uma mutação SCD.

[0064]Como usado aqui, o termo "genoma de correção" refere-se a um genoma de AAV recombinante que é capaz de integrar um elemento de edição (por exemplo, um ou mais nucleotídeos ou uma ligação internucleotídeos) através de recombinação homóloga em um *locus* alvo para corrigir um defeito genético em um gene HBB. Em certas modalidades, o *locus* alvo é no gene HBB humano. O técnico no assunto irá reconhecer que a porção de um genoma de correção

compreendendo o braço de homologia 5', elemento de edição, e braço de homologia 3' pode estar na orientação senso ou antissenso relativa para o *locus* alvo (por exemplo, o gene HBB humano).

[0065] Como usado aqui, o termo "elemento de edição" refere-se à porção de um genoma de correção que quando integrado em um *locus* alvo modifica o *locus* alvo. Um elemento de edição pode mediar inserção, deleção ou substituição de um ou mais nucleotídeos no *locus* alvo.

[0066] Como usado aqui, o termo "*locus* alvo" refere-se a uma região de um cromossomo ou uma ligação internucleotídeos que é modificada por um elemento de edição. Em certas modalidades, o *locus* alvo é uma região ou uma ligação internucleotídeos no gene HBB, opcionalmente em que o *locus* alvo compreende pelo menos uma mutação genética que compromete a expressão ou função de proteína beta globina. Em certas modalidades, o *locus* alvo é AAVS1. O *locus* AAVS1 está no cromossomo 19 qter13.3-13.4, nucleotídeos 55.090.913 a 55.117.600 de NCBI Sequência de Referência No. NC_000019.10, como descrito em Giraud et al., Proc Natl Acad Sci U S A. (1994) 91(21):10039-43; Linden et al., Proc Natl Acad Sci U S A. (1996) 93 (21): 11288-94; e Linden et al., Proc Natl Acad Sci U S A. (1996) 93(15):7966-72, cada um sendo aqui incorporado por referência em sua totalidade. Em certas modalidades, o *locus* alvo é um *locus* de abrigo seguro. Os *loci* de abrigo seguro são sítios no genoma capazes de acomodar a integração de novo material genético em um modo que assegura que o material genético recentemente inserido: (1) funcione de modo previsível; e (2) não cause alterações do

genoma do hospedeiro que possam colocar um risco para a célula ou organismo hospedeiro. Conseqüentemente, em certas modalidades, o *locus* alvo pode ser qualquer *locus* de abrigo seguro conhecido na técnica possa suportar expressão de transgene previsível, enquanto minimizando o risco de interações indesejadas com o genoma do hospedeiro.

[0067]Como usado aqui, o termo "gene alvo" refere-se a um gene em que um *locus* alvo ou uma porção do mesmo está localizado. Em certas modalidades, o *locus* alvo está totalmente no gene alvo. Em certas modalidades, o gene alvo é HBB. Em certas modalidades, o gene alvo é PPP1R12C humano. Em certas modalidades, o gene alvo é expressado em um progenitor de eritrócitos.

[0068]Como usado aqui, o termo "braço de homologia" refere-se a uma porção de um genoma de correção posicionado 5' ou 3' de um elemento de edição que é substancialmente idêntico ao genoma flanqueando um *locus* alvo. Em certas modalidades, o *locus* alvo está em um gene HBB humano, e o braço de homologia compreende uma sequência substancialmente idêntica ao genoma flanqueando *locus* alvo.

[0069]Como usado aqui, o termo "proteína de capsídeo Clade F" refere-se a uma proteína de capsídeo de AAV VP1, VP2, ou VP3 que compreende uma sequência de aminoácido tendo pelo menos 90% de identidade com a sequência de aminoácidos VP1, VP2, ou VP3 especificada, respectivamente, em aminoácidos 1-736, 138-736, e 203-736 de SEQ ID NO:1 aqui. Como usado aqui, a identidade entre duas sequências de nucleotídeos ou entre duas sequências de aminoácidos é determinada pelo número de nucleotídeos ou aminoácidos idênticos em alinhamento dividido pelo comprimento completo

da sequência de nucleotídeo ou de aminoácido mais longa.

[0070]Como usado aqui, o termo "uma doença ou distúrbio associado com uma mutação do gene HBB" refere-se a qualquer doença ou distúrbio causada por, exacerbada por, ou geneticamente ligada com variação de um gene HBB. Em certas modalidades, a doença ou distúrbio associado com uma mutação do gene HBB é uma hemoglobinopatia, tal como doença de célula falciforme ou beta talassemia.

[0071]Como usado aqui, o termo "alterada silenciosamente" refere-se à alteração de uma sequência de codificação ou uma sequência de codificação inserida em um fragmento *stuffer* de um gene (por exemplo, por substituição de nucleotídeo) sem mudar a sequência de aminoácido do polipeptídeo codificado pela sequência de codificação ou sequência de codificação inserida em fragmento *stuffer*. Esta alteração silenciosa é vantajosa em que ela reduz a possibilidade de integração do genoma de correção em *loci* de outros genes ou pseudogenes parálogos ao gene alvo (por exemplo, outro *locus* de gene globina ou um *locus* de pseudogene de beta globina). Tal alteração silente também reduz a homologia entre o elemento de edição e o gene alvo, assim reduzindo integração indesejada mediada pelo elemento de edição em vez de por um braço de homologia.

[0072]Como usado aqui, o termo "sequência de codificação" refere-se a uma porção de um DNA complementar (cDNA), ou uma sequência alterada silenciosamente do mesmo, que codifica um polipeptídeo, partindo em um códon de partida e terminando em um códon de parada. Um gene pode ter uma ou mais sequências de codificação de tipo selvagem devido à emenda alternativa e/ou iniciação de tradução

alternativa. Uma sequência de codificação de tipo selvagem de HBB exemplar é especificada em nucleotídeos 51-494 da Sequência de Referência NCBI: NM_000518.4.

[0073]Como usado aqui, o termo "nucleotídeo codificante" refere-se a um nucleotídeo de um gene que corresponde a um nucleotídeo em uma sequência de codificação do gene, exceto nucleotídeo 3' do códon de parada. Conseqüentemente, em certas modalidades, um nucleotídeo codificante do gene HBB é qualquer um dos nucleotídeos 1-443 do gene HBB.

[0074]Como usado aqui, o termo "sequência de codificação inserida em fragmento *stuffer*" de um gene refere-se a uma sequência de nucleotídeo compreendendo um ou mais íntrons inseridos em uma sequência de codificação do gene. Em certas modalidades, pelo menos um dos íntrons é um íntron não nativo, isto é, tendo uma sequência diferente de um íntron nativo do gene. Em certas modalidades, todos os íntrons na sequência de codificação inserida em fragmento *stuffer* são íntrons não nativos. Um íntron não nativo pode ter a sequência de um íntron de uma diferente espécie ou a sequência de um íntron em um gene diferente de espécies iguais. Alternativamente ou adicionalmente, pelo menos uma porção de uma sequência de íntron não nativo pode ser sintética. Um técnico no assunto irá reconhecer que as sequências de íntrons não nativos podem ser planejadas para mediar a emenda de RNA por introdução de quaisquer motivos de emenda de consenso conhecidos na técnica. Os motivos de emenda de consenso exemplares são providos em Sibley et al., (2016) Nature Reviews Genetics, 17, 407-21, que é incorporado por referência aqui em sua totalidade. Inserção

de um íntron não nativo pode promover a eficiência e robustez de empacotamento do vetor, como sequências de fragmento *stuffer* permitem ajustes do vetor para alcançar um tamanho ótimo (por exemplo, 4,5-4,8 kb). Em certas modalidades, pelo menos um dos íntrons é um íntron nativo do gene. Em certas modalidades, todos os íntrons na sequência de codificação inserida em fragmento *stuffer* são íntrons nativos do gene. Os íntrons não nativos ou nativos podem ser inseridos em qualquer ligação internucleotídeos na sequência de codificação. Em certas modalidades, um ou mais íntrons não nativos ou nativos são inseridos na ligações internucleotídeos previstas para promover emenda eficaz (ver, por exemplo, Zhang (1998) Human Molecular Genetics, 7(5):919-32, que é incorporado por referência aqui em sua totalidade). Em certas modalidades, um ou mais íntrons não nativos ou nativos são inseridos nas ligações internucleotídeos que ligam dois éxons endógenos.

[0075]Como usado aqui, o termo "elemento de salto ribossômico" refere-se a uma sequência de nucleotídeos codificando uma sequência de peptídeo curta capaz de causar geração de duas cadeias de peptídeo a partir da tradução de uma molécula de mRNA. Como usado aqui, o termo "peptídeo de salto ribossômico" refere-se a um peptídeo codificado por um elemento de salto ribossômico. Em certas modalidades, o peptídeo de salto ribossômico compreende um motivo de consenso de $X_1X_2EX_3NPGP$, em que X_1 é D ou G, X_2 é V ou I, e X_3 é qualquer aminoácido (SEQ ID NO: 49). Em certas modalidades, o peptídeo de salto ribossômico é selecionado dentre o grupo consistindo em peptídeo, vírus thossea-asigna 2A (T2A), peptídeo, teschovirus-1 porcino 2A (P2A),

peptídeo, vírus doença de febre aftosa 2A (F2A), peptídeo, vírus da rinite equina 2A (E2A), peptídeo, vírus polihedrose citoplásmica 2A (BmCPV 2A), e peptídeo, vírus flacherie de *B. mori* 2A (BmIFV 2A). Exemplos de sequências de aminoácidos de peptídeo T2A e peptídeo P2A são especificadas em SEQ ID NOs: 71 e 73, respectivamente. Sequências de nucleotídeos exemplares de elemento T2A e elemento P2A são especificadas em SEQ ID NOs: 72 e 74, respectivamente. Em certas modalidades, o elemento de salto ribossômico codifica um peptídeo que compreende adicionalmente uma sequência de Gly-Ser-Gly no término N, opcionalmente em que a sequência de Gly-Ser-Gly é codificada pela sequência de nucleotídeo de GGCAGCGGA (SEQ ID NO: 75). Embora não desejando ser limitado por teoria, é formulada a hipótese de que os elementos de salto ribossômico funcionam por: terminar a tradução da primeira cadeia de peptídeo e re-iniciar a tradução da segunda cadeia de peptídeo; ou por clivagem de uma ligação de peptídeo no peptídeo de salto ribossômico por uma atividade de protease intrínseca do peptídeo codificado ou por outra protease no ambiente (por exemplo, citossol).

[0076] Como usado aqui, o termo "sequência de poliadenilação" refere-se a uma sequência de DNA que quando transcrita em RNA constitui uma sequência de sinal de poliadenilação.

[0077] Na presente divulgação, posições de nucleotídeo em um gene são especificadas relativas ao primeiro nucleotídeo do códon de partida. O primeiro nucleotídeo de um códon de partida é posição 1, nucleotídeos 5' para o primeiro nucleotídeo de um códon de partida tem números

negativos, e os nucleotídeos 3' para o primeiro nucleotídeo de um códon de partida tem números positivos. Por exemplo, nucleotídeo 1 do gene HBB como usado aqui é nucleotídeo 70.595 da Sequência de Referência NCBI: NG_000007.3. O nucleotídeo adjacientemente 5' para o códon de partida é nucleotídeo -1.

[0078]Na presente divulgação, éxons e íntrons em um gene são especificados relativos ao éxon englobando o primeiro nucleotídeo do códon de partida. O éxon englobando o primeiro nucleotídeo do códon de partida é éxon 1. Exons 3' para éxon 1 são de 5' a 3': éxon 2, éxon 3, etc. Íntrons 3' para éxon 1 são de 5' a 3': íntron 1, íntron 2, etc. Conseqüentemente, um gene compreende de 5' a 3': éxon 1, íntron 1, éxon 2, íntron 2, éxon 3, etc. Um éxon 1 exemplar do gene HBB humano é de nucleotídeos 70.545 a 70.686 da Sequência de Referência NCBI: NG_000007.3. Um íntron 1 exemplar do gene HBB humano é de nucleotídeos 70.687 a 70.816 da Sequência de Referência NCBI: NG_000007.3. Um técnico no assunto irá reconhecer que um gene pode ser transcrito em múltiplos mRNAs diferentes. Como tal, um gene (por exemplo, HBB) pode ter múltiplos conjuntos diferentes de éxons e íntrons.

[0079]Como usado aqui, o termo "integração" refere-se à introdução de um elemento de edição em um *locus* alvo por recombinação homóloga entre um genoma de correção e o gene alvo. Integração de um elemento de edição pode resultar em substituição, inserção e/ou deleção de um ou mais nucleotídeos no gene alvo.

[0080]Como usado aqui, o termo "eficiência de integração do elemento de edição no *locus* alvo" refere-se à

porcentagem de células em uma população transduzida em que integração do elemento de edição no *locus* alvo tinha ocorrido.

[0081]Como usado aqui, o termo "frequência alélica de integração do elemento de edição no *locus* alvo" refere-se à porcentagem de alelos em uma população de células transduzidas em que integração do elemento de edição no *locus* alvo tinha ocorrido.

[0082]Como usado aqui, o termo "condições de transdução de AAV padrão" refere-se à transdução de 2×10^5 células-tronco humanas CD34+ com um AAV em um multiplicidade de infecção (MOI) de $1,5 \times 10^5$, em que as células são cultivadas em meio de Dulbecco modificado por Iscove (IMDM) suplementado com soro de bezerro fetal a 20% (FCS), 100 µg/mL estreptomicina, 100 U/mL penicilina, 2mmol/L L-glutamina, 10 ng/mL IL-3 humana, 10 ng/mL IL-6 humana, e 1 ng/mL SCF humano, a 37°C em um meio de incubador de dióxido de carbono a 5% (CO₂), em que o AAV é formulado em solução salina tamponada com fosfato (PBS), e em que o AAV é adicionado ao meio de cultura de células contendo as células CD34+ em um volume que é menor do que ou igual a 1/9 do volume do meio de cultura.

[0083]Como usado aqui, o termo "quantidade eficaz" no contexto da administração de um AAV para um indivíduo refere-se à quantidade do AAV que obtém um efeito profilático ou terapêutico desejado.

[0084]Como usado aqui, o termo "um progenitor de eritrócitos" refere-se a uma célula capaz de se diferenciar em um eritrócito. Em certas modalidades, a progenitor de eritrócitos é uma célula-tronco pluripotente. Em certas

modalidades, o progenitor de eritrócitos é uma célula tronco pluripotente induzida. Em certas modalidades, o progenitor de eritrócitos é uma célula-tronco hematopoiética. Em certas modalidades, o progenitor de eritrócitos é uma célula-tronco hematopoiética CD34⁺. Em certas modalidades, o progenitor de eritrócitos é uma célula de progenitor de mieloides. Em certas modalidades, o progenitor de eritrócitos é uma célula de progenitor de megacariócitos-eritroides. Em certas modalidades, o progenitor de eritrócitos é uma célula de precursor de eritroides.

II. Composições de vírus adeno-associados

[0085]Em um aspecto, a divulgação provê novas composições de AAVs defeituosos na replicação utilizáveis para corrigir mutações em um gene HBB. Os AAVs divulgados aqui geralmente compreendem: um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e um genoma de correção para edição de um *locus* alvo em um gene HBB.

[0086]Qualquer proteína de capsídeo de AAV Clade F ou derivado da mesma pode ser usada nas composições de AAV divulgadas aqui. Por exemplo, em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 80%, 85%, 90%, 95%, ou 99% de identidade de sequência com a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO:2. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 80%, 85%, 90%, 95%, ou 99% de identidade de sequência com a sequência de aminoácido de aminoácidos 203-736 de SEQ ID

NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na proteína de capsídeo

correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17.

[0087] Por exemplo, em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 80%, 85%, 90%, 95%, ou 99% de identidade de sequência com a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 80%, 85%, 90%, 95%, ou 99% de identidade de sequência com a sequência de

aminoácido de aminoácidos 138-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na

proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17.

[0088] Por exemplo, em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 90%, 95%, ou 99% de identidade de sequência com a sequência de aminoácido de aminoácidos

1-736 de SEQ ID NO:2. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 90%, 95%, ou 99% de identidade de sequência com a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 68 de SEQ ID NO: 2 é V; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de

SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é Y. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R. Em certas modalidades, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C. Em certas modalidades, a proteína de capsídeo de AAV Clade F compreende a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17.

[0089]Em certas modalidades, o capsídeo de AAV compreende duas ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17; e (c) uma proteína de capsídeo Clade F

compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17. Em certas modalidades, o capsídeo de AAV compreende: (a) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 203-736 de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17; (b) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17; e (c) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17.

[0090]Em certas modalidades, o capsídeo de AAV compreende um ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 8; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 8; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 8. Em certas modalidades, o capsídeo de AAV compreende duas ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 8; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 8; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 8. Em certas modalidades, o capsídeo de AAV compreende: (a) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 203-736 de SEQ ID NO:

8; (b) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 138-736 de SEQ ID NO: 8; e (c) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 1-736 de SEQ ID NO: 8.

[0091]Em certas modalidades, o capsídeo de AAV compreende um ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 13; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 13; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 13. Em certas modalidades, o capsídeo de AAV compreende duas ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 13; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 13; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 13. Em certas modalidades, o capsídeo de AAV compreende: (a) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 203-736 de SEQ ID NO: 13; (b) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 138-736 de SEQ ID NO: 13; e (c) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 1-736 de SEQ ID NO: 13.

[0092]Em certas modalidades, o capsídeo de AAV compreende um ou mais de: (a) uma proteína de capsídeo

Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 16; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 16; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 16. Em certas modalidades, o capsídeo de AAV compreende duas ou mais de: (a) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 16; (b) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 16; e (c) uma proteína de capsídeo Clade F compreendendo a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 16. Em certas modalidades, o capsídeo de AAV compreende: (a) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 203-736 de SEQ ID NO: 16; (b) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 138-736 de SEQ ID NO: 16; e (c) uma proteína de capsídeo Clade F tendo uma sequência de aminoácido consistindo em aminoácidos 1-736 de SEQ ID NO: 16.

[0093] Genomas de correção utilizáveis nas composições de AAV divulgadas aqui geralmente compreendem: (i) um elemento de edição para editar um *locus* alvo em um gene alvo, (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo, e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo. Em certas modalidades, o

genoma de correção compreende uma sequência de nucleotídeo 5' de repetição terminal invertida 5' (5' ITR) da sequência de nucleotídeo de braço de homologia 5', e uma sequência de nucleotídeo 3' de repetição terminal invertida 3' (3' ITR) da sequência de nucleotídeo de braço de homologia 3'.

[0094]Elementos de edição usados nos genomas de correção divulgados aqui podem mediar inserção, deleção, ou substituição de um ou mais nucleotídeos no *locus* alvo. O *locus* alvo pode se localizar completamente ou parcialmente em um gene alvo, que pode ser o gene HBB ou outro gene expressado em um progenitor de eritrócitos.

[0095]Em certas modalidades, quando corretamente integrado por recombinação homóloga no *locus* alvo, o elemento de edição corrige uma mutação em um gene HBB de volta para a sequência HBB de tipo selvagem ou para uma sequência silenciosamente alterada que codifica a proteína HBB de tipo selvagem ou um equivalente funcional da mesma. A maior parte das mutações no gene HBB pode ser corrigida por um elemento de edição como divulgado aqui. Em certas modalidades, o elemento de edição é um ou mais nucleotídeos que corrigem uma mutação de substituição ou deleção no gene HBB. Em certas modalidades, o elemento de edição é uma ligação internucleotídeos que deleta uma mutação de inserção no gene HBB. Em certas modalidades, o elemento de edição compreende um ou mais éxons codificantes de um gene HBB. Por exemplo, o elemento de edição pode compreender uma porção de um gene HBB que engloba a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3. Os éxons podem ser de tipo selvagem ou alterados silenciosamente

como divulgado aqui.

[0096]Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação ou sequência de codificação inserida em fragmento *stuffer* de um gene HBB. Em certas modalidades, o elemento de edição compreende toda ou substancialmente toda de uma sequência de codificação ou sequência de codificação inserida em fragmento *stuffer* de um gene HBB. Por exemplo, em certas modalidades, o elemento de edição compreende nucleotídeos 4 a 444 de uma sequência de codificação de HBB, ou uma porção de uma sequência de codificação de HBB inserida em fragmento *stuffer* de nucleotídeo 4 para o códon de parada, opcionalmente compreendendo adicionalmente uma sequência 3' de poliadenilação exógena para a porção de sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*. Em certas modalidades, o elemento de edição compreende uma sequência de nucleotídeo codificando SEQ ID NO: 48, e pode opcionalmente compreender adicionalmente uma sequência 3' de poliadenilação exógena para a sequência de nucleotídeo codificando SEQ ID NO: 48. Em certas modalidades, tais elementos de edição podem ser integrados no éxon 1 imediatamente após o códon de partida endógeno do gene HBB (por exemplo, entre nucleotídeo 3 e nucleotídeo 4 do gene HBB) por recombinação homóloga, pelo qual a integração do elemento de edição resulta na geração de uma sequência de codificação de HBB completa em quadro com o códon de partida do gene HBB endógeno. Em certas modalidades, tais elementos de edição podem ser integrados no éxon 1 imediatamente após o códon de partida endógeno de um gene alvo não HBB (por exemplo, entre nucleotídeo 3 e

nucleotídeo 4 do gene alvo) por recombinação homóloga, pelo que a integração do elemento de edição resulta na geração de uma sequência completa de codificação de HBB em quadro com o códon de partida do gene alvo endógeno. A porção da sequência de codificação de HBB no elemento de edição pode ser tipo selvagem ou silenciosamente mudada como divulgado aqui. A porção de sequência de aminoácido de HBB codificada pelo elemento de edição pode ser tipo selvagem ou um equivalente funcional do mesmo.

[0097]Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* (por exemplo, uma sequência de codificação de HBB completa ou sequência de codificação inserida em fragmento *stuffer*), e um elemento de salto ribossômico ou uma sequência de poliadenilação exógena. Em certas modalidades, o elemento de edição compreende 5' a 3' um elemento de salto ribossômico, e pelo menos uma porção de uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* (por exemplo, uma sequência de codificação de HBB completa ou sequência de codificação inserida em fragmento *stuffer*). Em certas modalidades, o elemento de edição compreende 5' a 3': um elemento de salto ribossômico; uma sequência de codificação de HBB completa ou sequência de codificação inserida em fragmento *stuffer*; e uma sequência de poliadenilação exógena. Em certas modalidades, o elemento de edição acima mencionado pode ser integrado adjacientemente 3' para um nucleotídeo codificante de um gene alvo (por exemplo, adjacientemente 5' para o códon de parada do gene HBB

nativo) por recombinação homóloga para produzir um gene alvo recombinante compreendendo 5' a 3': uma porção 5' do gene alvo terminando no nucleotídeo codificante, um elemento de salto ribossômico, uma sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*, e uma sequência de poliadenilação exógena, em que o elemento de salto ribossômico é posicionado de modo que ele está em quadro com a região de codificação do gene alvo e a sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*. Expressão deste gene alvo recombinante produz um primeiro polipeptídeo compreendendo a sequência de aminoácido codificada por uma porção N-terminal do gene alvo fusionada a uma primeira porção do peptídeo de salto ribossômico codificado, e um segundo polipeptídeo compreendendo uma segunda porção do peptídeo de salto ribossômico codificado (por exemplo, um resíduo único de prolina) fusionado à sequência de aminoácido de HBB completa. A sequência de codificação de HBB ou as regiões de codificação da sequência de codificação de HBB inserida em fragmento *stuffer* no elemento de edição pode ser tipo selvagem ou silenciosamente mudada, como divulgado aqui. A sequência de aminoácido de HBB codificada pelo elemento de edição pode ser tipo selvagem ou um equivalente funcional da mesma (por exemplo, faltando a primeira metionina). O *locus* alvo pode ser uma ligação internucleotídeos ou uma sequência de nucleotídeo adjacente 3' para um nucleotídeo codificante do gene alvo. Em certas modalidades, o *locus* alvo consiste no códon de parada nativo do gene HBB.

[0098]Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* (por exemplo, uma sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*), e um ou mais de: um sítio aceitador de emenda; um sítio doador de emenda; um elemento de salto ribossômico; e uma sequência de poliadenilação exógena. Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; e pelo menos uma porção de uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* (por exemplo, uma sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*). Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; uma sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*; e uma sequência de poliadenilação exógena. Em certas modalidades, o elemento de edição acima mencionado pode ser integrado em um íntron de um gene alvo (por exemplo, íntron 1 do gene HBB endógeno) por recombinação homóloga para produzir um gene HBB recombinante compreendendo 5' a 3': o um ou mais éxons 5' para o íntron do gene alvo; um a porção 5' do íntron incluindo o sítio doador de emenda endógeno; um sítio aceitador de emenda; um elemento de salto ribossômico, uma sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer*; e uma sequência de poliadenilação exógena, em que o

elemento de salto ribossômico é posicionado de modo que ele está em quadro com a sequência de codificação completa de HBB ou sequência de codificação inserida em fragmento *stuffer* e de modo que a emenda do sítio aceitador de emenda para o sítio doador de emenda endógeno do gene alvo ocorra em quadro com a região de codificação do gene alvo. Expressão deste gene alvo recombinante produz um primeiro polipeptídeo compreendendo o sequência de aminoácido de gene alvo codificada pelo(s) éxon(s) 5' endógeno(s) para o sítio de inserção fusionado a uma primeira porção do peptídeo de salto ribossômico codificado, e um segundo polipeptídeo compreendendo uma segunda porção do peptídeo de salto ribossômico codificado (por exemplo, um resíduo único de prolina) fusionado a uma sequência de aminoácido HBB completa. A sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* no elemento de edição pode ser tipo selvagem ou silenciosamente mudada como divulgado aqui. A sequência de aminoácido HBB codificada pelo elemento de edição pode ser tipo selvagem ou um equivalente funcional da mesma (por exemplo, faltando a primeira metionina). O *locus* alvo pode ser uma ligação internucleotídeos ou uma sequência de nucleotídeo adjacientemente 3' para um nucleotídeo em um intron do gene alvo.

[0099]Em certas modalidades, uma ou mais porções de uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* dentro de um elemento de edição pode ser alterada silenciosamente para ser não idêntica com os éxons correspondentes do gene HBB de tipo selvagem. Tal alteração silenciosa é vantajosa em que ela

reduz a possibilidade de integração do genoma de correção em *loci* de outros genes ou pseudogenes de globina, por exemplo, um *locus* de pseudogene de beta globina. Tal alteração silenciosa também reduz a homologia entre o elemento de edição e o gene alvo, assim reduzindo integração indesejada mediada pelo elemento de edição em vez de por um braço de homologia.

[00100]Consequentemente, em certas modalidades, o elemento de edição compreende regiões de codificação de um ou mais éxons de um gene HBB que foram alteradas silenciosamente para serem menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idênticas às correspondentes regiões de codificação de éxons do gene HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende regiões de codificação de um ou mais éxons de um gene HBB que foram alteradas silenciosamente para serem menos que 70% idênticas às correspondentes regiões de codificação de éxons do gene HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende regiões de codificação de um ou mais éxons de um gene HBB que foram alteradas silenciosamente para serem menos que 85% idênticas às correspondentes regiões de codificação de éxons do gene HBB de tipo selvagem.

[00101]Em certas modalidades, o elemento de edição compreende uma porção de um gene HBB englobando a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3, em que uma ou mais das regiões de codificação de éxon 1, o éxon completo 2, e a região de codificação de éxon 3

foram alteradas silenciosamente para serem menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idênticas às correspondentes regiões de éxons do gene HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende uma porção de um gene HBB englobando a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3, em que um ou mais dentre regiões de codificação de éxon 1, o éxon completo 2, e a região de codificação de éxon 3 foram alterados silenciosamente para serem menos que 70% idênticos às correspondentes regiões de éxons do gene HBB de tipo selvagem.

[00102]Em certas modalidades, o elemento de edição compreende uma porção de um gene HBB englobando a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3, em que cada um de região de codificação de éxon 1, o éxon completo 2, e a região de codificação de éxon 3 foi alterado silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntico às regiões correspondentes de éxons do gene HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende uma porção de um gene HBB englobando a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3, em que cada um dentre a região de codificação de éxon 1, o éxon completo 2, e a região de codificação de éxon 3 foi alterado silenciosamente para ser menos que 70% idêntico às regiões

correspondentes de éxons do gene HBB de tipo selvagem.

[00103]Em certas modalidades, o elemento de edição compreende uma ou mais dentre as sequências de nucleotídeos selecionadas dentre o grupo consistindo em SEQ ID NOs: 43-46 e 105-107. Em certas modalidades, o elemento de edição compreende duas ou mais das sequências de nucleotídeos selecionadas dentre o grupo consistindo em SEQ ID NOs: 43-46 e 105-107. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em SEQ ID NO: 46. Em certas modalidades, o elemento de edição compreende as sequências de nucleotídeos especificadas em SEQ ID NOs: 43, 44, e 45. Em certas modalidades, o elemento de edição compreende as sequências de nucleotídeos especificadas em SEQ ID NOs: 105, 106, e 107.

[00104]Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 70% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 85% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem.

[00105]Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a pelo menos uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação de menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a pelo menos uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação menos que 70% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende pelo menos uma porção de uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a pelo menos uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação menos que 85% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem.

[00106]Em certas modalidades, o elemento de edição compreende nucleotídeos 4 a 444 de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%,

80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende nucleotídeos 4 a 444 de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 70% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em SEQ ID NO: 47. Em certas modalidades, o elemento de edição compreende nucleotídeos 4 a 444 de uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 85% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em SEQ ID NO: 100. Tais elementos de edição podem compreender adicionalmente uma sequência 3' de poliadenilação exógena para a sequência de codificação de gene HBB. Em certas modalidades, o elemento de edição compreende 5' a 3': nucleotídeos 4 a 444 de uma sequência de codificação de HBB; e uma sequência de poliadenilação exógena, em que os nucleotídeos 4 a 444 de uma sequência de codificação de HBB foram alterados silenciosamente para serem menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idênticos à região da sequência de codificação de HBB de tipo selvagem correspondente.

[00107] Em certas modalidades, o elemento de edição compreende uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a porção de sequência de

codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em nucleotídeos 4 a 444 de uma sequência de codificação de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em nucleotídeos 4 a 444 de uma sequência de codificação de HBB que é menos que 70% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, a porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada na sequência de nucleotídeo especificada em SEQ ID NO: 47. Em certas modalidades, o elemento de edição compreende uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em nucleotídeos 4 a 444 de uma sequência de codificação de HBB que é menos que 85% idêntica à porção correspondente da sequência de codificação de HBB de tipo selvagem. Em certas modalidades, a porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada na sequência de nucleotídeo especificada em SEQ ID NO: 100. Tais elementos de edição podem compreender adicionalmente

uma sequência 3' de poliadenilação exógena para a sequência de codificação de gene HBB. Em certas modalidades, o elemento de edição compreende 5' a 3': uma porção de sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente; e uma sequência de poliadenilação exógena, em que a porção de sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em nucleotídeos 4 a 444 de uma sequência de codificação de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente.

[00108]Em certas modalidades, o elemento de edição compreende uma sequência de codificação completa de HBB que foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende uma sequência de codificação completa de HBB que foi alterada silenciosamente para ser menos que 70% idêntica à sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende 5' a 3' um códon de partida e a sequência de nucleotídeo especificada em SEQ ID NO: 47. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em SEQ ID NO: 28. Em certas modalidades, o elemento de edição compreende uma sequência de codificação completa de HBB de um gene HBB que foi alterada silenciosamente para ser menos que 85% idêntica à sequência de codificação de HBB de tipo selvagem. Em certas

modalidades, o elemento de edição compreende 5' a 3' um códon de partida e a sequência de nucleotídeo especificada em SEQ ID NO: 100. Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em SEQ ID NO: 99. Tais elementos de edição podem compreender adicionalmente um ou mais de: um sítio aceitador de emenda; um elemento de salto ribossômico; e uma sequência 3' de poliadenilação exógena para a sequência de codificação de gene HBB. Em certas modalidades, o elemento de edição compreende 5' a 3': um elemento de salto ribossômico; e uma sequência de codificação completa de HBB opcionalmente faltando o códon de partida, em que a sequência de codificação completa de HBB foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um elemento de salto ribossômico; uma sequência de codificação completa de HBB opcionalmente faltando o códon de partida; e uma sequência de poliadenilação exógena, em que a sequência de codificação completa de HBB foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; e uma sequência de codificação completa de HBB opcionalmente faltando o códon de partida, em que a sequência de codificação completa de

HBB foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; uma sequência de codificação completa de HBB opcionalmente faltando o códon de partida; e uma sequência de poliadenilação exógena, em que a sequência de codificação completa de HBB foi alterada silenciosamente para ser menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente.

[00109]Em certas modalidades, o elemento de edição compreende uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à sequência de codificação de HBB de tipo selvagem. Em certas modalidades, o elemento de edição compreende uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 70% idêntica à sequência de codificação de HBB de tipo selvagem. Em certas modalidades, a sequência de codificação completa de HBB

consiste em 5' a 3': um códon de partida e a sequência de nucleotídeo especificada em SEQ ID NO: 47. Em certas modalidades, a sequência de codificação completa de HBB consiste na sequência de nucleotídeo especificada em SEQ ID NO: 28. Em certas modalidades, a sequência de codificação de HBB inserida em fragmento *stuffer* compreende as sequências de nucleotídeos especificadas em SEQ ID NOs: 43-45. Em certas modalidades, o elemento de edição compreende uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, de modo que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 85% idêntica à sequência de codificação de HBB de tipo selvagem. Em certas modalidades, a sequência de codificação de HBB inserida em fragmento *stuffer* compreende 5' a 3': as sequências de nucleotídeos especificada em SEQ ID NOs: 105, 106, e 107. Tais elementos de edição podem compreender adicionalmente um ou mais de: um sítio aceitador de emenda; um elemento de salto ribossômico; e uma sequência 3' de poliadenilação exógena para a sequência de codificação de gene HBB. Em certas modalidades, o elemento de edição compreende 5' a 3': um elemento de salto ribossômico; e uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, em que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem

correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um elemento de salto ribossômico; uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente; e uma sequência de poliadenilação exógena, em que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; e uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente, em que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%) idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente. Em certas modalidades, o elemento de edição compreende 5' a 3': um sítio aceitador de emenda; um elemento de salto ribossômico; uma sequência de codificação de HBB inserida em fragmento *stuffer* que foi alterada silenciosamente; e uma sequência de poliadenilação exógena, em que a sequência de codificação de HBB inserida em fragmento *stuffer* pode ser transcrita e emendada em uma sequência de codificação completa de HBB que é menos que 100% (por exemplo, menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50%)

idêntica à região da sequência de codificação de HBB de tipo selvagem correspondente.

[00110]Qualquer um e todos dos elementos de edição divulgados aqui podem incluir adicionalmente uma sequência única não presente no gene alvo (por exemplo, gene HBB mutante ou tipo selvagem, ou um gene codificando um equivalente funcional do mesmo), assim permitindo a identificação de células que têm integração do elemento de edição no *locus* alvo. Tal sequência única pode ser uma sequência apropriada para análise de sequenciamento de ácido nucleico (por exemplo, por PCR ou sequenciamento de próxima geração) do *locus* alvo e suas regiões de flanqueando ou um ácido nucleico amplificado a partir daí. Tal sequência única também pode ser um sítio de endonuclease de restrição que permite a identificação de células que têm integração do elemento de edição no *locus* alvo com base na análise de polimorfismo de comprimento do fragmento de restrição do *locus* alvo e suas regiões flanqueando ou um ácido nucleico amplificado a partir do mesmo.

[00111]Qualquer um e todos dos elementos de edição divulgados aqui podem compreender uma ou mais alterações de nucleotídeo que causam uma ou mais modificações de aminoácido (por exemplo, substituição, inserção, ou deleção) em proteína beta globina quando integrada no *locus* alvo. Em certas modalidades, a proteína beta globina modificada é um equivalente funcional da beta globina de tipo selvagem, isto é, pode funcionar como uma beta globina de tipo selvagem. Em certas modalidades, a beta globina funcionalmente equivalente compreende adicionalmente pelo

menos uma característica não encontrada em beta globina de tipo selvagem, por exemplo, a capacidade para inibir a agregação de beta globina carregando a mutação SCD.

[00112]Em certas modalidades, um elemento de edição como descrito aqui compreende pelo menos 0, 1, 2, 10, 100, 200, 500, 1000, 1500, 2000, 3000, 4000, ou 5000 nucleotídeos. Em certas modalidades, o elemento de edição compreende ou consiste em 1 a 5000, 1 a 4500, 1 a 4000, 1 a 3000, 1 a 2000, 1 a 1000, 1 a 500, 1 a 200, 1 a 100, 1 a 50, ou 1 a 10 nucleotídeos.

[00113]Em certas modalidades, um elemento de edição como descrito aqui compreende ou consiste em um éxon, um íntron, uma região 5' não traduzida (UTR), um 3' UTR, um promotor, um doador de emenda, um aceitador de emenda, um elemento de salto ribossômico, uma sequência codificando um RNA não codificante, um isolante, um gene, ou uma combinação dos mesmos.

[00114]Em certas modalidades, o elemento de edição compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28. Em certas modalidades, o elemento de edição consiste na sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-26.

[00115]Braços de homologia usados nos genomas de correção divulgados aqui podem ser dirigidos para qualquer região do gene alvo (por exemplo, o gene HBB) ou um gene próximo do genoma. A identidade e o posicionamento precisos do braço de homologias são determinados pela identidade do elemento de edição e/ou o *locus* alvo.

[00116]Braços de homologia empregados nos genomas de correção divulgados aqui são substancialmente idênticos ao

genoma flanqueando um *locus* alvo (por exemplo, um *locus* alvo no gene HBB). Em certas modalidades, o braço de homologia 5' tem pelo menos cerca de 90% (por exemplo, pelo menos cerca de 95%, 96%, 97%, 98%, 99%, ou 99,5%) identidade de sequência de nucleotídeo para uma primeira região genômica 5' para o *locus* alvo. Em certas modalidades, o braço de homologia 5' tem 100% de identidade de sequência de nucleotídeo para a primeira região genômica. Em certas modalidades, o braço de homologia 3' tem pelo menos cerca de 90% (por exemplo, pelo menos cerca de 95%, 96%, 97%, 98%, 99%, ou 99,5%) identidade de sequência de nucleotídeo para uma segunda região genômica 3' para o *locus* alvo. Em certas modalidades, o braço de homologia 3' tem 100% de identidade de sequência de nucleotídeo para a segunda região genômica. Em certas modalidades, os braços de homologia 5' e 3' são cada um pelo menos cerca de 90% (por exemplo, pelo menos cerca de 95%, 96%, 97%, 98%, 99%, ou 99,5%) idênticos a uma primeira região genômica 5' para o *locus* alvo e uma segunda região genômica 3' para o *locus* alvo, respectivamente. Em certas modalidades, os braços de homologia 5' e 3' são cada 100% idênticos às primeira e segunda regiões genômicas, respectivamente. Em certas modalidades, as diferenças em sequências de nucleotídeos do braço de homologia 5' e a primeira região genômica, e/ou a diferença em sequências de nucleotídeos do braço de homologia 3' e a segunda região genômica, compreendem, consistem essencialmente em ou consistem em diferenças não codificantes em sequências de nucleotídeo.

[00117]O técnico no assunto reconhecerá que os braços

de homologia não precisam ser 100% idênticos à sequência genômica flanqueando o *locus* alvo para ser capaz de mediar a integração de um elemento de edição neste *locus* alvo por recombinação homóloga. O técnico no assunto reconhecerá ainda que em situações onde um braço de homologia não é 100% idêntico à sequência genômica flanqueando o *locus* alvo, recombinação homóloga entre o braço de homologia e o genoma pode alterar a sequência genômica flanqueando o *locus* alvo de modo que ela se torna idêntica à sequência do braço de homologia usado.

[00118]Em certas modalidades, a primeira região genômica 5' para o *locus* alvo está localizada em uma primeira janela de edição, em que a sequência de nucleotídeo da primeira janela de edição consiste em uma sequência selecionada dentre o grupo consistindo em SEQ ID NOs: 101-103. Em certas modalidades, a segunda região genômica 3' para o *locus* alvo está localizada em uma segunda janela de edição, em que a sequência de nucleotídeo da segunda janela de edição consiste em uma sequência selecionada dentre o grupo consistindo em SEQ ID NOs: 101-103. Em certas modalidades, a primeira região genômica 5' para o *locus* alvo está localizada em uma primeira janela de edição, em que a sequência de nucleotídeo da primeira janela de edição consiste em uma sequência selecionada dentre o grupo consistindo em SEQ ID NOs: 101-103; e a segunda região genômica 3' para o *locus* alvo está localizada em uma segunda janela de edição, em que a sequência de nucleotídeo da segunda janela de edição consiste em uma sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103.

[00119]Em certas modalidades, a primeira e segunda janelas de edição são diferentes. Em certas modalidades, a primeira janela de edição está localizada 5' para a segunda janela de edição. Em certas modalidades, a primeira janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101. Em certas modalidades, a segunda janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102. Em certas modalidades, a primeira janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, e na segunda janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102. Em certas modalidades, a primeira região genômica consiste em uma sequência que é mais curta do que a sequência da primeira janela de edição. Em certas modalidades, a primeira região genômica consiste na sequência da primeira janela de edição. Em certas modalidades, a segunda região genômica consiste em uma sequência que é mais curta do que a sequência da segunda janela de edição. Em certas modalidades, a segunda região genômica consiste na sequência da segunda janela de edição.

[00120]Em certas modalidades, a primeira e segunda janelas de edição são iguais. Em certas modalidades, o *locus* alvo é uma ligação internucleotídeos ou uma sequência de nucleotídeo na janela de edição, em que o primeiro locus genômico consiste em uma primeira porção da janela de edição 5' para o *locus* alvo, e o segundo *locus* genômico consiste em uma segunda porção da janela de edição 3' para o *locus* alvo. Em certas modalidades, a primeira porção da janela de edição consiste na sequência a partir da extremidade 5' da janela de edição para o nucleotídeo

adjacentemente 5' para o *locus* alvo. Em certas modalidades, a segunda porção da janela de edição consiste na sequência a partir do nucleotídeo adjacentemente 3' para o *locus* alvo para a extremidade 3' da janela de edição. Em certas modalidades, a primeira porção da janela de edição consiste na sequência a partir da extremidade 5' da janela de edição para o nucleotídeo adjacentemente 5' para o *locus* alvo, e a segunda porção da janela de edição consiste na sequência a partir do nucleotídeo adjacentemente 3' para o *locus* alvo para a extremidade 3' da janela de edição. Em certas modalidades, a janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 103. Em certas modalidades, a primeira e segunda porções das janelas de edição têm comprimentos substancialmente iguais (por exemplo, a razão do comprimento para a porção mais curta para o comprimento da porção mais longa é maior do que 0,5, 0,55, 0,6, 0,65, 0,7, 0,75, 0,8, 0,85, 0,9, 0,95, 0,96, 0,97, 0,98, ou 0,99).

[00121]Em certas modalidades, o braço de homologia 5' tem um comprimento de cerca de 50 a cerca de 4000 nucleotídeos (por exemplo, cerca de 100 a cerca de 3000, cerca de 200 a cerca de 2000, cerca de 500 a cerca de 1000 nucleotídeos). Em certas modalidades, o braço de homologia 5' tem um comprimento de cerca de 800 nucleotídeos. Em certas modalidades, o braço de homologia 5' tem um comprimento de cerca de 100 nucleotídeos. Em certas modalidades, o braço de homologia 3' tem um comprimento de cerca de 50 a cerca de 4000 nucleotídeos (por exemplo, cerca de 100 a cerca de 3000, cerca de 200 a cerca de 2000, cerca de 500 a cerca de 1000 nucleotídeos). Em certas

modalidades, o braço de homologia 3' tem um comprimento de cerca de 800 nucleotídeos. Em certas modalidades, o braço de homologia 3' tem um comprimento de cerca de 100 nucleotídeos. Em certas modalidades, cada um dos braços de homologia 5' e 3' independentemente tem um comprimento de cerca de 50 a cerca de 4000 nucleotídeos (por exemplo, cerca de 100 a cerca de 3000, cerca de 200 a cerca de 2000, cerca de 500 a cerca de 1000 nucleotídeos). Em certas modalidades, os braços de homologia 5' e 3' têm um comprimento de cerca de 800 nucleotídeos.

[00122]Em certas modalidades, os braços de homologia 5' e 3' têm comprimentos de nucleotídeo substancialmente iguais. Em certas modalidades, os braços de homologia 5' e 3' têm comprimentos de nucleotídeo assimétricos. Em certas modalidades, a assimetria em comprimento de nucleotídeo é definida por uma diferença entre os braços de homologia 5' e 3' de até 90% no comprimento, tal como até uma diferença de 80%, 70%, 60%, 50%, 40%, 30%, 20%, ou 10% no comprimento.

[00123]Em certas modalidades, o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42.

[00124]Em certas modalidades, o genoma de correção compreende 5' a 3': a sequência especificada em SEQ ID NO: 101; um elemento de salto ribossômico; uma sequência de codificação de HBB que foi alterada silenciosamente para ser menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idêntica à sequência de codificação de HBB de tipo selvagem; uma sequência de poliadenilação exógena (por exemplo, como especificada em SEQ ID NO: 76, 77, 78, ou

79); e a sequência especificada em SEQ ID NO: 102. Em certas modalidades, o genoma de correção compreende 5' a 3': a sequência especificada em SEQ ID NO: 101; um elemento de salto ribossômico; a sequência especificada em SEQ ID NO: 99; uma sequência de poliadenilação exógena (por exemplo, como especificada em SEQ ID NO: 76, 77, 78, ou 79); e a sequência especificada em SEQ ID NO: 102. Em certas modalidades, o genoma de correção compreende 5' a 3': a sequência especificada em SEQ ID NO: 101; um elemento de salto ribossômico; a região de codificação do primeiro éxon de HBB que foi alterada silenciosamente para ser menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idêntica a região de codificação de tipo selvagem do primeiro éxon de HBB; um primeiro íntron não nativo opcional; o segundo éxon de HBB que foi alterado silenciosamente para ser menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idêntico ao segundo éxon de HBB de tipo selvagem; um segundo íntron opcional; a região de codificação do terceiro éxon de HBB que foi alterada silenciosamente para ser menos que 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idêntica à região de codificação de tipo selvagem do terceiro éxon de HBB; uma sequência de poliadenilação exógena (por exemplo, como especificada em SEQ ID NO: 76, 77, 78, ou 79); e a sequência especificada em SEQ ID NO: 102. Em certas modalidades, o genoma de correção compreende 5' a 3': a sequência especificada em SEQ ID NO: 101; um elemento de salto ribossômico; a sequência especificada em SEQ ID NO: 105; um primeiro íntron não nativo opcional; a sequência especificada em SEQ ID NO: 106; um segundo íntron opcional; a sequência

especificada em SEQ ID NO: 107; uma sequência de poliadenilação exógena (por exemplo, como especificada em SEQ ID NO: 76, 77, 78, ou 79); e a sequência especificada em SEQ ID NO: 102. Em certas modalidades, o genoma de correção compreende a sequência de nucleotídeo especificada em SEQ ID NO: 104.

[00125]Em certas modalidades, os genomas de correção divulgados aqui compreendem adicionalmente uma sequência de nucleotídeo 5' de repetição terminal invertida 5' (5' ITR) da sequência de nucleotídeo de braço de homologia 5', e uma sequência de nucleotídeo 3' de repetição terminal invertida 3' (3' ITR) da sequência de nucleotídeo de braço de homologia 3'. As sequências de ITR de qualquer sorotipo de AAV ou variante do mesmo podem ser usadas nos genomas de correção divulgados aqui. Os 5' e 3' ITRs podem ser de um AAV de sorotipo igual ou de AAVs de diferentes sorotipos. Exemplares ITRs para uso nos genomas de correção divulgados aqui são especificados em SEQ ID NO: 18-21 aqui. Em certas modalidades, a sequência de nucleotídeo 5' ITR e a sequência de nucleotídeo 3' ITR são substancialmente complementares a cada outra (por exemplo, são complementares a cada outra, exceto pela falta de correspondência em posições 1, 2, 3, 4 ou 5 de nucleotídeos em 5' ou 3' ITR).

[00126]Em certas modalidades, 5' ITR ou 3' ITR é de AAV2. Em certas modalidades, ambos os 5' ITR e 3' ITR são de AAV2. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:18, ou a sequência

de nucleotídeo 3' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:19. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:18, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:19. Em certas modalidades, o genoma de correção compreende um elemento de edição tendo a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28, uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:18, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:19. Em certas modalidades, o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42, uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:18, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:19. Em certas modalidades, o genoma de correção consiste em 5' a 3' uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:18, a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:19.

[00127]Em certas modalidades, os 5' ITR ou 3' ITR são de AAV5. Em certas modalidades, ambos os 5' ITR e 3' ITR são de AAV5. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% (por exemplo, pelo

menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:20, ou a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:21. Em certas modalidades, a sequência de nucleotídeo 5' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:20, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% (por exemplo, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99%, ou 100%) identidade de sequência para SEQ ID NO:21. Em certas modalidades, o genoma de correção compreende um elemento de edição tendo a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28, uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:20, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:21. Em certas modalidades, o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42, uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:20, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:21. Em certas modalidades, o genoma de correção consiste em 5' a 3' uma sequência de nucleotídeo 5' ITR tendo a sequência de SEQ ID NO:20, a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42, e uma sequência de nucleotídeo 3' ITR tendo a sequência de SEQ ID NO:21.

[00128]Em certas modalidades, o genoma de correção divulgado aqui tem um comprimento de cerca de 0,5 a cerca de 8 kb, e qualquer faixa encerrada entre os mesmos (por

exemplo, cerca de 1 a cerca de 5, cerca de 2 a cerca de 5, cerca de 3 a cerca de 5, cerca de 4 a cerca de 5, cerca de 4,5 a cerca de 4,8, ou cerca de 4,7 kb).

[00129]Os genomas de correção divulgados aqui podem ser configurados para integrar um elemento de edição em qualquer *locus* alvo desejado do gene HBB. Em certas modalidades, o *locus* alvo é uma mutação (por exemplo, inserção, deleção ou substituição de um ou mais nucleotídeos) em uma sequência de gene HBB relativa à correspondente sequência de gene HBB de tipo selvagem. Em certas modalidades, o *locus* alvo está em uma mutação de ponto de nucleotídeo ou deleção no gene HBB. As mutações de ponto de HBB exemplares ou deleção incluem, sem limitação, G em posição -87, G em posição -31, A em posição -30, G em posição -29, G em posição -28, T em posição -10, C em posição 1, A em posição 1, G em posição 2, deleção de C e T em posições 17 e 18, A em posição 19, deleção de A em posição 20, T em posição 20, deleção de A e A em posições 25 e 26, adição de G após posição 26, A em posição 47, A em posição 48, deleção de C em posição 51, A em posição 52, G em posição 58, G em posição 59, A em posição 79, T em posição 82, adição de C após posição 84, T em posição 93, A em posição 93, C em posição 97, C em posição 98, G em posição 202, G em posição 208, C em posição 222, deleção de T em posição 241 ou 242, deleção de T e T e C e T em posições 254 a 257, T em posição 260, deleção de C em posição 264 ou 265, adição de A após posição 343, deleção de G e T em posições 399 e 400, T em posição 401, adição de A após posição 417, A em posição 446, T em posição 1099, A em posição 1293, T em posição 1344. Em certas modalidades,

o *locus* alvo é a mutação da doença de célula falciforme (isto é, T em posição 20 do gene HBB). Em certas modalidades, o *locus* alvo é uma região de cromossomo ou uma ligação internucleotídeos em éxon 1 do gene HBB, por exemplo, imediatamente seguindo o códon de partida endógeno (por exemplo, a ligação internucleotídeos entre nucleotídeo 3 e nucleotídeo 4 do gene HBB). Em certas modalidades, o *locus* alvo é uma região de cromossomo ou uma ligação internucleotídeos em íntron 1 do gene HBB. Em certas modalidades, o *locus* alvo consiste no códon de parada nativo de um gene HBB de tipo selvagem ou nos nucleotídeos correspondentes de um gene HBB mutante. Em certas modalidades, o *locus* alvo consiste na ligação internucleotídeos adjacente 5' para o códon de parada de um gene HBB de tipo selvagem ou a ligação internucleotídeos correspondente de um gene HBB mutante.

[00130]As composições de AAV divulgadas aqui são particularmente vantajosas em que elas são capazes de corrigir uma mutação em um gene HBB em uma célula com elevada eficiência tanto *in vivo* como *in vitro*. Em certas modalidades, a eficiência de integração do elemento de edição no *locus* alvo é pelo menos 0,1% (por exemplo pelo menos 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, 1%, 1,5%, 2%, 2,5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, ou 95%) quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco hematopoiéticas humanas CD34⁺ sob condições de transdução de AAV padrão. Em certas modalidades, a frequência alélica de integração do elemento de edição no

locus alvo é pelo menos 0,05% (por exemplo pelo menos 0,1%, 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, 1%, 1,5%, 2%, 2,5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, ou 95%) quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco hematopoiéticas humanas CD34⁺ sob condições de transdução de AAV padrão.

[00131]Quaisquer métodos de determinar a eficiência de edição de genes podem ser empregados. Em certas modalidades, células individuais são separadas da população de células transduzida e submetidas a PCR de célula única usando iniciadores de PCR que podem identificar a presença de um elemento de edição corretamente integrado no *locus* alvo. Tal método pode compreender adicionalmente PCR de célula única das mesmas células usando iniciadores de PCR que amplificam seletivamente um *locus* alvo não modificado. Deste modo, o genótipo das células pode ser determinado. Por exemplo, se o PCR de célula única mostrou que uma célula tem tanto um *locus* alvo editado como um *locus* alvo não modificado, então a célula seria considerada heterozigótica para o gene HBB editado.

[00132]Adicionalmente ou alternativamente, em certas modalidades, PCR mediado por amplificação linear (LAM-PCR), PCR quantitativo (qPCR) ou Droplet Digital PCR (ddPCR) pode ser realizado em DNA extraído de uma população de células transduzidas (por exemplo, um tecido ou órgão) para avaliar a frequência alélica de integração. Em certas modalidades, o DNA extraído é analisado por Droplet Digital PCR (ddPCR) usando pelo menos dois pares de iniciadores que detectam

diferentes sequências. Por exemplo, o ddPCR pode empregar um primeiro par de iniciadores que detectam o *loci* alvo não editado e *loci* alvo editado, um segundo par de iniciadores que detecta uma sequência presente nos vetores não integrados e integrados, e opcionalmente um terceiro par de iniciadores que detecta uma sequência no braço de homologia, que está presente no *loci* alvo não editado e editado, assim como vetores não integrado e integrado. Após a correção da possibilidade da co-divisão de um DNA genômico não editado e um vetor não integrado, a porcentagem de gotículas positivas com tanto o primeiro como o segundo par de iniciadores corresponde à frequência alélica de integração. Um exemplo deste método é descrito aqui no Exemplo 1.

[00133]Adicionalmente ou alternativamente, em certas modalidades, o *locus* de HBB pode ser amplificado a partir do DNA extraído de uma população de células transduzidas (por exemplo, um tecido ou órgão) ou por PCR usando iniciadores que se ligam a regiões do gene HBB flanqueando a região genômica englobada pelo genoma de correção, ou por PCR mediado por amplificação linear (LAM-PCR) usando um iniciador que se liga a uma região dentro do genoma de correção (por exemplo, uma região compreendendo uma sequência não nativa exógena para o *locus*). Os amplicons de PCR resultantes podem ser individualmente sequenciados usando técnicas de sequenciamento de próxima geração de molécula única (NGS) para determinar o número relativo de alelos HBB editados e não editados presentes na população de células transduzidas. Estes números podem ser usados para determinar a frequência alélica de integração do

elemento de edição no *locus* alvo.

[00134]Em outro aspecto, a presente divulgação provê composições farmacêuticas compreendendo um AAV como divulgado aqui junto com um excipiente farmacêuticamente aceitável, adjuvante, diluente, veículo ou carreador, ou uma combinação dos mesmos. Um "carreador farmacêuticamente aceitável" inclui qualquer material que, quando combinado com um ingrediente ativo de uma composição, permite ao ingrediente reter atividade biológica e sem causar reações fisiológicas disruptivas, tal como uma reação imune não desejada. Os carreadores farmacêuticamente aceitáveis incluem água, solução salina tamponada com fosfato, emulsões tais com uma emulsão óleo/água e agentes umectantes. Composições compreendendo tais carreadores são formuladas por métodos convencionais bem conhecidos, tais como os especificados em Remington's Pharmaceutical Sciences, Ed. atual, Mack Publishing Co., Easton Pa. 18042, USA; A. Gennaro (2000) "Remington: The Science and Practice of Pharmacy", 20^a. ed., Lippincott, Williams, & Wilkins; Pharmaceutical Dosage Forms and Drug Delivery Systems (1999) H. C. Ansel et al, 7^a. ed., Lippincott, Williams, & Wilkins; e Handbook of Pharmaceutical Excipients (2000) A. H. Kibbe et al, 3^a. ed. Amer. Pharmaceutical Assoc.

III. Método de uso

[00135]Em outro aspecto, a presente divulgação provê métodos para corrigir uma mutação em um gene HBB em uma célula. Os métodos geralmente compreendem transduzir a célula com um AAV defeituoso na replicação como divulgado aqui. Tais métodos são altamente eficazes na correção de mutações em um gene HBB e não requerem clivagem do genoma

no *locus* alvo pela ação de uma nuclease exógena (por exemplo, uma meganuclease, uma nuclease 'dedo-de-zinco', uma nuclease tipo ativador transcripcional (TALEN), ou uma nuclease RNA-guiada tal como uma Cas9) para facilitar tal correção. Consequentemente, em certas modalidades, os métodos divulgados aqui envolvem transduzir a célula com um AAV defeituoso na replicação como divulgado aqui sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[00136]Os métodos divulgados aqui podem ser aplicados a qualquer célula abrigando uma mutação em um gene HBB. O técnico no assunto irá reconhecer que as células que são capazes de diferenciação em eritrócitos são de particular interesse. Consequentemente, em certas modalidades, o método é aplicado a células-tronco, incluindo sem limitação, células-tronco pluripotentes, células-tronco pluripotentes induzidas (iPSCs), e células-tronco hematopoiéticas (HSC). HSCs exemplares para os quais o métodos podem ser aplicados incluem, sem limitação, HSCs CD34⁺.

[00137]Os métodos divulgados aqui podem ser realizados *in vitro* para fins de pesquisa ou podem ser realizados *ex vivo* ou *in vivo* para fins terapêuticos.

[00138]Em certas modalidades, a célula a ser transduzida é retirada de um indivíduo e transduzida para corrigir uma mutação no gene HBB *ex vivo* de acordo com os métodos divulgados aqui, e a célula transduzida é subsequentemente administrada de volta ao indivíduo. Consequentemente, em certas modalidades, a presente

divulgação provê um método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método compreendendo: transduzir uma célula-tronco (por exemplo, uma célula-tronco hematopoiética CD34+) *ex vivo* com um AAV defeituoso na replicação como divulgado aqui para obter uma célula transduzida; e administrar a célula transduzida para o indivíduo. As células transduzidas podem ser selecionada para integração genética correta e/ou cultivadas para expansão clonal antes da administração para o indivíduo. Em certas modalidades, a célula-tronco a ser transduzida é obtida da medula óssea, sangue do cordão umbilical ou sangue periférico, a célula-tronco é opcionalmente selecionada por um método com base em um ou mais marcadores celulares (por exemplo, tamanho da célula, densidade da célula e marcadores de superfície, tal como CD34). Em certas modalidades, a célula-tronco é autóloga, isto é, do indivíduo ao qual as células, após a transdução do AAV, serão administradas. Em certas modalidades, a célula-tronco é alogênica para o indivíduo em necessidade da mesma, isto é, a célula-tronco é obtida a partir de um doador que é geneticamente não idêntico ao indivíduo recipiente. Conseqüentemente, em certas modalidades, a presente divulgação provê um método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método compreendendo: transduzir uma célula-tronco alogênica (por exemplo, uma célula-tronco hematopoiética CD34+) *ex vivo* com um AAV defeituoso na replicação, como divulgado aqui, para obter uma célula transduzida; e administrar a célula transduzida para o indivíduo. Em certas modalidades, a

célula-tronco alogênica é derivada de um doador compatível. O técnico no assunto reconhecerá que, para aplicações alogênicas, a célula transduzida pode requerer modificações adicionais antes da administração, por exemplo, modificações genéticas para evitar e/ou reduzir a ocorrência de doença enxerto-versus-hospedeiro (GVHD). O indivíduo pode ser um indivíduo humano ou um indivíduo roedor (por exemplo, um camundongo) contendo células de precursor de eritrócitos humanas. Os indivíduos camundongos apropriados incluem, com limitação, camundongos em que células-tronco humanas (por exemplo, HSCs CD34⁺ humanas) foram enxertadas. Qualquer doença ou distúrbio associado com uma mutação do gene HBB pode ser tratada usando os métodos divulgados aqui. As doenças ou distúrbios apropriadas incluem, sem limitação, beta talassemia ou doença de célula falciforme. Em certas modalidades, a célula é transduzida sem co-transdução de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[00139]Em certas modalidades, a célula a ser transduzida é de um indivíduo mamífero e o AAV é administrado ao indivíduo em uma quantidade eficaz para transduzir a célula no indivíduo. Consequentemente, em certas modalidades, a presente divulgação provê um método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método geralmente compreendendo administrar ao indivíduo uma quantidade eficaz de um defeito de replicação, como divulgado aqui. O indivíduo pode ser um indivíduo humano, um indivíduo primata não humano (por exemplo, *Macaca*

fascicularis), ou um indivíduo roedor (por exemplo, um camundongo) contendo células de precursor de eritrócitos humano. Os indivíduos camundongos apropriados incluem, sem limitação, camundongos em que células-tronco humanas (por exemplo, HSCs CD34⁺ humanas) foram enxertadas. Qualquer doença ou distúrbio associado com uma mutação do gene HBB pode ser tratado usando os métodos divulgados aqui. As doenças ou distúrbios apropriados incluem, sem limitação, beta talassemia ou doença de célula falciforme. Em certas modalidades, a célula é transduzida sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

[00140]Os métodos divulgados aqui são particularmente vantajosos em que eles são capazes de corrigir um gene HBB em uma célula com elevada eficiência tanto *in vivo* como *in vitro*. Em certas modalidades, a eficiência de integração do elemento de edição no *locus* alvo é pelo menos 0,1% (por exemplo pelo menos 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, 1%, 1,5%, 2%, 2,5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, ou 95%) quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão. Em certas modalidades, a frequência alélica de integração do elemento de edição no *locus* alvo é pelo menos 0,05% (por exemplo pelo menos 0,1%, 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, 1%, 1,5%, 2%, 2,5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, ou 95%) quando o AAV é contatado *in vitro* na ausência

de uma nuclease exógena com uma população de células-tronco humanas CD34+ sob condições de transdução de AAV padrão. Quaisquer métodos de determinação da eficiência de edição de genes podem ser empregados, incluindo, sem limitação, os aqui descritos.

[00141]Em certas modalidades, transdução de uma célula com uma composição de AAV, como divulgado aqui, pode ser realizada como provido aqui, ou por qualquer método de transdução conhecido do técnico no assunto. Em certas modalidades, a célula pode ser contatada com o AAV em uma multiplicidade de infecção (MOI) de 50.000; 100.000; 150.000; 200.000; 250.000; 300.000; 350.000; 400.000; 450.000; ou 500.000, ou em qualquer MOI que possibilite uma transdução ótima da célula. Em certas modalidades, o indivíduo pode ser administrado com o AAV em uma dose de cerca de 10^{11} , 10^{12} , 10^{13} , 10^{14} , ou 10^{15} genomas de vetor por kg de peso corporal.

[00142]Uma composição de AAV divulgada aqui pode ser administrada a um indivíduo por qualquer via apropriada, incluindo, sem limitação, vias intravenosa, intraperitoneal, subcutânea, intramuscular, intranasal, tópica ou intradérmica. Em certas modalidades, a composição é formulada para administração via injeção intravenosa ou injeção subcutânea.

IV. Sistemas de empacotamento de AAV

[00143]Em outro aspecto, a presente divulgação provê sistemas de empacotamento para preparação recombinante de um AAV defeituoso em replicação divulgados aqui. Tais sistemas de empacotamento geralmente compreendem: uma sequência de nucleotídeo Rep codificando uma ou mais

proteínas Rep de AAV; uma sequência de nucleotídeo Cap codificando uma ou mais proteínas de capsídeo de AAV Clade F como divulgado aqui; e um genoma de correção para correção de uma mutação no gene HBB como divulgado aqui, em que o sistema de empacotamento é operativo em uma célula para encerrar o genoma de correção no capsídeo para formar o AAV.

[00144]Em certas modalidades, o sistema de empacotamento compreende um primeiro vetor compreendendo a sequência de nucleotídeo Rep e a sequência de nucleotídeo Cap, e um segundo vetor compreendendo o genoma de correção. Como usado no contexto de um sistema de empacotamento como descrito aqui, um "vetor" refere-se a uma molécula de ácido nucleico que é um veículo para introduzir ácidos nucleicos em uma célula (por exemplo, um plasmídeo, um vírus, um cosmídeo, um cromossomo artificial, etc.).

[00145]Qualquer proteína Rep de AAV pode ser empregada nos sistemas de empacotamento divulgados aqui. Em certas modalidades do sistema de empacotamento, a sequência de nucleotídeo Rep codifica uma proteína Rep de AAV2. As proteínas Rep de AAV2 incluem, sem limitação, Rep 78/68 ou Rep 68/52. Em certas modalidades do sistema de empacotamento, a sequência de nucleotídeo codificando a proteína Rep de AAV2 compreende uma sequência de nucleotídeo que codifica uma proteína tendo uma identidade percentual mínima de sequência para a sequência de aminoácido Rep de AAV2 de SEQ ID NO:22, em que a identidade percentual mínima de sequência é de pelo menos 70% (por exemplo, pelo menos 75%, pelo menos 80%, pelo menos 85%, pelo menos 90%, pelo menos 95%, pelo menos 98%, pelo menos

99%, ou 100%) através do comprimento da sequência de aminoácido da proteína Rep de AAV2. Em certas modalidades do sistema de empacotamento, a proteína Rep de AAV2 tem a sequência de aminoácido especificada em SEQ ID NO:22.

[00146]Em certas modalidades do sistema de empacotamento, o sistema de empacotamento compreende adicionalmente um terceiro vetor, por exemplo, um vetor de vírus auxiliar. O terceiro vetor pode ser um terceiro vetor independente, integrante com o primeiro vetor, ou integrante com o segundo vetor. Em certas modalidades, o terceiro vetor compreende genes codificando proteínas de vírus auxiliar.

[00147]Em certas modalidades do sistema de empacotamento, o vírus auxiliar é selecionado dentre o grupo consistindo em adenovírus, vírus de herpes (incluindo vírus herpes simplex (HSV)), poxvírus (tal como vírus de vaccinia), citomegalovírus (CMV), e baculovirus. Em certas modalidades do sistema de empacotamento, onde o vírus auxiliar é adenovírus, o genoma do adenovírus compreende um ou mais genes de RNA de adenovírus selecionados dentre o grupo consistindo em E1, E2, E4 e VA. Em certas modalidades do sistema de empacotamento, onde o vírus auxiliar é HSV, o genoma de HSV compreende um ou mais de genes de HSV selecionados dentre o grupo consistindo em UL5/8/52, ICP0, ICP4, ICP22 e UL30/UL42.

[00148]Em certas modalidades do sistema de empacotamento, o primeiro, segundo, e/ou terceiro vetores estão contidos dentro de um ou mais plasmídeos transfectantes. Em certas modalidades, o primeiro vetor e o terceiro vetor estão contidos dentro de um primeiro

plasmídeo transfectante. Em certas modalidades, o segundo vetor e o terceiro vetor estão contidos dentro de um segundo plasmídeo transfectante.

[00149]Em certas modalidades do sistema de empacotamento, o primeiro, segundo, e/ou terceiro vetores estão contidos dentro de um ou mais vírus auxiliares recombinantes. Em certas modalidades, o primeiro vetor e o terceiro vetor estão contidos dentro de um vírus auxiliar recombinante. Em certas modalidades, o segundo vetor e o terceiro vetor estão contidos dentro de um vírus auxiliar recombinante.

[00150]Em um aspecto adicional, a divulgação provê um método para preparação recombinante de um AAV como descrito aqui, em que o método compreende transfectar ou transduzir uma célula com um sistema de empacotamento, como descrito sob condições operativas para encerrar o genoma de correção no capsídeo para formar o AAV, como descrito aqui. Métodos exemplares para a preparação recombinante de um AAV incluem transfecção transiente (por exemplo, com um ou mais plasmídeos de transfecção contendo um primeiro, e um segundo, e opcionalmente um terceiro vetor, como descrito aqui), infecção viral (por exemplo com um ou mais vírus auxiliares recombinantes, tal como um adenovírus, poxvírus (tal como vírus de vaccinia), vírus de herpes (incluindo HSV, citomegalovírus, ou baculovírus, contendo um primeiro, e um segundo e, opcionalmente, um terceiro vetor, como descrito aqui), e infecção ou transfecção de linhagens de células produtoras estáveis (por exemplo, com uma célula produtora estável, tal como uma célula de mamífero ou inseto, contendo uma sequência de nucleotídeo Rep

codificando uma ou mais proteínas Rep de AAV e/ou uma sequência de nucleotídeo Cap codificando uma ou mais proteínas de capsídeo de AAV Clade F como descrito aqui, e com um genoma de correção como descrito aqui sendo entregue na forma de um plasmídeo transfectante ou um vírus auxiliar recombinante).

V. Exemplos

[00151]Os vetores de AAV recombinantes divulgados aqui medeiam a edição de genes altamente eficiente *in vitro* e *in vivo* via um mecanismo com base em reparo dependente de homologia livre de nuclease. Os seguintes exemplos demonstram a correção eficiente de um gene HBB que é mudado em algumas doenças humanas, tal como doença de célula falciforme (SCD) e beta talassemia, usando um vetor à base de AAV como divulgado aqui. Estes exemplos são oferecidos a título de ilustração e não de limitação.

Exemplo 1: Seleção de capsídeo de AAV para edição de genes de células mutantes HBB

[00152]Este exemplo caracteriza a eficiência de integração de um vetor de AAV para edição de genes, AAVS1-FP, empacotado em capsídeos de AAV clade F, tal como AAVHSC7, AAVHSC15, e AAVHSC17, em células mutantes-HBB. AAVHSC7, AAVHSC15, e AAVHSC17, também conhecidos como AAVF7, AAVF15, e AAVF17, respectivamente, são completamente descritos em WO2016049230A1, que é incorporado aqui por referência em sua totalidade.

[00153]AAVS1-FP, o vetor de edição de genes empregado aqui, foi completamente descrito em WO2016049230A1. Ele compreende de 5' a 3': uma repetição terminal invertida 5' de AAV2 (ITR), um braço de homologia 5' consistindo em 800

nucleotídeos tendo a sequência do DNA a montante de um *locus* alvo, um aceitador de emenda, um elemento 2A, uma sequência de codificação de uma proteína fluorescente (FP), um braço de homologia 3' consistindo em 800 nucleotídeos tendo a sequência do DNA a jusante de um *locus* alvo, e um AAV2 3' ITR, em que o *locus* alvo está em íntron 1 de PPP1R12C humano no *locus* no cromossomo 19 de AAVS1, e em que, após a recombinação homóloga entre o vetor AAVS1-FP e o genoma humano, éxon 1 de PPP1R12C, o elemento 2A, e a sequência de codificação FP estão em quadro. Porque a sequência de codificação FP no vetor é sem promotor, células transduzidas com este vetor irão expressar o FP somente quando o vetor é integrado no genoma. AAVHSC-scEGFP, um vetor de AAV auto-complementar compreendendo ITRs de AAV2 e um promotor operativamente ligado a uma proteína fluorescente verde intensificada (EGFP), serviram como um controle para eficiência de transdução (ver, por exemplo, Patente US No. 8.628.966, que é incorporada por referência aqui em sua totalidade).

[00154]Células GM16265, uma linhagem celular linfoblastoide (LCL) tendo a mutação A para T em posição 20 em íntron 1 de HBB que é característica de doença de célula falciforme (SCD), foram obtidos do Coriell Institute for Medical Research (Camden, NJ). LCLs foram cultivadas em RPMI suplementado com 15% FCS e 2mM L-glutamina. Células foram semeadas a aproximadamente 200.000 células por ml e divididas quando alcançando 500.000 a 1.000.000 células por ml. Antes de colocar as células, a quantidade de vírus necessária foi calculada para cada transdução. O volume de transdução do vírus não excedeu 10% do volume total da

cavidade. No dia de transdução, células em fase log foram contadas e colocadas em placas. Células foram transduzidas com vírus em uma multiplicidade de infecção (MOI) de $1,5 \times 10^5$. As partículas de AAV empacotadas foram descongeladas em gelo sonicadas em gelo caso necessário antes da transdução, e foram adicionadas a cada cavidade individualmente. As células foram coletadas 48 horas após a transdução.

[00155]Células GM16265 foram transduzidas com o vetor AAVS1-FP empacotado em capsídeos AAVHSC7, AAVHSC15, ou AAVHSC17. A eficiência de integração foi avaliada por citometria de fluxo usando o seguinte método: células foram coletadas usando tampão FACS (1X PBS, 2% FCS, 0,1% azida de sódio) e centrifugadas a 1200 RPM durante 10 minutos. Sobrenadante em excesso foi decantado de modo que aproximadamente 200 μ l permaneceram. 4',6-Diamidino-2-Fenilindol (DAPI) foi adicionado de um estoque de trabalho de 100 μ M imediatamente antes da análise de citometria de fluxo para uma concentração final de 3 μ M.

[00156]Como mostrado em Figura 1A, as porcentagens de células positivas para FP dentre todas as células vivas transduzidas por AAVHSC7-AAVS1-FP e AAVHSC17-AAVS1-FP foram 24,3% (34,0% menos nível de fundo 9,7%) e 7,8% (17,5% menos nível de fundo 9,7%), respectivamente. Como mostrado em Figura 1B, em outra experiência, a porcentagens de células positivas para FP dentre todas as células vivas transduzidas por AAVHSC15-AAVS1-FP e AAVHSC17-AAVS1-FP foram 25,1% (29,8% menos nível de fundo 4,7%) e 37,6% (42,3% menos nível de fundo 4,7%), respectivamente. Estes dados mostram que as células GM16265 podem ser transduzidas

de modo eficiente por AAVS1-FP empacotado em capsídeos AAVHSC7, AAVHSC15, e AAVHSC17.

[00157]A eficiência de integração deste vetor AAVS1-FP empacotado em um capsídeo AAVHSC17 foi também examinada em células-tronco hematopoiéticas primárias humanas CD34⁺ (HSCs). HSCs CD34⁺ primárias humanas foram purificadas a partir de células de sangue periférico humano de doadores com SCD por enriquecimento com as microcontas Miltenyi CD34 duas vezes, ou foram obtidos de ReachBio Inc. que seguiu um procedimento de enriquecimento duplo similar. As células foram cultivadas em meio de Dulbecco modificado por Iscove (IMDM) suplementado com 20% soro de bezerro fetal (FCS), 100 µg/mL estreptomicina, 100 U/mL penicilina, 2mmol/L L-glutamina, 10 ng/mL IL-3 humana, 10 ng/mL IL-6 humana, e 1 ng/mL SCF humana. Cerca de 200.000 células foram colocadas em placas em um meio de 500 µl. As partículas de AAV foram adicionadas diretamente ao meio. As células foram coletadas 48 horas após a transdução.

[00158]A eficiência de edição foi medida por Droplet Digital PCR (ddPCR) usando o sistema de PCR BioRad QX200™ Droplet Digital™. Dois conjuntos de iniciadores e sondas, como mostrado em Tabela 1, foram planejados para quantificar a integração por ddPCR. O conjunto AAVS1_Genomic detectou uma sequência no *locus* de AAVS1 que estava presente fora do braço de homologia no genoma não editado e o genoma editado após a integração alvo-marcada da sequência de codificação FP no *locus* de AAVS1 do genoma. O conjunto AAVS1_FP detectou uma sequência na região de codificação de FP, que estava presente somente no genoma editado após a integração alvo marcada da sequência de

codificação FP no *Alocus* AVS1 do genoma. As duas sondas foram conjugadas para porções fluorescentes de diferentes comprimentos de onda.

Tabela 1: Iniciadores para análise de ddPCR

Nome do iniciador	SEQ NO:	ID	Sequência de Nucleotídeo
AAVS1_Genomic, iniciador dianteiro	90		GCGTTAGAGGGCAGAGTTC
AAVS1_Genomic, iniciador reverso	91		AGCTCCCATAGCTCAGTCT
AAVS1_Genomic, sonda	92		CATTGTCACTTTGCCTGCCCTC
AAVS1_FP iniciador dianteiro	93		GCAATAGCATCACAAATTTTAC
AAVS1_FP, iniciador reverso	94		GATCCAGACATGATAAGATACATTG
AAVS1_FP, sonda	95		TCACTGCATTCTAGTTGTGGTTTGTCCA

[00159] Amostras tendo 100 pg/ μ l de DNA foram divididas em gotículas de óleo. A maior parte das gotículas de óleo não continham molécula de DNA ou uma única molécula de DNA – genoma não editado, positivo para somente o conjunto AAVS1_Genomic; um vetor não integrado, positivo para apenas o conjunto AAVS1_FP; ou um genoma editado, positivos para ambos os conjuntos de iniciador/sonda. A possibilidade de co-divisão foi determinada por várias amostras padrão (ver Regan et al., *A rapid molecular approach for chromosomal phasing*, PLoS One. (2015) 10(3):e0118270, incorporado aqui por referência em sua totalidade). As amostras padrão continham 100 genomas não editados por μ l, 1000 vetor episomal por μ l, e uma faixa de alelos positivos clonados a 1, 5, 10, 15, 20, e 25 alelos editados por μ l, respectivamente. Uma curva padrão de co-divisão contra a razão de alelo não editado para alelo editado foi traçada em gráfico ($R^2 = 0,972$, correlação de Pearson $p < 0.001$).

[00160] Cada amostra foi analisada por ddPCR em pelo

menos três experiências, e as quantidades de gotículas AAVS1_Genomic positivas, AAVS1_FP positivas, e duplo positivas em cada amostra foram medidas e traçadas em gráfico contra a razão conhecida de alelo não editado para alelo editado em cada amostra. Como mostrado em Figura 1C, integração da sequência de codificação FP no genoma foi detectada em cerca de 30% de todos os alelos dos HSCs CD34⁺ primárias humanas. Conseqüentemente, AAVS1-FP empacotado no capsídeo AAVHSC17 eficientemente transduziu HSCs CD34⁺ primárias humanas.

Exemplo 2: Correção *in vitro* de uma mutação de HBB

[00161]Um vetor de correção de HBB à base de AAV chamado hHBB-hL-014, como mostrado em Figura 2, foi gerado. Este vetor de correção foi planejado para corrigir uma mutação de HBB, por exemplo, a mutação de A para T em nucleotídeo 20 na região de codificação de éxon 1 (partindo do códon de partida) do gene HBB em doença de célula falciforme. O vetor hHBB-hL-014 incluía 5' e 3' ITRs de AAV2, flanqueando uma porção de HBB e sua sequência genômica vizinha, em que a mutação de A para T em nucleotídeo 20 em éxon 1 foi invertida. A porção de sequência genômica de HBB foi obtida de HBB de tipo selvagem e seu loci vizinho usando os iniciadores de amplificação como mostrado em Tabela 2, que foram planejados usando Iniciador NCBI Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/). O produto de PCR cobriu todos os éxons e íntrons de HBB, e incluiu adicionalmente 1678 nucleotídeos a montante a partir do sítio de iniciação de transcrição de HBB e 234 nucleotídeos a jusante da sequência de poliadenilação de HBB. Relativo à

mutação de HBB em doença de célula falciforme, este vetor continha braços de homologia (as sequências genômicas 5' e 3' para o sítio de reversão de mutação) de cerca de 1,7 kb de comprimento cada. Integridade dos ITRs foi confirmada por triagem do digesto de restrição usando *Bgl*III, *Msc*I, e *Sma*I e sequenciamento usando um protocolo de sequenciamento específico de ITR (Mroske *et al.*, Hum Gene Ther Methods (2012) 23(2):128-36). O inserto foi verificado por digesto de restrição e sequenciamento Sanger usando os iniciadores mostrados em Tabela 3 aqui. Este vetor foi capaz de corrigir não somente mutações nos éxons e íntrons de HBB, mas também mutações nas regiões 5' e 3' não traduzidas que afetam a expressão de HBB como observado em beta talassemia.

Tabela 2: Iniciadores de amplificação de região genômica de beta globina humana

Nome do iniciador	SEQ NO:	ID	Sequência de Nucleotídeo
HBB2M Gib 5' fwd	50		AGGGGTGGAGTCGTGACGTGCCAAATCAAGCCTCTACTTGAATCC
HBB2M Gib 3' rev	51		AATGATTAACCCGCCATGCTACTTATCTACGTAAACCTAGGCTCCAGATAGCCA

Tabela 3: Iniciadores de sequenciamento de beta globina humana

Nome Iniciador	SEQ NO:	ID	Sequência de Nucleotídeo
Fwd3 seq HBB2	52		GGAAGCAGAACTCTGCAC
Fwd4 seq HBB2	53		GCATTAAGAGGTCTCTAGTTTTTTATC
Fwd5 seq HBB2	54		GATGGTATGGGGCCAAGAGATATATC
Fwd6 seq HBB2	55		GTCTACCCTTGGACCCAGAG
Fwd7 seq HBB2	56		CAGTCTGCCTAGTACATTACTATTTG
Fwd8 seq HBB2	57		CATGTTTCATACCTCTTATCTTCC
Fwd9 seq HBB2	58		GCAAACAGCTAATGCACATTGG
Rev3 seq HBB2	59		CAGAATCCAGATGCTCAAGGCC
Rev4 seq HBB2	60		CCCTGATTTGGTCAATATGT
Rev5 seq HBB2	61		CATCAAGCGTCCCATAGACTCAC
Rev6 seq HBB2	62		GCAGACTTCTCCTCAGGAGTC
Rev7 seq HBB2	63		CTTACAGGACAGAATGGATGAAAAC

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
Rev8 seq HBB2	64	GAAAAGGTCTTCTACTTGGCTC
Rev9 seq HBB2	65	GGTTAACCAAAGAACTGG

[00162] Para facilitar a detecção do gene corrigido, um ligante de 12-bp tendo a sequência de ACTAGTATCGAT (SEQ ID NO: 80), contendo um sítio de restrição *ClaI* e um sítio de restrição *SpeI*, foi inserido no gene HBB. Esta sequência de ligante estava localizada em íntron 1, 117 bp do códon de partida e 97 bp do sítio de reversão de mutação, estabeleceu uma forte ligação genética entre o ligante e a correção de gene desejada. Disrupção de sítios chave de doador e receptor em íntron 1 foi evitada para manter emenda de mRNA de HBB corrigida.

[00163] O vetor hHBB-hL-014 foi empacotado em proteínas de capsídeos AAVHSC15 ou AAVHSC17 usando o método de empacotamento descrito em Chatterjee *et al.*, (1993) *Methods* 5:51-59, que é incorporado por referência aqui em sua totalidade. O título do vírus empacotado foi determinado por qPCR usando os iniciadores e sondas vistos em Tabela 4.

Tabela 4: Iniciadores qPCR de vetor de correção de HBB e sonda

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
qHBB2M Iniciador Dianteiro	66	TGCAGATTAGTCCAGGCAGAAA
qHBB2M Iniciador Reverso	67	GGGTAATCAGTGGTGTCAAATAGGA
qHBB2M Sonda	68	AGTTAGATGTCCCAGTTAA

[00164] Os vírus AAVHSC15-hHBB-hL-014 e AAVHSC17-hHBB-hL-014 foram testados por sua capacidade de editar o gene HBB em células GM16265 usando um teste de integração (TI) alvo-marcado. Neste teste, células foram centrifugadas a 4000 RPM durante 10 minutos, lavadas com 1X PBS, e os grânulos foram congelados a 80°C para uso subsequente.

Grânulos de células congelados foram re-colocados em suspensão em 200 µl para 100.000 células. 1 µl de RNase livre de DNase foi adicionado e incubado a 37°C durante uma hora. 10 µl de 10% SDS e 1,2 µl de proteinase K foram adicionados e incubados a 56°C durante a noite. DNA de alto peso molecular foi extraído por extração padrão com fenol e clorofórmio. Para um alto rendimento de DNA, a extração reversa foi realizada com um volume de 0,5X de tampão Tris-EDTA (TE) (pH 8,0) e adicionado ao tubo final. DNA foi precipitado com acetato de amônio 10M a uma concentração final de 2,5M. Foi adicionado um volume de aproximadamente 4X de etanol 100% gelado. DNA foi deixado precipitar a -80°C por pelo menos uma hora. DNA foi lavado com etanol a 70%, secado, recolocado em suspensão em aproximadamente 30 a 50 µ de TE e quantificado por Nanodrop. Após a quantificação, o DNA foi submetido a um ensaio de "integração alvo-marcada" (TI) baseada em PCR, usando iniciadores apropriados para confirmar a recombinação correta entre o vetor hHBB-hL-014 e o gene HBB.

[00165]Os iniciadores tendo as sequências especificadas na Tabela 5 foram usados para o ensaio de TI neste exemplo. O iniciador HBB2MTI100 alvo-marcou o ligante e sua região vizinha e o iniciador HBB350 alvo-marcou uma sequência genômica fora do braço de homologia. A reação de PCR geraria um amplicon de 2.219 pb a partir de DNA isolado de genomas editados, mas não amplificaria substancialmente a partir de DNA isolado de células não transduzidas ou do vetor hHBB-hL-014 sozinho.

[00166]A reação PCR foi definida como a seguir: até 50 µl de água para PCR; 10 µl de tampão 5X Q5; 5 µl de

betaína; 1 µl de 10mM dNTPs; 1 µl de Iniciador Dianteiro HBB2MTI100 (concentração de 25µM); 1 µl de Iniciador Reverso HBB 350 (concentração de 25 µM); 100 ng a 1 µg de DNA genômico; 1 µl de polimerase NEB Q5 High Fidelity Polymerase. A máquina de PCR foi ajustada como a seguir: desnaturação inicial a 95°C durante 5 minutos; 15 ciclos de desnaturação a 95°C durante 10 segundos, anelamento a 70°C durante 30 segundos, diminuir 0,5 graus a cada ciclo, e extensão a 72°C durante 2 minutos; 20 ciclos de desnaturar a 95°C durante 10 segundos; anelamento a 65°C durante 30 segundos, e extensão a 72°C durante 2 minutos; e uma extensão final a 72°C durante 5 minutos. Os produtos de PCR foram analisados por eletroforese de gel.

[00167]Amplicons com tamanho aparente de 2,2 kb foram isolados e ligados de modo obtuso em uma estrutura dorsal pUC118. Os resultantes plasmídeos foram analisados por digestão de restrição usando endonucleases *ClaI* ou *SpeI* para identificar clones com inserção correta do ligante. Os clones positivos identificados por digestão de restrição foram analisados ainda por sequenciamento de DNA usando iniciadores de oligonucleotídeo M13F e M13R.

Tabela 5: Iniciadores para ensaio de integração alvo-marcado de HBB

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
HBB2MTI100 Iniciador Dianteiro (TM 54,1°C)	69	CTATTGGTCTCCTTAAAATCGATACTAGT
HBB350 Iniciador Reverso (TM 54,8°C)	70	ATATTCAAACCTCCGCAGAACACT

[00168]Como mostrado em Figura 3A, uma banda de PCR de 2,2 kb (indicativa de edição correta por o vetor hHBB-hL-014) foi detectada no ensaio TI em células GM16265

transduzidas com vírus AAVHSC15-hHBB-hL-014 e AAVHSC17-hHBB-hL-014, enquanto este produto de PCR não foi detectado em células GM16265 não transduzidas.

[00169]A correção da mutação A para T na posição 20 do gene HBB nas células GM16265 foi verificada por sequenciamento de clones positivos identificados no ensaio TI. A análise de sequência mostrou que após a transdução com vírus AAVHSC15-hHBB-hL-014 ou AAVHSC17-hHBB-hL-014, a mutação T foi corrigida para A e as mutações assintomáticas próximas também foram corrigidas para o tipo selvagem. Sequenciamento da fita oposta de PCR confirmou a correção das mutações. A inserção do ligante no *locus* de HBB foi detectada em todos os clones. Além disso, nenhum dos clones examinados mostrou quaisquer mutações indesejadas (por exemplo, inserção, deleção ou inversão adicional) nas regiões genômicas correspondentes aos terminais dos braços de homologia.

[00170]Para confirmar adicionalmente a capacidade do vetor hHBB-hL-014 para editar o gene HBB, duas LCLs, GM16266 e GM16267 adicionais, foram usadas. Estas LCLs foram também obtidas do Coriell Institute for Medical Research (Camden, NJ), e foram coletadas de diferentes doadores dos de GM16265. Ambas as LCL tinham a mutação A para T em posição 20 em íntron 1 de HBB. Seguindo os mesmos métodos para cultura e transdução de GM16265, as células foram transduzidas com hHBB-hL-014 empacotado em um capsídeo AAVHSC17.

[00171]Como mostrado em Figura 3B, um amplicon de PCR de cerca de 2,2kb foi detectado a partir de células transduzidas com AAVHSC17-hHBB-hL-014, mas não a partir de

células não transduzidas. Os resultados do sequenciamento mostraram que a mutação SCD foi corrigida nas células transduzidas de todas as três LCLs (incluindo as células GM16265), e não foram detectadas mutações indesejadas (por exemplo, inserção, deleção, ou inversão adicional) nas regiões genômicas correspondendo aos térmicos dos braços de homologia. Os resultados do sequenciamento também confirmaram a inserção de ligante apropriada e a transição sem junção do braço de homologia ao genoma.

[00172]Os resultados acima mostram que o vetor hHBB-hL-014 é capaz de reverter um gene HBB mutante para a sequência de tipo selvagem em linhagens celulares de SCD múltiplas. Conseqüentemente, mutações nos éxons, introns, ou seqüências regulatórias do gene HBB em outras doenças genéticas, tal como beta talassemia, também deveriam ser corrigíveis usando o vetor hHBB-hL-014.

Exemplo 3: Vetores de correção de HBB compreendendo uma seqüência genômica de HBB

[00173]Este exemplo mostra vetores de correção à base de AAV de HBB compreendendo uma seqüência genômica de HBB capaz de corrigir mutações no gene HBB.

a) Vetor de correção de HBB hHBB-hL-001

[00174]O vetor de correção de HBB hHBB-hL-001, como mostrado em Figura 4A, contém seqüência genômica de HBB incluindo todos os éxons, todos os introns, e a seqüência de poliadenilação. Este vetor adicionalmente contém uma região 5' englobando 800 bp a montante do sítio de iniciação de transcrição de HBB (referido como "HBB HAL" na Figura 4A), e uma região 3' englobando 800 bp a jusante da seqüência de poliadenilação de HBB ("HBB HAR" em Figura

4A). O vetor hHBB-hL-001 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 31 (com um ligante TI RE) ou SEQ ID NO:32 (sem um ligante TI RE), e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19). Este vetor corrige não somente mutações nos éxons e introns de HBB, mas mutações nas regiões 5' e 3' não traduzidas que afetam a expressão de HBB, como observado em beta talassemia.

b) Vetor de correção de HBB hHBB-hLW-013

[00175]O vetor de correção de HBB hHBB-hLW-013, como mostrado em Figura 4B, contém elementos genéticos iguais como no vetor de correção de HBB hHBB-hL-001, exceto que as sequências de DNA das regiões de codificação de éxons 1, 2 e 3 são alteradas silenciosamente para ser cerca de 67% idênticas, em vez de completamente idênticas, às sequências de tipo selvagem correspondentes. A identidade de sequência reduzida resulta de alteração dos códons, em que códons degenerados são substituídos por códons originais sem mudar os aminoácidos codificados. Esta alteração silenciosa de códons não é esperada alterar de modo significativo o nível de expressão de HBB. Ao contrário, ela reduz a homologia de éxons de HBB com outros genes ou pseudogenes de globina, tal como epsilon globina (HBE), delta globina (HBD), gama globina 1 (HBG1), gama globina 2 (HBG2), e HBB pseudogene HBPP, assim reduzindo a possibilidade de recombinação indesejada deste vetor em outros *loci* genômicos. Em um exemplo específico, as sequências alteradas silenciosamente da região de codificação de éxon 1, éxon 2, e éxon 3 são especificadas em SEQ ID NOs: 43,44 e 45, respectivamente. O

mostrado em Figura 4C, com
como no vetor de correção de HBB
3' região a jusante da sequência
(referida como "HBB HAR" em Figura 4C)
comprimento. Esta modificação é feita
inclusão das sequências de promotor de
outros genes (por exemplo, GATA1, MYC, etc.)
localizadas cerca de 100 bp a jusante da sequência
poliadenilação de HBB, porque as sequências de pro

TI RE), e
ITR (por exemplo, tendo a
NO: 18) e um 3' ITR (por exemplo, tendo
NCIA de SEQ ID NO: 19).

d) Vetor de correção de HBB hHBB-hLW-012

[00177]O vetor de correção de HBB hHBB-hLW-012, como
mostrado em Figura 4D, contém elementos genéticos iguais

como no vetor de correção de HBB hHBB-hL-011 exceto que as sequências de DNA das regiões de codificação de éxons 1, 2 e 3 são alteradas silenciosamente para serem 67% idênticas às correspondentes sequências de tipo selvagem. Em um exemplo específico, as sequências alteradas silenciosamente da região de codificação de éxon 1, éxon 2, e éxon 3 são especificadas em SEQ ID NOs: 43, 44 e 45, respectivamente. O vetor hHBB-hLW-012 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 37 (com um ligante TI RE) ou 38 (sem um ligante TI RE), e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19).

[00178]Cada um dos quatro vetores de correção de HBB neste exemplo contém uma sequência de ligante, que compreende sítios de reconhecimento e clivagem para as endonucleases de restrição única, para facilitar detecção do gene corrigido. Esta sequência de ligante está localizada em íntron 1, 117 bp do códon de partida. Disrupção de sítios de doador e receptor chave em íntron 1 é evitada para manter emenda de mRNA de HBB corrigida.

[00179]Cada um dos quatro vetores de correção de HBB descritos acima foi gerado e empacotado em AAVHSC17. HSCs CD34⁺ humanas primárias foram transduzidas com os vírus usando o método como descrito em Exemplo 1, e integração foi avaliada por ensaio TI como descrito em Exemplo 2. Como mostrado em Figuras 5A e 5B, hHBB-hL-001, hHBB-hL-011, e hHBB-hLW-012 foram todos capazes de editar o gene HBB.

[00180]A eficiência de edição do genoma foi medida quantitativamente por sequenciamento de próxima geração (NGS). As reações de PCR foram realizadas usando

iniciadores específicos para uma região do genoma fora dos braços de homologia, como mostrado em Tabela 6. Estes iniciadores irão amplificar um produto de 2,342 bp de alelos não editados e editados, mas não irá amplificar o vetor de AAV.

Tabela 6: Iniciadores para a preparação de amostra de NGS

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
HBB350 Iniciador Reverso	70	ATATTCAAACCTCCGCAGAACACT
HBB L NGS S1	81	CCTCTGGGTCCAAGGGTAGA

[00181]A reação PCR foi especificada como a seguir: até 50 µl de água para PCR; 10 µl de tampão 5X Q5; 5 µl de Betaína; 1 µl de 10mM dNTPs; 1 µl de Iniciador Reverso de HBB350 (25 µM); 1 µl de iniciador HBB L NGS S1 (25 µM); 200 ng de DNA genômico; e 1 µl de polimerase Q5 Hifidelity Polymerase. A máquina para o PCR foi ajustada como a seguir: desnaturação inicial a 98°C durante 30 minutos; 30 ciclos de desnaturar a 95°C durante 10 segundos, anelamento a 65°C durante 30 segundos, e extensão a 72°C durante 2 minutos; e uma extensão final a 72°C durante 5 minutos.

[00182]O produto PCR do tamanho correto foi isolado por eletroforese em gel e foi extraído usando o kit Qiagen Qiaquick Gel Extraction Kit de acordo com o protocolo padrão. A ausência de genomas do vetor de amplicons extraídos com gel foi confirmada por PCR usando iniciadores específicos de vetor e um número conhecido de gabaritos do genoma específicos do vetor como controle positivo. Para confirmar ausência de genomas do vetor, as seguintes condições de PCR foram usadas: desnaturação inicial a 98°C durante 30 minutos; 30 ciclos de desnaturar a 98°C durante 10 segundos, anelamento a 66°C durante 30 segundos, e extensão a 72°C durante 1 minuto; e uma extensão final a

72°C durante 2 minutos. O iniciador dianteiro usado foi AAAGTCAGGGCAGAGCCATC (SEQ ID NO: 108) e o iniciador reverso usado foi AATGATTAACCCGCCATGCT (SEQ ID NO: 109), e irá amplificar um amplicon de 1.797 pares de bases.

[00183]O produto PCR extraído foi usado para o sequenciamento de NGS, e/ou quantificação PCR digital, como descrito abaixo. No caso de sequenciamento NGS, o produto de PCR extraído foi submetido a um ciclo abrigado de PCR usando os iniciadores mostrados em Tabela 7. Cada amostra tinha uma combinação singular de iniciadores dianteiro e reverso e o tamanho de cada produto de PCR correto foi cerca de 388 bp. A máquina de PCR foi ajustada como a seguir: desnaturação inicial a 98°C durante 30 minutos; 30 ciclos de desnaturar a 98°C durante 10 segundos, anelamento a 72°C durante 30 segundos, e extensão a 72°C durante 30 segundos; e uma extensão final a 72°C durante 2 minutos.

Tabela 7: Iniciadores para a preparação da amostra de NGS

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
HBB NGS Nest 5 Adaptador Dianteiro 1	82	AATGATACGGCGACCACCGAGATCTACACAAGTAGAGTCTT TCCCTACACGACGCTCTTCCGATCTGGGCATAAAAAGTCAGG GCAGA
HBB NGS Nest 5 Adaptador Dianteiro 2	83	AATGATACGGCGACCACCGAGATCTACACCATGCTTATCTT TCCCTACACGACGCTCTTCCGATCTtGGGCATAAAAAGTCAG GGCAGA
HBB NGS Nest 5 Adaptador Dianteiro 3	84	AATGATACGGCGACCACCGAGATCTACACGCACATCTTCTT TCCCTACACGACGCTCTTCCGATCTatGGGCATAAAAAGTCA GGGCAGA
HBB NGS Nest 5 Adaptador Dianteiro 4	85	AATGATACGGCGACCACCGAGATCTACACTGCTCGACTCTT TCCCTACACGACGCTCTTCCGATCTgatGGGCATAAAAAGTC AGGGCAGA
HBB NGS Nest 5 Adaptador Reverso 1	86	CAAGCAGAAGACGGCATAACGAGATCATGATCGGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTGTCTCCACATGCCCCAG TTTCTA
HBB NGS Nest 5 Adaptador Reverso 2	87	CAAGCAGAAGACGGCATAACGAGATAGGATCTAGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTtGTCTCCACATGCCCCA GTTTCTA
HBB NGS Nest 5 Adaptador Reverso 3	88	CAAGCAGAAGACGGCATAACGAGATGACAGTAAGTGACTGGA GTTCAGACGTGTGCTCTTCCGATCTatGTCTCCACATGCCCC AGTTTCTA

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
HBB NGS Nest 5 Adaptador Reverso 4	89	CAAGCAGAAGACGGCATAACGAGATCCTATGCCGTGACTGGA GTTGAGACGTGTGCTCTTCCGATCTgaGTCTCCACATGCC AGTTTCTA

[00184] Os tamanhos do amplicon foram confirmados por eletroforese em gel, e os produtos de PCR foram purificados usando o kit de Purificação PCR Qiagen Qiaquick de acordo com o protocolo padrão. Amplicons de todas as amostras testadas foram misturados em molaridades iguais, e a concentração foi confirmada por um bioanalisador Advanced Analytical. As amostras foram sequenciadas usando um kit MiSeq V2 300 cycle kit.

[00185] A Tabela 8 mostra o número de leituras aceitáveis com e sem a sequência de ligante presente, assim como o número de leituras com sequências de HBB, como determinado a partir da análise NGS. Porque a presença da sequência de ligante indicou a integração do vetor de AAV, as porcentagens de alelos com ligante representaram as frequências alélicas de integração destes vetores, que estão geralmente entre 0,1% e 1%.

Tabela 8: Frequência alélica de integração de vetores de correção de HBB

Amostra	# ligantes	# sem ligantes	# sequência de ligante adjacente	% com ligante	% leituras capturadas*
HBB-hL-014 #1	4	2703	2710	0,14	99,88
HBB-hL-011 #1	3101	475533	479142	0,64	99,89
HBB-hL-011 #2	5	1917	1925	0,26	99,84
HBB-hL-001 #1	3	449	455	0,66	99,34
HBB-hL-011 #3	2	235	235	0,84	100,85
Controle negativo	0	1	1	0	1
Não transduzida	0	0	0	0	0

*% leituras capturadas é uma métrica de controle de qualidade que indica a porcentagem de leituras passando

pelo controle de qualidade contendo a sequência adjacente à sequência do ligante dividida pela soma nas leituras contendo a sequência do ligante e nenhuma sequência de ligante.

[00186]No caso de PCR digital, o produto de PCR extraído foi submetido à análise de PCR digital usando o sistema PCR de BioRad QX200™ Droplet Digital™. A edição foi determinada por cálculo da ligação entre o alvo genômico e a carga útil de vetor inserido (ligante) e foi medida por detecção da quantidade de gotículas divididas que contém tanto o vetor como o genoma em relação à expectativa de coincidência por possibilidade do emprego de métodos usados para ligação genética entre variantes (ver, por exemplo, Regan et al., *A rapid molecular approach for chromosomal phasing*, *PLoS One*. (2015) 10(3):e0118270, incorporado aqui por referência em sua totalidade). Uma concentração de 0,1 ng/ul de DNA genômico foi analisada através de um mínimo de três experiências por amostra para ligação e medidas usando ddPCR multiplexada com um conjunto de sonda específica do vetor e um conjunto de sonda específica genômica. O conjunto de iniciador e sonda é como a seguir:

Tabela 9: Conjunto de iniciador e sonda usado para PCRdigital

	SEQ ID NO:	Sequência de Nucleotídeo
<i>Conjunto de iniciador e sonda específico de ligante</i>		
HBB Sonda	110	AACTGGGCATGTGGAGACAGAGAA
HBB Ligante F	111	GTTACAAGACAGGACTAGTATCGAT
HBB R	112	TAGACCAATAGGCAGAGAGAGT
<i>Conjunto de iniciador e sonda específico genômico</i>		
HBB0101GDNA F	113	CTGAGCCAAGTAGAAGACCTTT
HBB0101GDNA R	114	CTGTTTCTGCCTGGACTAATCT
HBB0101GDNA Sonda	115	CCCTACTTTCTAAGTCACAGAGGCT

[00187]A fim de medir a edição/ligação de genoma como pelo método acima contra uma quantidade conhecida de

material editado, uma série padrão de DNA foi criada. O padrão consistia de 100 genomas não editados por ul, 1000 vetores episomais por ul e uma faixa de alelos positivos clonados a 1 por ul, 5 por ul, 10 por ul, 15 por ul, 20 por ul, e 25 alelos editados por ul, respectivamente. A quantidade de ligação genética em cada amostra foi medida e traçada em gráfico contra a razão conhecida de alelos não editados para editados em cada amostra ($R^2 = 0,972$, correlação de pearson $p < 0,001$).

[00188]A edição do *locus* de HBB em células de sangue de cordão umbilical misturadas CD34+ primárias transduzidas com vários vetores de edição de HBB AAVHSC7 em um MOI de $1,5 \times 10^5$ foi medida por PCR digital. Células foram coletadas a 48 horas após-transdução e testadas para a porcentagem de alelos editados por um PCR 'out/out' e análise de PCR digital de produtos de PCR. Figura 6 mostra a fração de *loci* editados através das amostras como indicado.

Exemplo 4: Vetores de correção de HBB compreendendo uma sequência de codificação de HBB ou uma porção da mesma

[00189]Este exemplo fornece vetores de correção de HBB que são capazes de inserir uma sequência de codificação de HBB ou uma porção do mesmo no gene HBB, por exemplo, após o códon de partida ou em íntron 1. A sequência inserida pode ser transcrita e traduzido a partir do *locus* nativo sob o controle dos elementos nativos regulatórios de transcrição, assim restaurando a expressão de uma proteína HBB funcional.

[00190]Cada um dos vetores de correção de HBB continha uma sequência de codificação de HBB ou uma porção da mesma

do segundo códon para o códon de parada. A sequência de codificação de HBB ou porção da mesma é seguida por uma sequência de poliadenilação de SV40, que é forte o suficiente para suportar a expressão apropriada e reduzir de modo significativo outra transcrição do resto do gene HBB endógeno. Um cassete de restrição de integração alvo-marcada ("cassete TI RE") compreendendo sítios de reconhecimento e clivagem para uma endonuclease de restrição única é opcionalmente inserido a jusante da sequência de poliadenilação, facilitando a detecção da recombinação homóloga desejada.

a) Vetor de correção de HBB hHBB-hA-009

[00191]O vetor de correção de HBB hHBB-hA-009, como mostrado em Figura 7A, contém uma porção de uma sequência de codificação de HBB de tipo selvagem do segundo códon para o códon de parada (nucleotídeos 4-444 de SEQ ID NO: 27), seguido por uma sequência de poliadenilação de SV40 como descrito acima. O vetor contém adicionalmente um 5' braço de homologia (referido como "HBB HAL" em Figura 7A) compreendendo a sequência genômica de tipo selvagem a montante de e incluindo o códon de partida de HBB, e um 3' braço de homologia (referido como "HBB HAR" em Figura 7A) compreendendo a sequência genômica de tipo selvagem a jusante de, mas não incluindo o códon de partida de HBB. Cada um do braço de homologia 5' e do braço de homologia 3' tem cerca de 800 bp de comprimento. O vetor hHBB-hA-009 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 39, e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19). O braço

de homologia 5' tem a capacidade para corrigir mutações no códon de partida e/ou região 5' não traduzida (UTR) que afetam a expressão de HBB como observado em alguns pacientes com beta talassemia. Como um resultado, a integração de vetor de correção de HBB hHBB-hA-009 pode restaurar a expressão de HBB de tipo selvagem que foi afetada por mutações em 5' UTR, sequência de codificação, ou 3' UTR.

b) Vetor de correção de HBB hHBB-hAW-002

[00192]O vetor de correção de HBB hHBB-hAW-002, como mostrado em Figura 7B, contém os mesmos elementos genéticos como em vetor de correção de HBB hHBB-hA-009, exceto que a porção de sequência de codificação de HBB é alterada silenciosamente para SEQ ID NO: 47, 67% idêntica à correspondente região de sequência de cDNA de tipo selvagem. Como descrito em Exemplo 3, esta alteração de códons não é esperada alterar de modo significativo o nível de expressão de HBB. Ao contrário, ela reduz a homologia de éxons de HBB com outros genes e pseudogenes de globina, assim reduzindo a recombinação indesejada deste vetor em outros *loci* genômicos. O vetor hHBB-hAW-002 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 40, e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19).

c) Vetor de correção de HBB hHBB-h1-010

[00193]O vetor de correção de HBB hHBB-h1-010, como mostrado em Figura 7C, é planejado para inserir uma sequência de codificação de HBB de tipo selvagem em íntron 1 por recombinação homóloga. Especificamente, o sítio de

inserção está entre nucleotídeos 160 e 161 do gene HBB, e a inserção evita a disrupção de sítio chave doador de emendas em íntron 1. O elemento de edição (região inserida) deste vetor contém de 5' a 3' um sítio aceitador de emenda ("SA" em Figura 7C, por exemplo, SEQ ID NO: 14), um elemento de salto ribossômico ("T2A" em Figura 7C, por exemplo, SEQ ID NO: 72) em quadro com o códon de partida de HBB quando da integração, uma sequência de codificação de HBB de tipo selvagem (SEQ ID NO: 27), e uma sequência de poliadenilação SV40. Quando da integração, o pre-mRNA transcrito do locus de HBB contém de 5' a 3': éxon 1 de HBB endógeno; os primeiros 68 nucleotídeos de íntron 1, incluindo o doador de emenda endógeno na extremidade 5' de íntron 1; o aceitador de emenda introduzido pelo vetor hHBB-h1-010; o elemento de salto ribossômico; a sequência de codificação de HBB; e uma cauda poly(A). Após emenda, o mRNA contém de 5' a 3': éxon 1 de HBB endógeno; o elemento de salto ribossômico em quadro; a sequência de codificação de HBB; e uma cauda poly(A). O elemento de salto ribossômico leva à geração de dois polipeptídeos: um peptídeo HBB truncado terminado no final de éxon 1 fusionado com um peptídeo de salto ribossômico parcial, e uma prolina do peptídeo de salto ribossômico fusionado ao N-terminal de um polipeptídeo de HBB de comprimento completo.

[00194]O vetor hHBB-h1-010 compreende adicionalmente um braço de homologia 5' (referido como um "HBB HAL" em Figura 7C) compreendendo a sequência genômica a montante de tipo selvagem do sítio de inserção, e um braço de homologia 3' (referido como "HBB HAR" em Figura 7C) compreendendo a sequência genômica a jusante de tipo selvagem do sítio de

inserção. Cada um do braço de homologia 5' e do braço de homologia 3' tem cerca de 800 bp de comprimento. O vetor hHBB-h1-010 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 41, e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19). O braço de homologia 5' tem a capacidade para corrigir mutações no 5' UTR que afetam a expressão de HBB como observado em alguns pacientes com beta talassemia, e restauram a expressão de HBB tipo selvagem que foi conferida por mutações em 5' UTR, sequência de codificação, ou 3' UTR.

d) Vetor de correção de HBB hHBB-h1W-008

[00195]Vetor de correção de HBB hHBB-h1W-008, como mostrado em Figura 7D, contém os mesmos elementos genéticos como em vetor de correção de HBB hHBB-h1-010, exceto que a sequência de codificação de HBB é alterada silenciosamente para ser 67% idêntica à correspondente região da sequência de cDNA de tipo selvagem. Como descrito em Exemplo 3, esta sequência modificação não é esperada alterar de modo significativo o nível de expressão de HBB. Ao contrário, ela reduz a homologia de éxons de HBB com outros genes e pseudogenes de globina, assim reduzindo o indesejado deste vetor para outros *loci* genômicos. Em um exemplo específico, a sequência de codificação de HBB alterada silenciosamente é especificada em SEQ ID NO: 47. O vetor hHBB-h1W-008 compreende a sequência de nucleotídeo especificada em SEQ ID NO: 42, e compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19).

e) Vetor de correção de HBB hHBB-hE3C-001

[00196]O vetor de correção de HBB hHBB-hE3C-001, como mostrado em Figura 7E, é planejado para inserir uma sequência de codificação de HBB em éxon 3 do gene HBB, diretamente após o códon de parada, por recombinação homóloga. O elemento de edição (região inserida) deste vetor contém de 5' a 3' um elemento de salto ribossômico ("P2A" em Figura 7E, por exemplo, SEQ ID NO: 74) em quadro com uma sequência de codificação de HBB alterada silenciosamente (SEQ ID NO: 99, 85% idêntica à sequência de codificação de HBB de tipo selvagem), e uma sequência de poliadenilação SV40 (SEQ ID NO: 77). A alteração silenciosa da sequência de codificação de HBB é planejada para aumentar o nível de proteína expressada a partir da sequência de codificação, remover sequências de baixa complexidade que poderiam levar a uma marcação fora do alvo do vetor para *loci* genômico indesejado, e/ou reduzir a homologia entre o elemento de edição e o genoma, assim reduzindo a integração indesejada mediada pelo elemento de edição em vez de por um braço de homologia.

[00197]Quando da integração, o mRNA transcrito do *locus* de HBB contém de 5' a 3': uma porção do mRNA nativo de HBB adjacientemente 5' para o códon de parada, o elemento de salto ribossômico, a sequência de codificação de HBB alterada silenciosamente, e a sequência de poliadenilação de SV40. O elemento de salto ribossômico leva à geração de dois polipeptídeos: um peptídeo de HBB de comprimento completo nativo, fusionado com uma parte N-terminal do peptídeo de salto ribossômico, e um resíduo de prolina do peptídeo de salto ribossômico fusionado ao N-terminal de um

de polipeptídeo de HBB de comprimento completo de tipo selvagem.

[00198]O vetor hHBB-hE3C-001 vetor compreende adicionalmente um braço de homologia 5' (referido como "HAL" em Figura 7E) compreendendo a sequência de SEQ ID NO: 101, uma sequência genômica de tipo selvagem a montante do sítio de inserção, e um 3' braço de homologia (referido como "HAR" em Figura 7E) compreendendo a sequência de SEQ ID NO: 102, uma sequência genômica de tipo selvagem a jusante do sítio de inserção.

[00199]A sequência de nucleotídeos de hHBB-hE3C-001 é especificada em SEQ ID NO: 104. O vetor compreende adicionalmente um 5' ITR (por exemplo, tendo a sequência de SEQ ID NO: 18) e um 3' ITR (por exemplo, tendo a sequência de SEQ ID NO: 19).

[00200]A integração eficiência de hHBB-h1-010 e hHBB-h1W-008 foi avaliada em células RKO e LCL. As células GM16265 LCL foram transduzidas com vetores de correção de HBB empacotados em AAVHSC7 usando o método descrito em Exemplo 1.

[00201]As células RKO foram obtidas de ATCC. As células foram cultivadas em DMEM suplementado com 10% FCS e 2mM L-glutamina. Elas foram colocadas em placas em uma densidade de 750.000 células por cavidade em uma placa de 6 cavidades. As células foram transfectadas usando o seguinte método: 24 horas após colocação em placas, as células foram transfectadas em meio OptiMEM por adição de uma mistura de transfecção, que foi preparada por mistura e incubação de (a) 2 µg de plasmídeo editando HBB diluído em 250 µl OptiMEM e (b) 5 µl Lipofectamine 2000 diluído em 250 µl

OptiMEM durante 15 minutos. As células foram coletadas 24 horas após a transfecção.

[00202]Integração foi avaliada por um ensaio TI usando os iniciadores tendo as sequências especificadas em Tabela 9. Os iniciadores SA-2A-FM1 e SA-2A-FM2 foram específicos para o aceitador de emenda e elementos T2A inseridos no genoma pelo vetor hHBB-h1-010 ou hHBB-h1W-008, e o iniciador HBB-Out-RM2 foi específico para uma região do genoma a jusante do braço de homologia 3'. Os pares de iniciador de HBB-Out-RM2 com SA-2A-FM1 ou SA-2A-FM2 não amplificam um produto em células não transduzidas ou um produto a partir do vetor de correção sozinho. A reação de PCR usando SA-2A-FM1 e HBB-Out-RM2 poderia gerar um amplicon de 1.881 bp se o vetor hHBB-h1-010 ou hHBB-h1W-008 fosse integrado por recombinação homóloga através dos braços de homologia 5' e 3', e poderia gerar um amplicon de 1.188 bp se o vetor fosse integrado por recombinação homóloga através do braço de homologia 5' e a sequência de éxon 2 no elemento de edição. A reação de PCR usando SA-2A-FM2 e HBB-Out-RM2 também poderia gerar tamanhos diferentes de amplicon a partir destes dois modos de integração.

[00203]A reação de PCR foi ajustada como a seguir: até 50 µl de água para PCR; 5 µl de 10X Tampão PCR; 1 µl de 10mM dNTPs; 1 µl de 50mM MgCl₂; 10 µl de 5X Reagente Q; 2,5 µl de Iniciador Dianteiro TI (concentração 5 µM); 2,5 µl de Iniciador Reverso TI (concentração 5 µM); 100 ng DNA genômico; e 0,5 µl de HotStarTaq Polymerase. A máquina de PCR foi ajustada como a seguir: desnaturar inicial a 95°C durante 15 minutos; 40 ciclos de desnaturar a 94°C durante 10 segundos, anelamento a 58°C durante 30 segundos, e

extensão a 72°C durante 3 minutos; e uma extensão final a 68°C durante 10 minutos. Os produtos PCR foram analisados por eletroforese de gel.

Tabela 9: Iniciadores para ensaio de integração alvo-marcada de HBB

Nome Iniciador	SEQ ID NO:	Sequência de Nucleotídeo
SA-2A-FM1	96	GCTTCTGACCTCTTCTCTTCCTCCC
SA-2A-FM2	97	GCGGTGACGTGGAGGAGAATC
HBB-Out-RM2	98	GCAGAATGGTAGCTGGATTGTAGC

[00204] Como mostrado em Figura 8, usando os iniciadores SA-2A-FM1 e HBB-Out-RM2, o produto PCR de integração alvo-marcada tendo um comprimento de 1.874 nucleotídeos foi detectado em células RKO transduzidas com hHBB-h1W-002, indicando a integração com sucesso do vetor no modo desejado. Em contraste, um produto de PCR encurtado, gerado por recombinação de éxon 2 n elemento de edição (em vez de no braço de homologia 3') com o *locus* de HBB, foi detectado em células transduzidas RKO e LCL com hHBB-h1-010. Similares resultados foram obtidos de PCR usando os iniciadores SA-2A-FM2 e HBB-Out-RM2. Este resultado indica que a alteração de códons silenciosa no elemento de edição reduz ou elimina a recombinação indesejada, assim assegurando uma edição precisa do gene HBB.

Exemplo 5: Correção *in vivo* de mutações HBB

[00205] Este exemplo oferece um modelo animal para examinar os vetores de correção de HBB, tal como os descritos nos exemplos acima. Um camundongo NSG tendo o genótipo de NOD.*Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ* foi subletalmente irradiado e sofreu reconstituição hematopoiética com transplante com HSCs CD34⁺ humanas primárias de tipo selvagem. Os níveis de enxerto de HSCs CD34⁺ nos camundongos NSG foram determinados 12 semanas após

transplante por teste de sangue periférico pela presença de células CD45⁺ de humanos e murinos por citometria de fluxo. Camundongos tendo mais do que 25% de células humanas circulantes no sangue periférico foram usados para testar a eficiência de integração de vetores de AAV específicos para transduzir HSCs CD34⁺ humanas primárias *in vivo*.

[00206]O vetor AAVS1-FP foi empacotado em AAVHSC7 e AAVHSC17, e as partículas virais foram administradas ao camundongo NSG reconstituído intravenosamente em uma dose entre $1,22 \times 10^{13}$ e $1,54 \times 10^{13}$ genomas de vetor por kg. Amostras de sangue, medula óssea, e baço foram coletadas 4 semanas após a administração. DNA foi purificado a partir das amostras pelo método de extração de fenol/clorofórmio conhecido na técnica, e o DNA extraído foi analisado por ddPCR usando o método como descrito em Exemplo 1.

[00207]Dados dos grupos AAVHSC7 e AAVHSC17 foram reunidos. Como mostrado em Figura 9, nos camundongos NSG transplantados administrados com vetor AAVS1-FP, a frequência alélica de integração foi cerca de 3% no sangue e cerca de 1% na medula óssea. Este resultado sugeriu que os capsídeos de AAVHSC7 e AAVHSC17 entregam de modo eficiente um vetor para integração no *locus* AAVS1 e podem ser potencialmente usados para a entrega de vetores terapêuticos de edição de HBB

[00208]Um modelo animal modificado reconstituído com HSCs CD34⁺ humanas primárias defeituosas de HBB é utilizável para testar correção de mutações de HBB usando um vetor de correção ou vetor de correção, como descrito nos exemplos acima. Por exemplo, o vetor de correção empacotado em um capsídeo de AVV Clade F, como um capsídeo

AAVHSC7, AAVHSC15 ou AAVHSC17, pode ser administrado ao animal reconstituído. A eficiência da integração pode ser medida coletando amostras de sangue ou medula óssea e quantificando a porcentagem de células em que a recombinação homóloga desejada tinha ocorrido em uma população ampla ou em um tipo específico de células, como progenitores de eritrócitos.

[00209]O animal reconstituído com HSCs CD34⁺ humanas primárias defeituosas em HBB é esperado manifestar hemoglobinopatia devido à falta de um gene HBB e pode ser usado para determinar a eficácia e segurança de vetores de correção de HBB empacotados em vários capsídeos de AAV. Eficácia é avaliada através da medição da contagem de reticulócitos, contagens completas de sangue (CBCs), esfregaços de sangue e integração alvo-marcada da sequência de vetor. Segurança é avaliada medindo-se os níveis de transaminases hepáticas, como aspartato transaminase (AST) e alanina transaminase (ALT).

[00210]Este modelo também pode ser usado para avaliar a longevidade de correção de HBB após cada administração, assim otimizando o regime de dosagem.

[00211]A invenção não deve ser limitada em escopo pelas modalidades específicas aqui descritas. De fato, várias modificações da invenção além das descritas serão evidentes para os versados na técnica a partir da descrição acima e das figuras em anexo. Tais modificações são destinadas a estar dentro do escopo das reivindicações em anexo.

[00212]Todas as referências (por exemplo, publicações ou patentes ou pedidos de patentes) citadas aqui são incorporadas aqui por referência em sua totalidade e para

todos os fins na mesma extensão como se cada referência individual (por exemplo, publicação ou patente ou pedido de patente) fosse especificamente e individualmente indicada para ser incorporada por referência em sua totalidade para todos os fins. Outras modalidades estão dentro das seguintes reivindicações.

REIVINDICAÇÕES

1. Método para corrigir uma mutação em um gene de beta globina (HBB) em uma célula, o método **caracterizado** pelo fato de que compreende transduzir a célula com um vírus adeno-associado defeituoso na replicação (AAV) compreendendo:

a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e

b) um genoma de correção compreendendo: (i) um elemento de edição para editar um *locus* alvo em um gene alvo; (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo,

em que a célula é transduzida sem co-transdução ou co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

2. Método, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a célula é uma célula-tronco pluripotente.

3. Método, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a célula é uma célula-tronco hematopoiética.

4. Método, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a célula é uma célula-tronco hematopoiética CD34⁺.

5. Método, de acordo com qualquer uma das reivindicações 1-4, **caracterizado** pelo fato de que a célula

é de um indivíduo mamífero e o AAV é administrado ao indivíduo em uma quantidade eficaz para transduzir a célula no indivíduo.

6. Método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método **caracterizado** pelo fato de que compreende:

a) transduzir uma célula progenitora de eritrócitos a partir do indivíduo *ex vivo* com um AAV defeituoso na replicação compreendendo:

(i) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e

(ii) um genoma de correção compreendendo: (A) um elemento de edição para editar um *locus* alvo em um gene alvo; (B) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (C) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo, assim obtendo uma célula transduzida tendo um gene HBB corrigido, e;

b) administrar a célula transduzida para o indivíduo, em que a célula é transduzida sem co-transdução de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

7. Método, de acordo com a reivindicação 6, **caracterizado** pelo fato de que a célula progenitora de eritrócitos é uma célula-tronco pluripotente.

8. Método, de acordo com a reivindicação 6, **caracterizado** pelo fato de que a célula progenitora de eritrócitos é uma célula-tronco hematopoiética.

9. Método, de acordo com a reivindicação 6, **caracterizado** pelo fato de que a célula progenitora de eritrócitos é célula-tronco hematopoiética CD34+.

10. Método para tratamento de um indivíduo tendo uma doença ou distúrbio associado com uma mutação do gene HBB, o método **caracterizado** pelo fato de que compreende administrar ao indivíduo uma quantidade eficaz de um AAV defeituoso na replicação compreendendo:

a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e

b) um genoma de correção compreendendo: (i) um elemento de edição para editar um *locus* alvo em um gene alvo; (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo; e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo,

sem co-administração de uma nuclease exógena ou uma sequência de nucleotídeo que codifica uma nuclease exógena.

11. Método, de acordo com qualquer uma das reivindicações 6-10, **caracterizado** pelo fato de que a doença ou distúrbio é talassemia ou doença de célula falciforme.

12. Método, de acordo com qualquer uma das reivindicações 6-10, **caracterizado** pelo fato de que o indivíduo é um indivíduo humano.

13. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o gene alvo é o gene HBB.

14. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o *locus* alvo está em uma mutação de ponto de nucleotídeo, inserção ou deleção no gene HBB.

15. Método, de acordo com a reivindicação 14, **caracterizado** pelo fato de que a mutação de ponto de nucleotídeo ou deleção no gene HBB é selecionada dentre o grupo consistindo em G em posição -87, G em posição -31, A em posição -30, G em posição -29, G em posição -28, T em posição -10, C em posição 1, A em posição 1, G em posição 2, deleção de C e T em posições 17 e 18, A em posição 19, deleção de A em posição 20, T em posição 20, deleção de A e A em posições 25 e 26, adição de G após posição 26, A em posição 47, A em posição 48, deleção de C em posição 51, A em posição 52, G em posição 58, G em posição 59, A em posição 79, T em posição 82, adição de C após posição 84, T em posição 93, A em posição 93, C em posição 97, C em posição 98, G em posição 202, G em posição 208, C em posição 222, deleção de T em posição 241 ou 242, deleção de T e T e C e T em posições 254 a 257, T em posição 260, deleção de C em posição 264 ou 265, adição de A após posição 343, deleção de G e T em posições 399 e 400, T em posição 401, adição de A após posição 417, A em posição 446, T em posição 1099, A em posição 1293, T em posição 1344.

16. Método, de acordo com a reivindicação 14 ou 15, **caracterizado** pelo fato de que o elemento de edição compreende uma porção do gene HBB de tipo selvagem que corresponde à mutação.

17. Método, de acordo com qualquer uma das

reivindicações 1-16, **caracterizado** pelo fato de que o elemento de edição compreende as regiões de codificação de um ou mais éxons de um gene HBB.

18. Método, de acordo com qualquer uma das reivindicações 1-16, **caracterizado** pelo fato de que o elemento de edição compreende uma porção de um gene HBB compreendendo a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3.

19. Método, de acordo com a reivindicação 17 ou 18, **caracterizado** pelo fato de que as regiões de codificação foram alteradas silenciosamente para serem menos que 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idênticas aos éxons correspondentes do gene HBB de tipo selvagem.

20. Método, de acordo com a reivindicação 19, **caracterizado** pelo fato de que o elemento de edição compreende pelo menos uma das sequências de nucleotídeos selecionadas dentre o grupo consistindo em SEQ ID NOs: 43-46 e 105-107.

21. Método, de acordo com qualquer uma das reivindicações 1-12, **caracterizado** pelo fato de que o *locus* alvo é AAVS1.

22. Método, de acordo com qualquer uma das reivindicações 1-16 e 21, **caracterizado** pelo fato de que o elemento de edição compreende uma sequência de nucleotídeo codificando SEQ ID NO: 48 ou uma porção da mesma.

23. Método, de acordo com a reivindicação 22, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste em nucleotídeos 4 a 444

de SEQ ID NO: 27.

24. Método, de acordo com a reivindicação 22, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 é alterada silenciosamente para ser menos que 70%, 75%, 80%, 85%, ou 90% idêntica a nucleotídeos 4 a 444 de SEQ ID NO: 27.

25. Método, de acordo com a reivindicação 24, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste na sequência de SEQ ID NO: 47 ou 100.

26. Método, de acordo com a reivindicação 22, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste em uma sequência de codificação inserida em um fragmento *stuffer* do gene HBB.

27. Método, de acordo com qualquer uma das reivindicações 22-26, **caracterizado** pelo fato de que o *locus* alvo é a ligação internucleotídeos entre nucleotídeo 3 e nucleotídeo 4 do gene alvo, pelo que a integração do elemento de edição no *locus* alvo resulta no *locus* alvo compreendendo uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* partindo com o códon de partida do gene alvo.

28. Método, de acordo com qualquer uma das reivindicações 22-26, **caracterizado** pelo fato de que o elemento de edição compreende uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* consistindo em 5' a 3' um códon de partida e a sequência de nucleotídeo codificando SEQ ID NO: 48, ou uma porção da sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

29. Método, de acordo com a reivindicação 28, **caracterizado** pelo fato de que o *locus* alvo está em um íntron do gene alvo, e em que o elemento de edição compreende 5' a 3' um sítio aceitador de emenda, um elemento de salto ribossômico, e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

30. Método, de acordo com a reivindicação 29, **caracterizado** pelo fato de que o *locus* alvo está em íntron 1 do gene HBB.

31. Método, de acordo com a reivindicação 28, **caracterizado** pelo fato de que o *locus* alvo está adjacientemente 3' para um nucleotídeo codificante do gene alvo, e em que o elemento de edição compreende 5' a 3' um elemento de salto ribossômico e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

32. Método, de acordo com a reivindicação 31, **caracterizado** pelo fato de que o *locus* alvo é o códon de parada de um gene alvo de tipo selvagem ou os nucleotídeos correspondentes de um gene alvo mutante.

33. Método, de acordo com qualquer uma das reivindicações 22-32, **caracterizado** pelo fato de que o elemento de edição compreende adicionalmente uma sequência 3' de poliadenilação exógena para a sequência de nucleotídeo codificando SEQ ID NO: 48.

34. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o elemento de edição compreende adicionalmente um sítio de endonuclease de restrição não presente no gene alvo.

35. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o elemento de edição compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28.

36. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a sequência de nucleotídeo de braço de homologia 5' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à primeira região genômica.

37. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a sequência de nucleotídeo de braço de homologia 3' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à segunda região genômica.

38. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a primeira região genômica está localizada em uma primeira janela de edição, e a segunda região genômica está localizada em uma segunda janela de edição.

39. Método, de acordo com a reivindicação 38, **caracterizado** pelo fato de que a primeira e segunda janelas de edição são diferentes.

40. Método, de acordo com a reivindicação 38, **caracterizado** pelo fato de que a primeira e segunda janelas de edição são iguais

41. Método, de acordo com qualquer uma das reivindicações 38-40, **caracterizado** pelo fato de que a primeira janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103.

42. Método, de acordo com qualquer uma das

reivindicações 38-41, **caracterizado** pelo fato de que a segunda janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103.

43. Método, de acordo com qualquer uma das reivindicações 38-42, **caracterizado** pelo fato de que a primeira região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101.

44. Método, de acordo com qualquer uma das reivindicações 38-43, **caracterizado** pelo fato de que a segunda região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102.

45. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o braço de homologia 5' consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101.

46. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o braço de homologia 3' consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102.

47. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que cada uma das sequências de nucleotídeos de braços de homologia 5' e 3' independentemente tem um comprimento de cerca de 100 a cerca de 2000 nucleotídeos.

48. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o genoma de correção compreende adicionalmente uma sequência de nucleotídeo 5' de repetição terminal invertida 5' (5' ITR) da sequência de nucleotídeo de braço de homologia 5', e uma sequência de nucleotídeo 3' de

repetição terminal invertida 3' (3' ITR) da sequência de nucleotídeo de braço de homologia 3'.

49. Método, de acordo com a reivindicação 48, **caracterizado** pelo fato de que a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:18, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:19.

50. Método, de acordo com a reivindicação 48, **caracterizado** pelo fato de que a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:20, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:21.

51. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42 e 104.

52. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na

proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

53. Método, de acordo com a reivindicação 52, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

54. Método, de acordo com a reivindicação 52, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17.

55. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

56. Método, de acordo com a reivindicação 55, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de

SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

57. Método, de acordo com a reivindicação 55, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17.

58. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T; o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 65 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 68 de SEQ ID NO: 2 é V; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o

aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

59. Método, de acordo com a reivindicação 58, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é Y;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S;

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(f) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo

correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(g) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(h) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(i) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

60. Método, de acordo com a reivindicação 58, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17.

61. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que a eficiência de integração do elemento de edição para o locus alvo é pelo menos 1% quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão.

62. Método, de acordo com qualquer uma das reivindicações precedentes, **caracterizado** pelo fato de que

a frequência alélica de integração do elemento de edição no *locus* alvo é pelo menos 0,5% quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão.

63. Vírus adeno-associado defeituoso na replicação (AAV), **caracterizado** pelo fato de que compreende:

a) um capsídeo de AAV compreendendo uma proteína de capsídeo de AAV Clade F; e

b) um genoma de correção compreendendo (i) um elemento de edição para editar um *locus* alvo em um gene alvo, (ii) uma sequência de nucleotídeo 5' do braço de homologia 5' do elemento de edição tendo homologia para uma primeira região genômica 5' para o *locus* alvo, e (iii) uma sequência de nucleotídeo 3' do braço de homologia 3' do elemento de edição tendo homologia para uma segunda região genômica 3' para o *locus* alvo.

64. AAV, de acordo com a reivindicação 63, **caracterizado** pelo fato de que o gene alvo é o gene HBB.

65. AAV, de acordo com a reivindicação 63 ou 64, **caracterizado** pelo fato de que o *locus* alvo está em uma mutação de ponto de nucleotídeo ou deleção no gene HBB.

66. AAV, de acordo com a reivindicação 65, **caracterizado** pelo fato de que a mutação de ponto de nucleotídeo ou deleção no gene HBB é selecionada dentre o grupo consistindo em G em posição -87, G em posição -31, A em posição -30, G em posição -29, G em posição -28, T em posição -10, C em posição 1, A em posição 1, G em posição 2, deleção de C e T em posições 17 e 18, A em posição 19, deleção de A em posição 20, T em posição 20, deleção de A e

A em posições 25 e 26, adição de G após posição 26, A em posição 47, A em posição 48, deleção de C em posição 51, A em posição 52, G em posição 58, G em posição 59, A em posição 79, T em posição 82, adição de C após posição 84, T em posição 93, A em posição 93, C em posição 97, C em posição 98, G em posição 202, G em posição 208, C em posição 222, deleção de T em posição 241 ou 242, deleção de T e T e C e T em posições 254 a 257, T em posição 260, deleção de C em posição 264 ou 265, adição de A após posição 343, deleção de G e T em posições 399 e 400, T em posição 401, adição de A após posição 417, A em posição 446, T em posição 1099, A em posição 1293, T em posição 1344.

67. AAV, de acordo com a reivindicação 65 ou 66, **caracterizado** pelo fato de que o elemento de edição compreende uma porção do gene HBB de tipo selvagem que corresponde à mutação.

68. AAV, de acordo com qualquer uma das reivindicações 63-67, **caracterizado** pelo fato de que o elemento de edição compreende as regiões de codificação de um ou mais éxons de um gene HBB.

69. AAV, de acordo com qualquer uma das reivindicações 63-68, **caracterizado** pelo fato de que o elemento de edição compreende uma porção de um gene HBB compreendendo a região de codificação de éxon 1, o íntron completo 1, o éxon completo 2, o íntron completo 2, e a região de codificação de éxon 3.

70. AAV, de acordo com a reivindicação 68 ou 69, **caracterizado** pelo fato de que as regiões de codificação foram alteradas silenciosamente para serem menos que 100%,

95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, ou 50% idênticas aos éxons correspondentes do gene HBB de tipo selvagem.

71. AAV, de acordo com a reivindicação 70, **caracterizado** pelo fato de que o elemento de edição compreende pelo menos uma das sequências de nucleotídeos selecionadas dentre o grupo consistindo em SEQ ID NOs: 43-46 e 105-107.

72. AAV, de acordo com a reivindicação 63, **caracterizado** pelo fato de que o *locus* alvo é AAVS1.

73. AAV, de acordo com qualquer uma das reivindicações 63-66 e 72, **caracterizado** pelo fato de que o elemento de edição compreende uma sequência de nucleotídeo codificando SEQ ID NO: 48 ou uma porção da mesma.

74. AAV, de acordo com a reivindicação 73, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste em nucleotídeos 4 a 444 de SEQ ID NO: 27.

75. AAV, de acordo com a reivindicação 73, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 é alterada silenciosamente para ser menos que 70%, 75%, 80%, 85%, ou 90% idêntica a nucleotídeos 4 a 444 de SEQ ID NO: 27.

76. AAV, de acordo com a reivindicação 75, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste na sequência de SEQ ID NO: 47 ou 100.

77. AAV, de acordo com a reivindicação 73, **caracterizado** pelo fato de que a sequência de nucleotídeo codificando SEQ ID NO: 48 consiste em uma sequência de

codificação inserida em um fragmento *stuffer* do gene HBB.

78. AAV, de acordo com qualquer uma das reivindicações 73-77, **caracterizado** pelo fato de que o *locus* alvo é a ligação internucleotídeos entre nucleotídeo 3 e nucleotídeo 4 do gene alvo, pelo que a integração do elemento de edição no *locus* alvo resulta no *locus* alvo compreendendo uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* partindo com o códon de partida do gene alvo.

79. AAV, de acordo com qualquer uma das reivindicações 73-77, **caracterizado** pelo fato de que o elemento de edição compreende uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer* consistindo em 5' a 3' um códon de partida e a sequência de nucleotídeo codificando SEQ ID NO: 48, ou uma porção do sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

80. AAV, de acordo com a reivindicação 79, **caracterizado** pelo fato de que o *locus* alvo está em um íntron do gene alvo, e em que o elemento de edição compreende 5' a 3' um sítio aceitador de emenda, um elemento de salto ribossômico, e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

81. AAV, de acordo com a reivindicação 80, **caracterizado** pelo fato de que o *locus* alvo está em íntron 1 do gene HBB.

82. AAV, de acordo com a reivindicação 79, **caracterizado** pelo fato de que o *locus* alvo está adjacientemente 3' para um nucleotídeo codificante do gene

alvo, e em que o elemento de edição compreende 5' a 3' um elemento de salto ribossômico e uma sequência de codificação de HBB ou sequência de codificação inserida em fragmento *stuffer*.

83. AAV, de acordo com a reivindicação 82, **caracterizado** pelo fato de que o *locus* alvo é o códon de parada de um gene alvo de tipo selvagem ou os nucleotídeos correspondentes de um gene alvo mutante.

84. AAV, de acordo com qualquer uma das reivindicações 73-83, **caracterizado** pelo fato de que o elemento de edição compreende adicionalmente uma sequência 3' de poliadenilação exógena à sequência de nucleotídeo codificando SEQ ID NO: 48.

85. AAV, de acordo com qualquer uma das reivindicações 63-84, **caracterizado** pelo fato de que o elemento de edição compreende adicionalmente um sítio de endonuclease de restrição não presente no gene alvo.

86. AAV, de acordo com qualquer uma das reivindicações 63-85, **caracterizado** pelo fato de que o elemento de edição compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 23-28.

87. AAV, de acordo com qualquer uma das reivindicações 63-86, **caracterizado** pelo fato de que a sequência de nucleotídeo de braço de homologia 5' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à primeira região genômica.

88. AAV, de acordo com qualquer uma das reivindicações 63-87, **caracterizado** pelo fato de que a sequência de nucleotídeo de braço de homologia 3' é pelo menos 90%, 95%, 96%, 97%, 98%, ou 99% idêntica à segunda região genômica.

89. AAV, de acordo com qualquer uma das reivindicações

62-88, **caracterizado** pelo fato de que a primeira região genômica está localizada em uma primeira janela de edição, e a segunda região genômica está localizada em uma segunda janela de edição.

90. AAV, de acordo com a reivindicação 89, **caracterizado** pelo fato de que a primeira e segunda janelas de edição são diferentes.

91. AAV, de acordo com a reivindicação 89, **caracterizado** pelo fato de que a primeira e segunda janelas de edição são iguais.

92. AAV, de acordo com qualquer uma das reivindicações 89-91, **caracterizado** pelo fato de que a primeira janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103.

93. AAV, de acordo com qualquer uma das reivindicações 89-92, **caracterizado** pelo fato de que a segunda janela de edição consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101, 102, ou 103.

94. AAV, de acordo com qualquer uma das reivindicações 89-93, **caracterizado** pelo fato de que a primeira região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101.

95. AAV, de acordo com qualquer uma das reivindicações 89-94, **caracterizado** pelo fato de que a segunda região genômica consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102.

96. AAV, de acordo com qualquer uma das reivindicações 63-95, **caracterizado** pelo fato de que o braço de homologia 5' consiste na sequência de nucleotídeo especificada em SEQ ID NO: 101.

97. AAV, de acordo com qualquer uma das reivindicações 63-96, **caracterizado** pelo fato de que o braço de homologia 3' consiste na sequência de nucleotídeo especificada em SEQ ID NO: 102.

98. AAV, de acordo com qualquer uma das reivindicações 63-97, **caracterizado** pelo fato de que cada uma das sequências de nucleotídeos de braços de homologia 5' e 3' independentemente tem um comprimento de cerca de 100 a cerca de 2000 nucleotídeos.

99. AAV, de acordo com qualquer uma das reivindicações 63-98, **caracterizado** pelo fato de que o genoma de correção compreende adicionalmente uma sequência de nucleotídeo 5' de repetição terminal invertida 5' (5' ITR) da sequência de nucleotídeo de braço de homologia 5', e uma sequência de nucleotídeo 3' de repetição terminal invertida 3' (3' ITR) da sequência de nucleotídeo de braço de homologia 3'.

100. AAV, de acordo com a reivindicação 99, **caracterizado** pelo fato de que a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:18, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:19.

101. AAV, de acordo com a reivindicação 99, **caracterizado** pelo fato de que a sequência de nucleotídeo 5' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:20, e a sequência de nucleotídeo 3' ITR tem pelo menos 95% de identidade de sequência para SEQ ID NO:21.

102. AAV, de acordo com qualquer uma das reivindicações 63-101, **caracterizado** pelo fato de que o genoma de correção compreende a sequência de nucleotídeo especificada em qualquer uma de SEQ ID NOs: 29-42 e 104.

103. AAV, de acordo com qualquer uma das reivindicações 63-102, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 203-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de

SEQ ID NO: 2 é G.

104. AAV, de acordo com a reivindicação 103, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

105. AAV, de acordo com a reivindicação 103, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 203-736

de SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, ou 17.

106. AAV, de acordo com qualquer uma das reivindicações 63-105, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na

proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

107. AAV, de acordo com a reivindicação 106, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo

correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

108. AAV, de acordo com a reivindicação 106, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 138-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, ou 17.

109. AAV, de acordo com qualquer uma das reivindicações 63-108, **caracterizado** pelo fato de que a proteína de capsídeo de AAV Clade F compreende uma sequência de aminoácido tendo pelo menos 95% de identidade de sequência com a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO:2, em que: o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 68 de SEQ ID NO: 2 é V; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 151 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 160 de SEQ ID NO: 2 é D; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 206 de SEQ ID NO: 2 é C; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A; o aminoácido na proteína de capsídeo correspondendo ao

aminoácido 464 de SEQ ID NO: 2 é N; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 590 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G ou Y; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K; o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C; ou, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G.

110. AAV, de acordo com a reivindicação 109, **caracterizado** pelo fato de que:

(a) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 2 de SEQ ID NO: 2 é T, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 312 de SEQ ID NO: 2 é Q;

(b) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 65 de SEQ ID NO: 2 é I, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é Y;

(c) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 77 de SEQ ID NO: 2 é R, e o aminoácido na

proteína de capsídeo correspondendo ao aminoácido 690 de SEQ ID NO: 2 é K;

(d) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 119 de SEQ ID NO: 2 é L, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 468 de SEQ ID NO: 2 é S;

(e) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 626 de SEQ ID NO: 2 é G, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 718 de SEQ ID NO: 2 é G;

(f) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 296 de SEQ ID NO: 2 é H, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 464 de SEQ ID NO: 2 é N, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 681 de SEQ ID NO: 2 é M;

(g) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 687 de SEQ ID NO: 2 é R;

(h) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 346 de SEQ ID NO: 2 é A, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R; ou

(i) o aminoácido na proteína de capsídeo correspondendo ao aminoácido 501 de SEQ ID NO: 2 é I, o aminoácido na proteína de capsídeo correspondendo ao aminoácido 505 de SEQ ID NO: 2 é R, e o aminoácido na proteína de capsídeo correspondendo ao aminoácido 706 de SEQ ID NO: 2 é C.

111. AAV, de acordo com a reivindicação 109, **caracterizado** pelo fato de que a proteína de capsídeo compreende a sequência de aminoácido de aminoácidos 1-736 de SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, ou 17.

112. AAV, de acordo com qualquer uma das reivindicações 63-111, **caracterizado** pelo fato de que a eficiência de integração do elemento de edição no *locus* alvo é pelo menos 1% quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão.

113. AAV, de acordo com qualquer uma das reivindicações 63-112, **caracterizado** pelo fato de que a frequência alélica de integração do elemento de edição no *locus* alvo é pelo menos 0,5% quando o AAV é contatado *in vitro* na ausência de uma nuclease exógena com uma população de células-tronco humanas CD34⁺ sob condições de transdução de AAV padrão.

114. Composição farmacêutica, **caracterizada** pelo fato de que compreende um AAV, conforme definido em qualquer uma das reivindicações 62-111.

115. Sistema de empacotamento para preparação recombinante de um AAV, **caracterizado** pelo fato de que o sistema de empacotamento compreende

a) uma sequência de nucleotídeo Rep codificando uma ou mais proteínas Rep de AAV;

b) uma sequência de nucleotídeo Cap codificando uma ou mais proteínas de capsídeo de AAV Clade F, como especificado com qualquer uma das reivindicações 103-111; e

c) um genoma de correção, como especificado com qualquer uma das reivindicações 63-102, em que o sistema de

empacotamento é operativo em uma célula para encerrar o genoma de correção no capsídeo para formar o AAV.

116. Sistema de empacotamento, de acordo com a reivindicação 115, **caracterizado** pelo fato de que o sistema de empacotamento compreende um primeiro vetor compreendendo a sequência de nucleotídeo Rep e a sequência de nucleotídeo Cap, e um segundo vetor compreendendo o genoma de correção.

117. Sistema de empacotamento, de acordo com a reivindicação 115 ou 116, **caracterizado** pelo fato de que a sequência de nucleotídeo Rep codifica uma proteína Rep de AAV2.

118. Sistema de empacotamento, de acordo com a reivindicação 117, **caracterizado** pelo fato de que a proteína Rep de AAV2 é 78/68 ou Rep 68/52.

119. Sistema de empacotamento, de acordo com a reivindicação 117 ou 118, **caracterizado** pelo fato de que a proteína Rep de AAV2 compreende uma sequência de aminoácido tendo uma identidade percentual mínima de sequência para a sequência de aminoácido Rep de AAV2 de SEQ ID NO:22, em que a identidade percentual mínima de sequência é pelo menos 70% através do comprimento da sequência de aminoácido codificando a proteína Rep de AAV2.

120. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 115-119, **caracterizado** pelo fato de que compreendendo adicionalmente um terceiro vetor, em que o terceiro vetor é um vetor de vírus auxiliar.

121. Sistema de empacotamento, de acordo com a reivindicação 120, **caracterizado** pelo fato de que o vetor de vírus auxiliar é um terceiro vetor independente.

122. Sistema de empacotamento, de acordo com a

reivindicação 120, **caracterizado** pelo fato de que o vetor de vírus auxiliar é integrante com o primeiro vetor.

123. Sistema de empacotamento, de acordo com a reivindicação 120, **caracterizado** pelo fato de que o vetor de vírus auxiliar é integrante com o segundo vetor.

124. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-123, **caracterizado** pelo fato de que o terceiro vetor compreende genes codificando proteínas de vírus auxiliar.

125. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-124 **caracterizado** pelo fato de que o vírus auxiliar é selecionado dentre o grupo consistindo em adenovírus, vírus de herpes, vírus de vaccinia, e citomegalovírus (CMV).

126. Sistema de empacotamento, de acordo com a reivindicação 125, **caracterizado** pelo fato de que o vírus auxiliar é adenovírus.

127. Sistema de empacotamento, de acordo com a reivindicação 126, **caracterizado** pelo fato de que o genoma do adenovírus compreende um ou mais genes de RNA de adenovírus selecionados dentre o grupo consistindo em E1, E2, E4 e VA.

128. Sistema de empacotamento, de acordo com a reivindicação 125, **caracterizado** pelo fato de que o vírus auxiliar é vírus herpes simplex (HSV).

129. Sistema de empacotamento, de acordo com a reivindicação 128, **caracterizado** pelo fato de que o genoma de HSV compreende um ou mais de genes de HSV selecionados dentre o grupo consistindo em UL5/8/52, ICPO, ICP4, ICP22 e UL30/UL42.

130. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-129, **caracterizado** pelo fato de que o primeiro vetor e o terceiro vetor estão contidos dentro de um primeiro plasmídeo transfectante.

131. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-129, **caracterizado** pelo fato de que os nucleotídeos do segundo vetor e do terceiro vetor estão contidos dentro de um segundo plasmídeo transfectante.

132. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-129, **caracterizado** pelo fato de que os nucleotídeos do primeiro vetor e do terceiro vetor são clonados em um vírus auxiliar recombinante.

133. Sistema de empacotamento, de acordo com qualquer uma das reivindicações 120-129, **caracterizado** pelo fato de que os nucleotídeos do segundo vetor e do terceiro vetor são clonados em um vírus auxiliar recombinante.

134. Método para preparação recombinante de um AAV, o método **caracterizado** pelo fato de que compreende introduzir o sistema de empacotamento, conforme definido em qualquer uma das reivindicações 115-133, em uma célula sob condições operativas para encerrar o genoma de correção no capsídeo para formar o AAV.

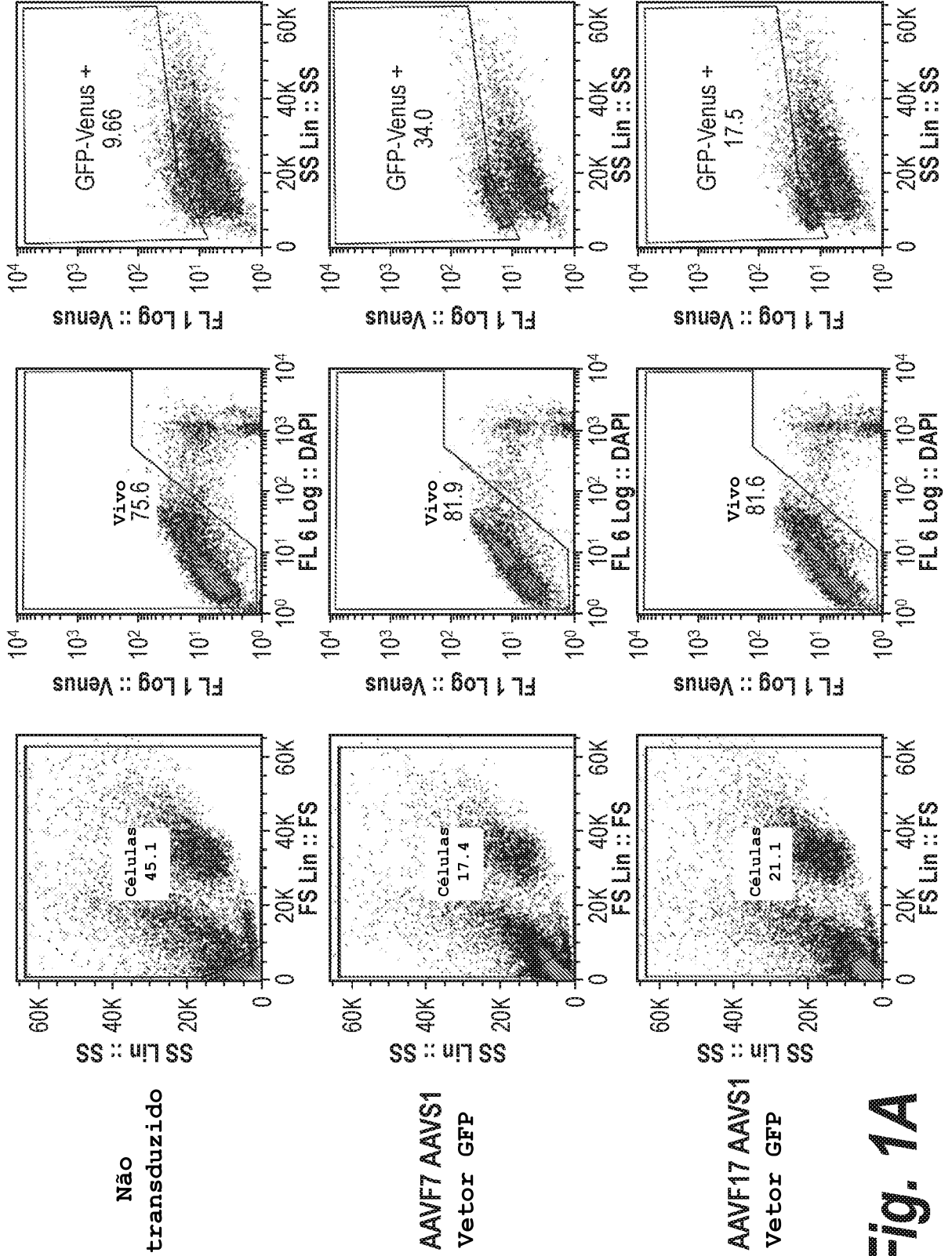


Fig. 1A

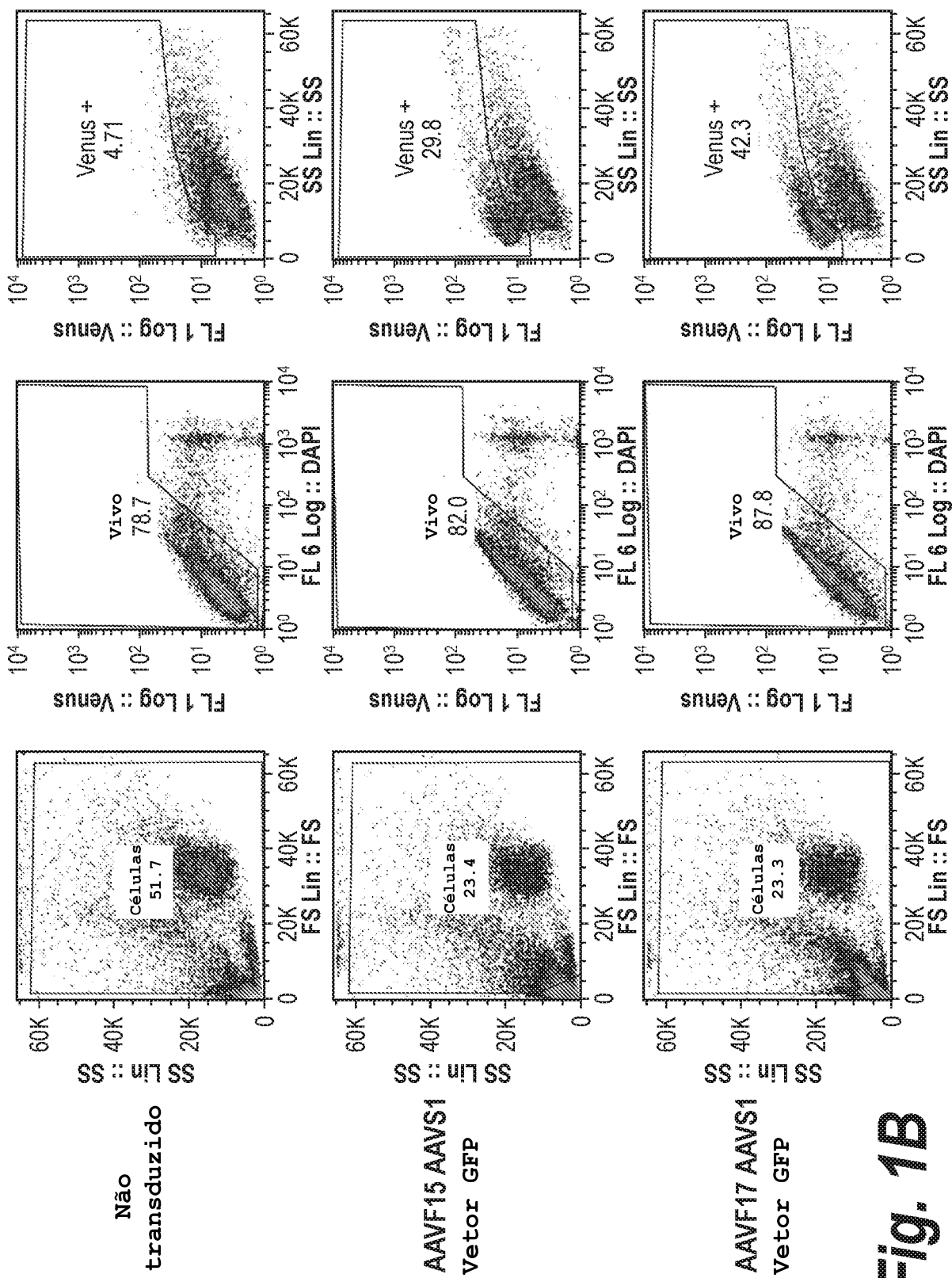


Fig. 1B

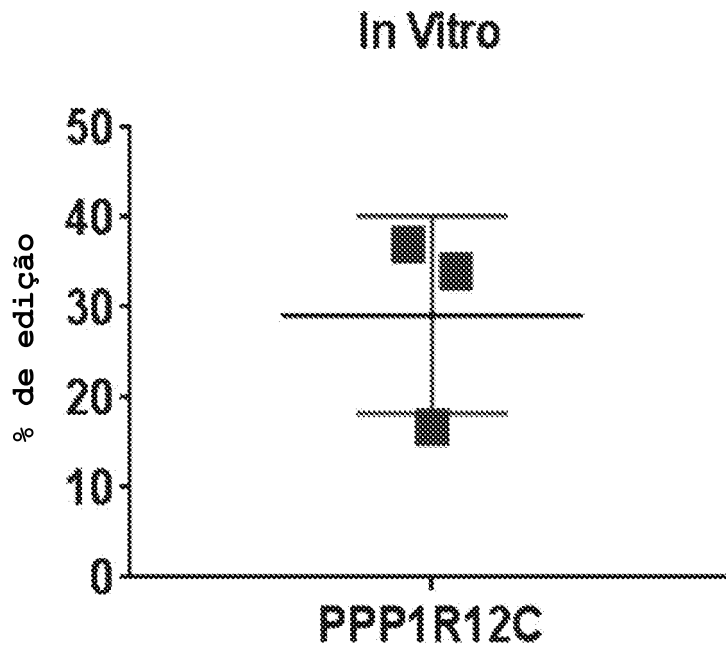


Fig. 1C

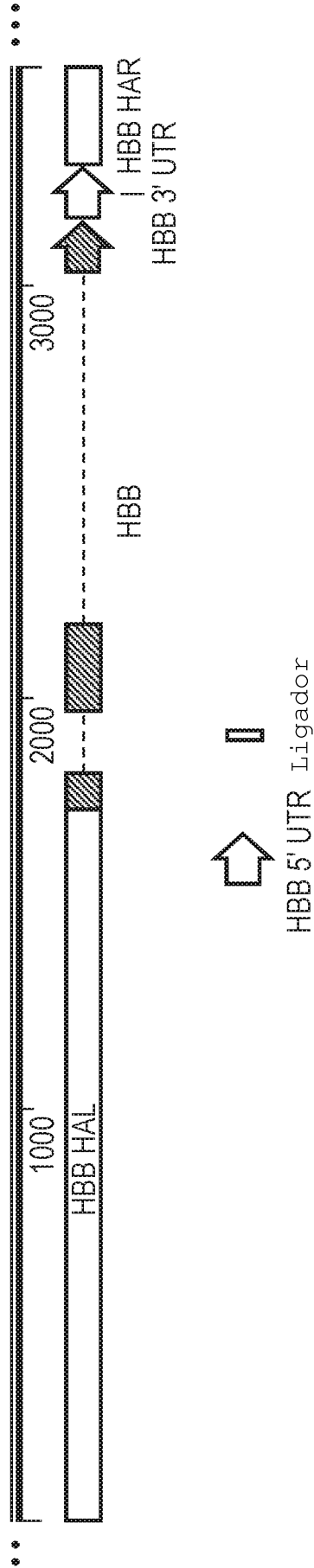
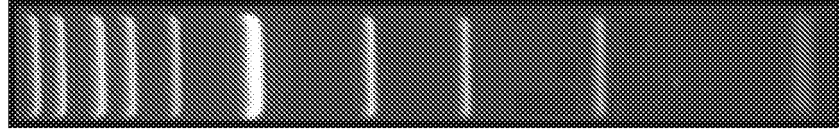
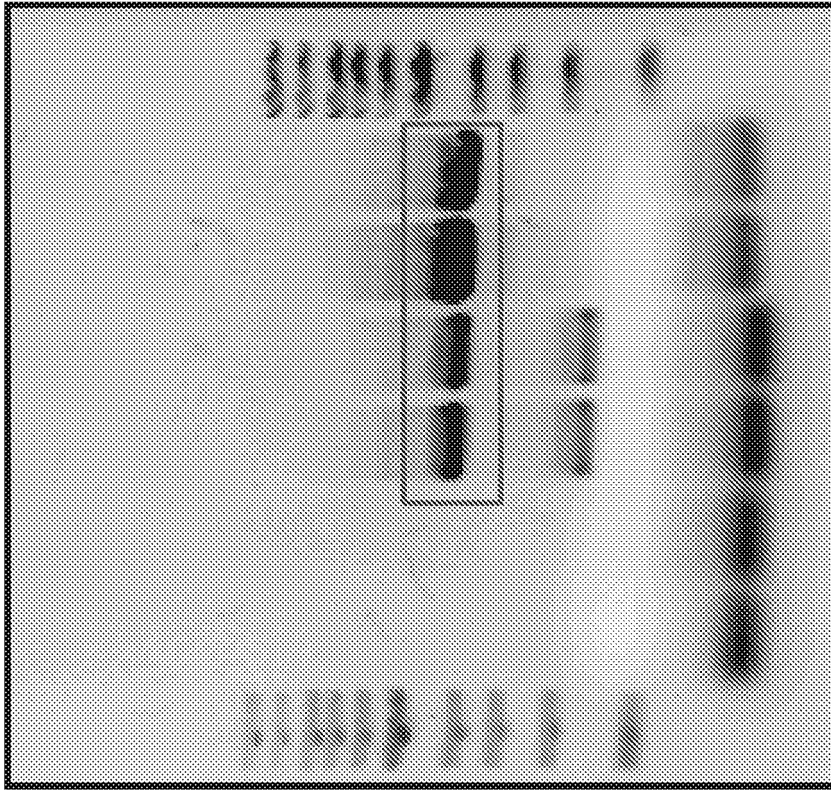


Fig. 2

Kilobases Massa (ng)

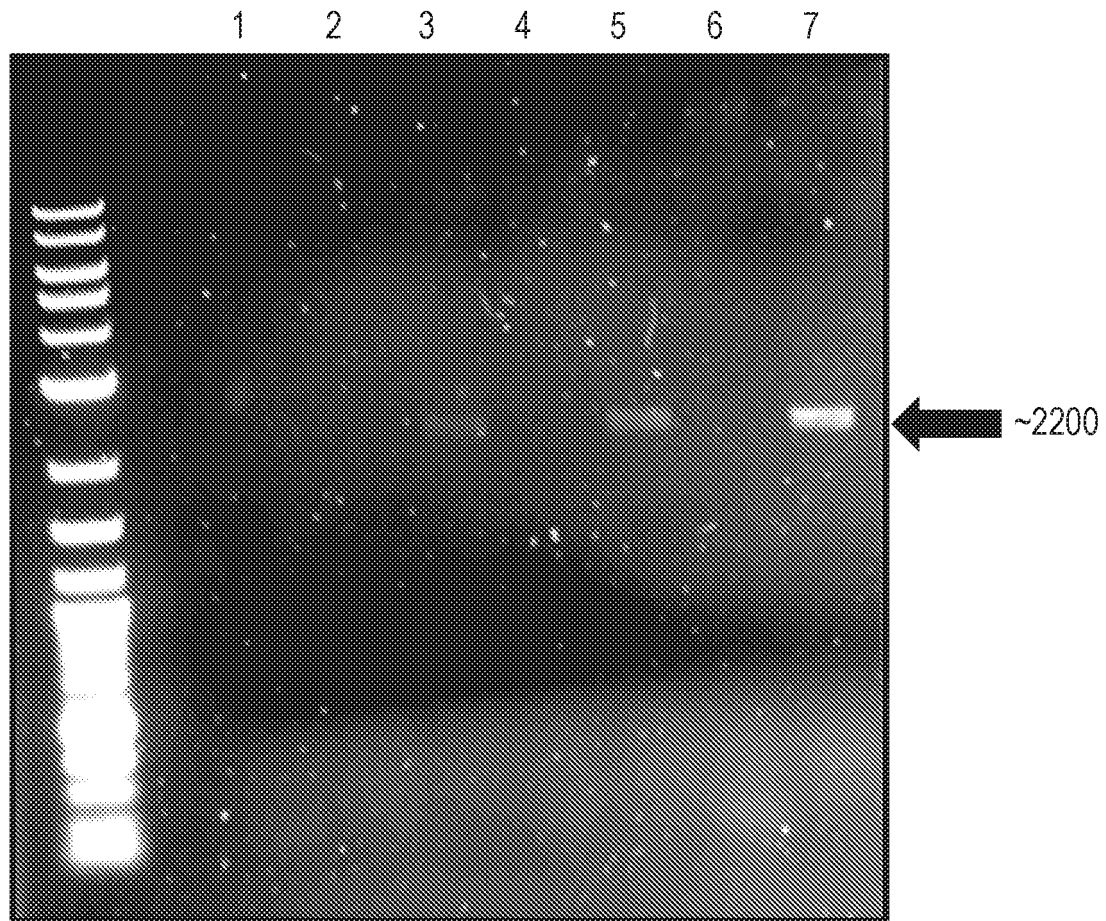


Banda
* de PCR*
~2200 bp



Não transduzido AAVF15 AAVF17
Ligador de HBB Ligador de HBB

Fig. 3A



Pistas:

1. Controle negativo
2. LCLs 16265 SCD não transduzidas
3. LCLs 16265 SCD transduzidas com vetor AAVHSC17-hHBB-ht-014
4. LCLs 16266 SCD não transduzidas
5. LCLs 16266 SCD transduzidas com vetor AAVHSC17-hHBB-hL-014
6. LCLs 16267 SCD não transduzidas
7. LCLs 16267 SCD transduzidas com vetor AAVHSC17-hHBB-hL-014

Fig. 3B

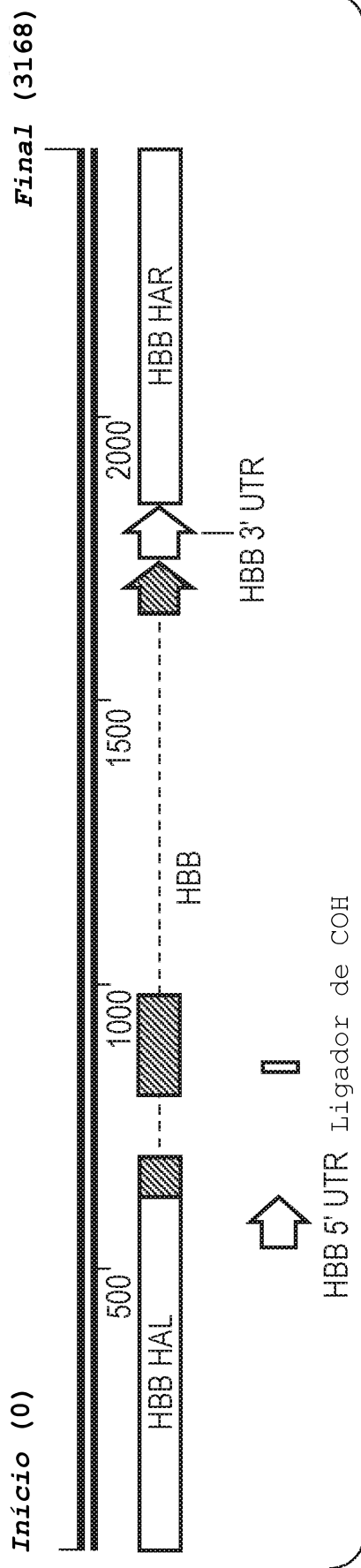


Fig. 4C

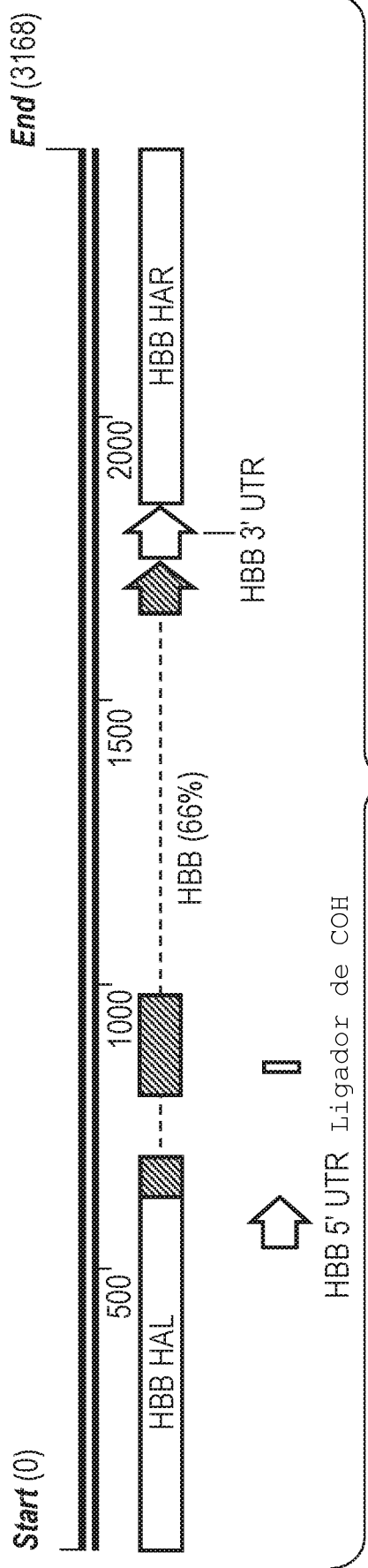


Fig. 4D

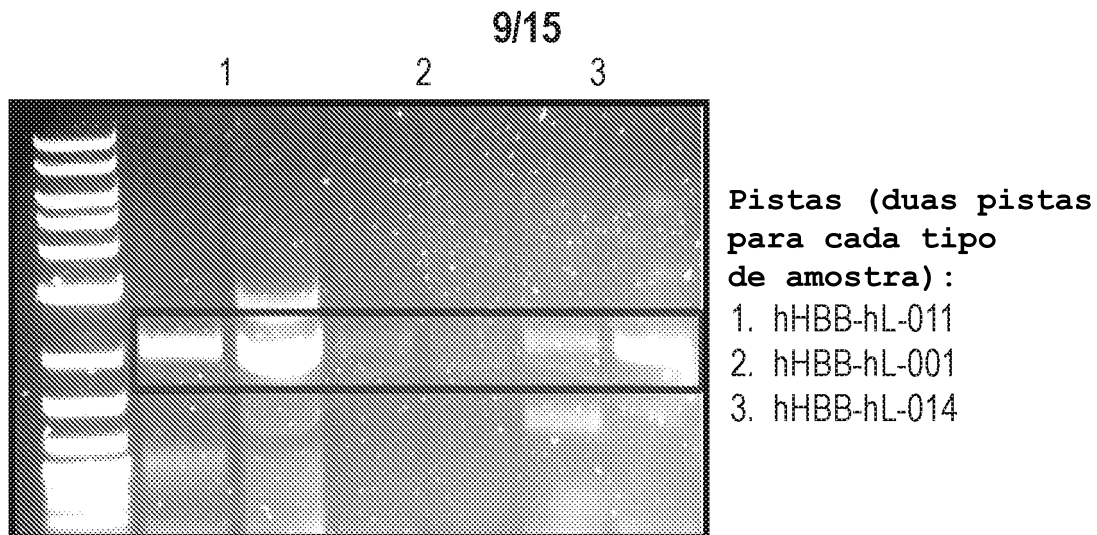


Fig. 5A

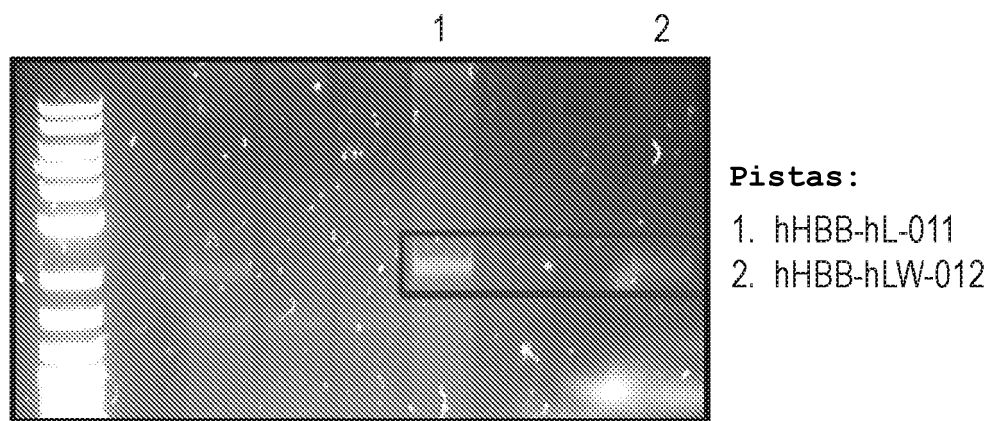


Fig. 5B

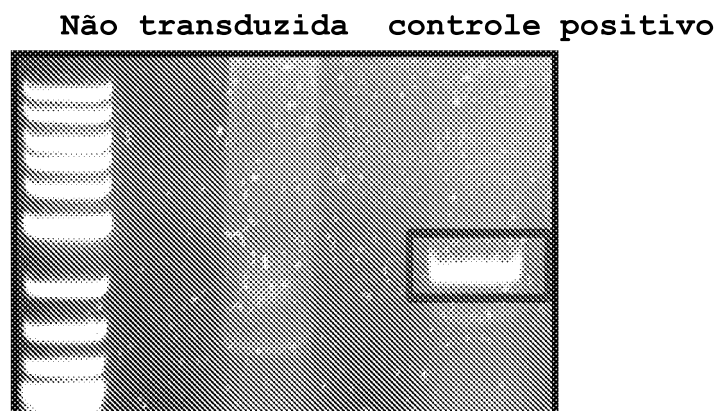


Fig. 5C

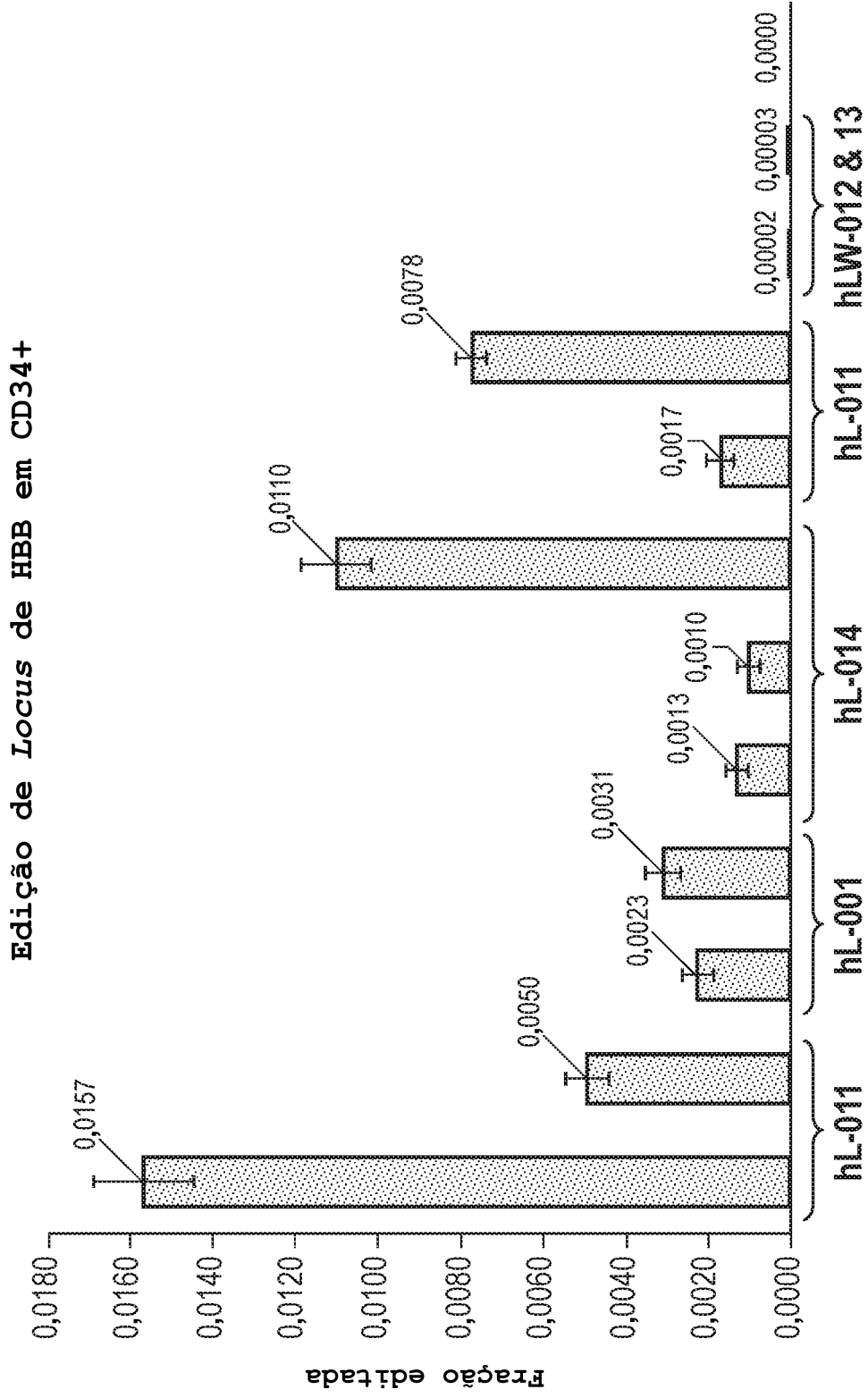


Fig. 6

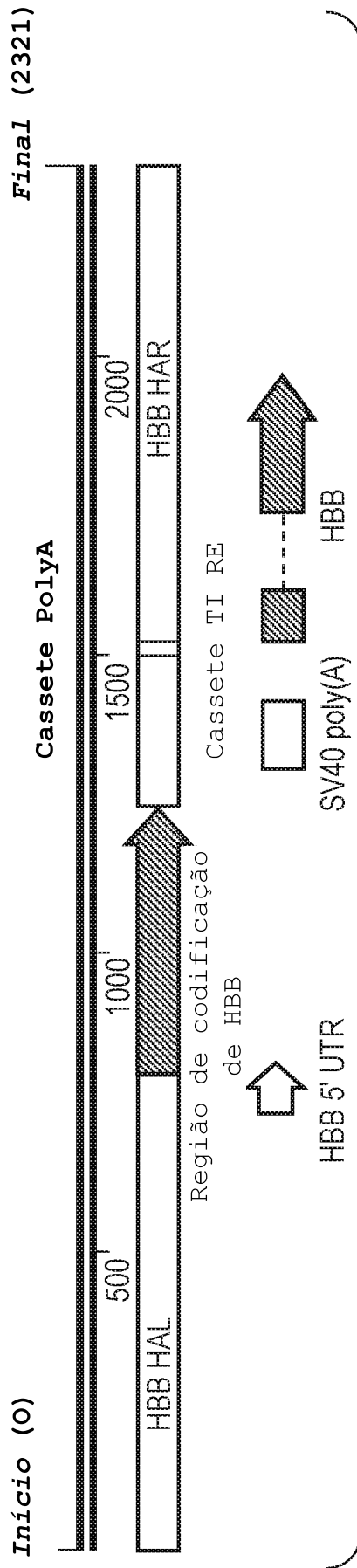


Fig. 7A

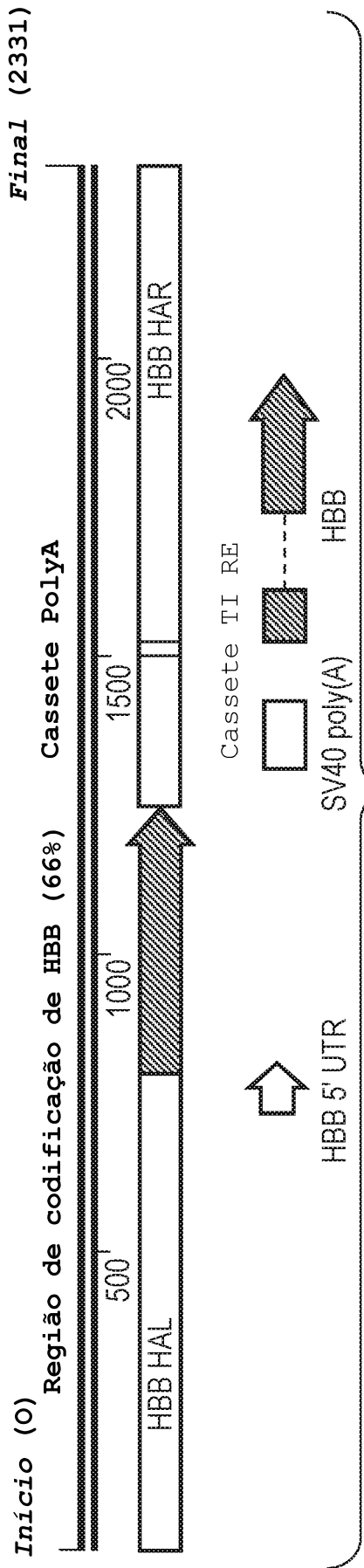


Fig. 7B

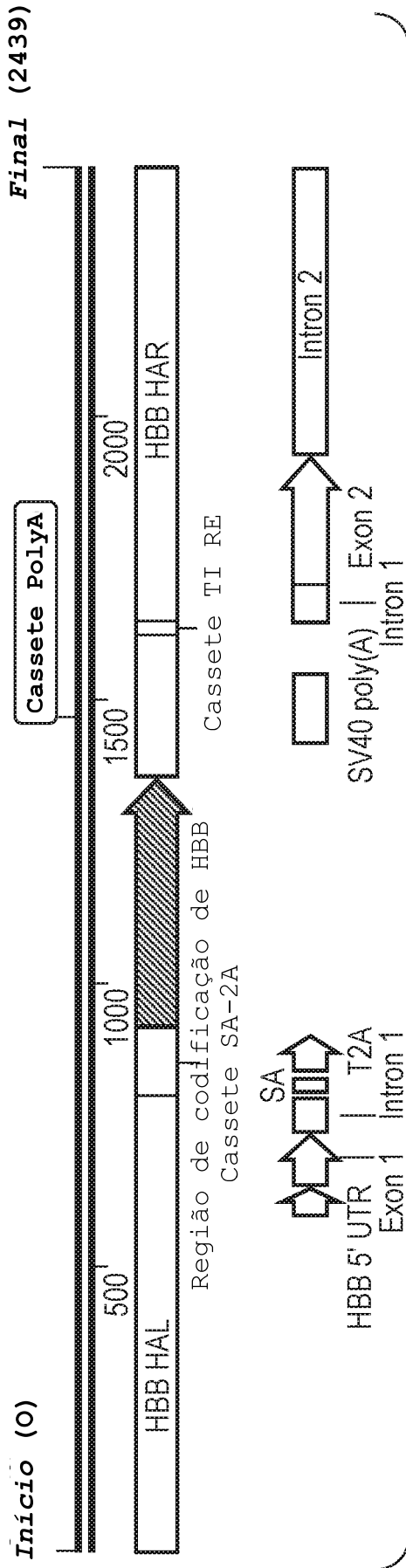


Fig. 7C

12/15

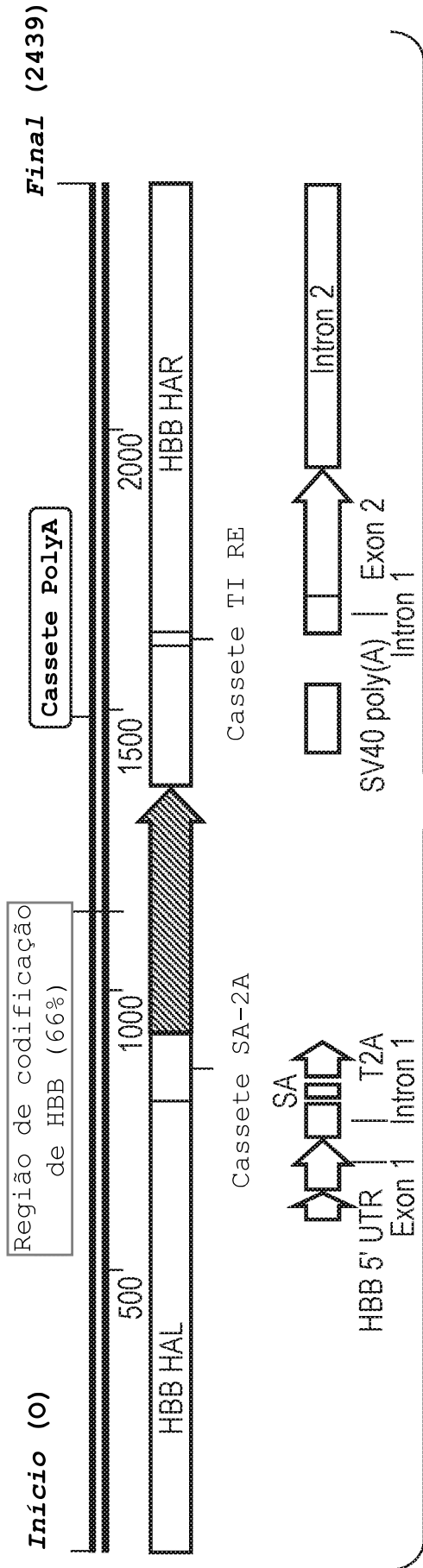


Fig. 7D

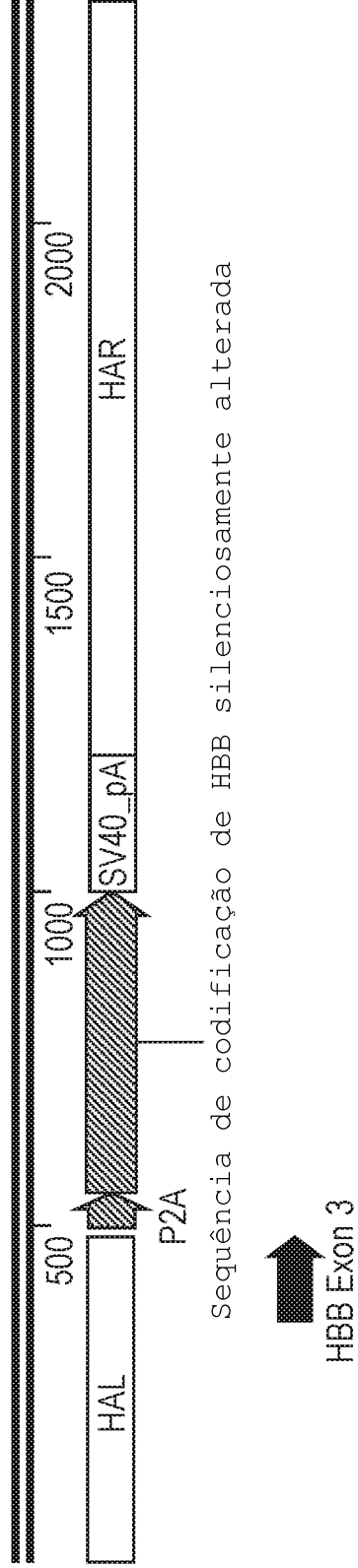
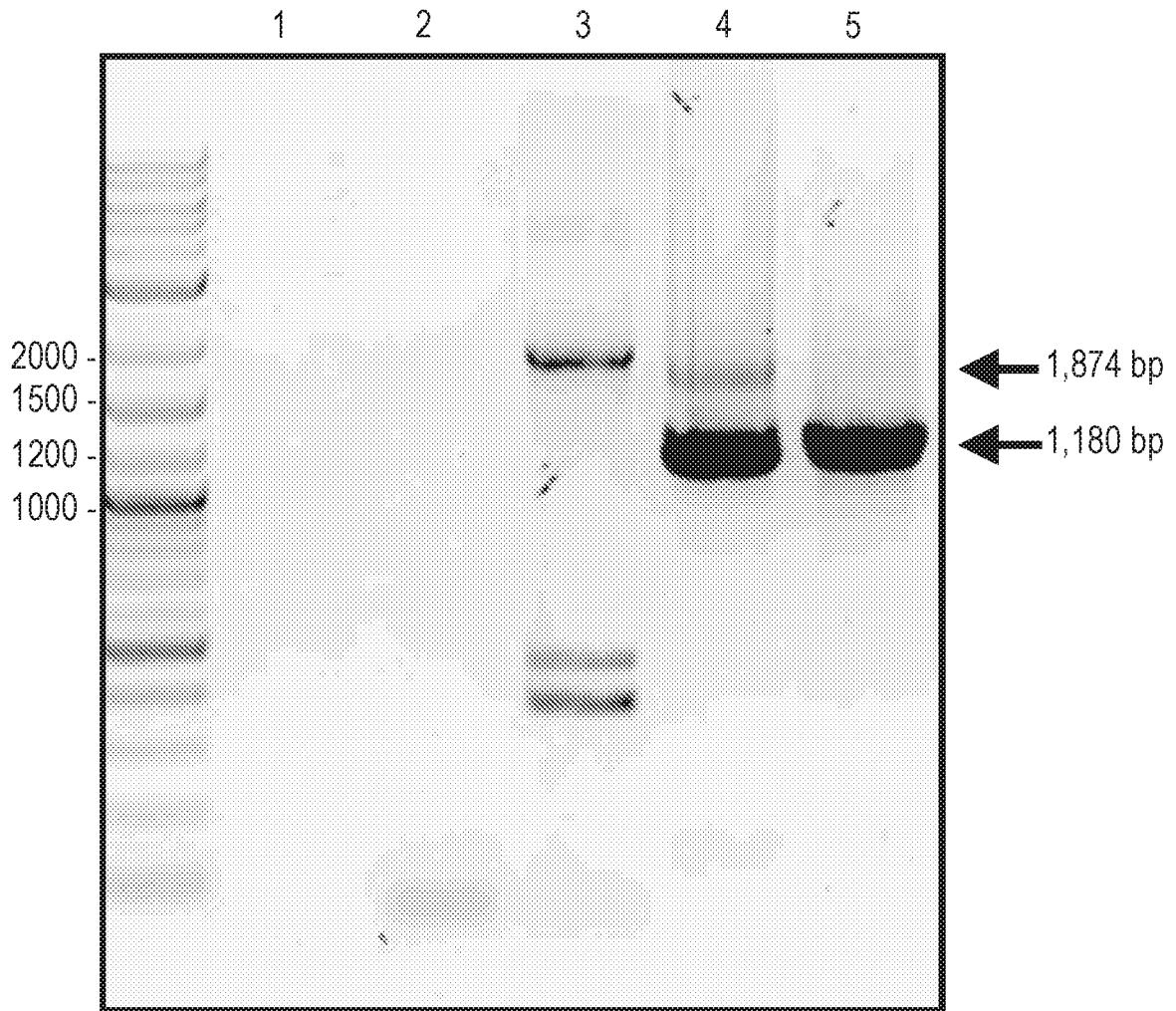


Fig. 7E



Pistas:

1. controle negativo
2. células não transduzidas
3. células RKO transduzidas com hHBB-h1W-002
4. células RKO transduzidas com hHBB-h1-010
5. células LCL transduzidas com hHBB-h1-010

Fig. 8

PPP1R12C

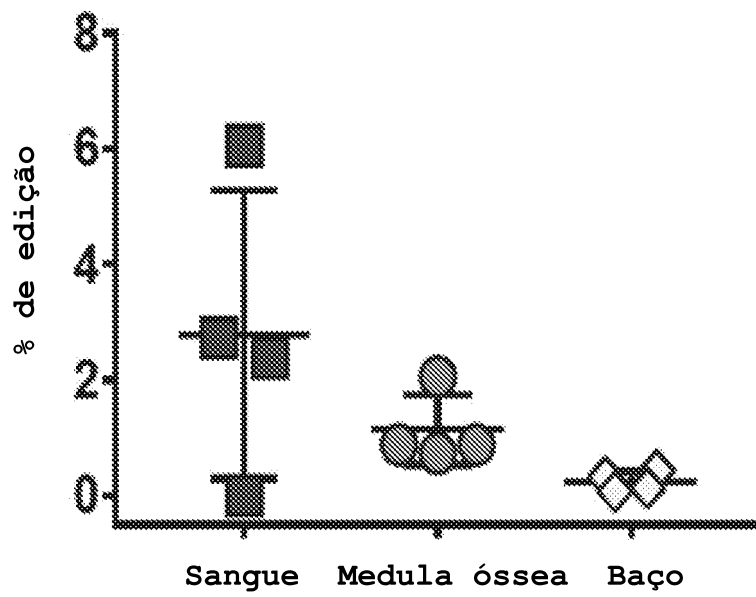


Fig. 9

RESUMO

**COMPOSIÇÕES DE VÍRUS ADENO-ASSOCIADOS PARA RESTAURAR A
FUNÇÃO DO GENE HBB E MÉTODOS DE USO DAS MESMAS**

São providas aqui composições de vírus adeno-associados (AAV) para corrigir uma mutação em um gene de beta globina (HBB) e métodos de usar os mesmos para corrigir uma mutação do gene HBB em uma célula. Também são providos sistemas de empacotamento para produzir composições de vírus adeno-associados.

Este anexo apresenta o código de controle da listagem de sequências biológicas.

Código de Controle

Campo 1



Campo 2



Outras Informações:

- Nome do Arquivo: 202000441 LISTAGEM.txt
- Data de Geração do Código: 17/04/2020
- Hora de Geração do Código: 17:28:50
- Código de Controle:
 - Campo 1: 90F3601B4E2A3236
 - Campo 2: 3DD1E942EA82A5D9