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

(21) BR 112020007502-7 A2

(22) Data do Depósito: 17/10/2018

(43) Data da Publicação Nacional: 06/10/2020

(54) Título: COMPOSIÇÕES E MÉTODOS PARA EDIÇÃO GÊNICA PARA HEMOFILIA A

(51) Int. Cl.: C12N 15/113; C12N 9/22; C07K 14/755; A61P 7/04.

(30) Prioridade Unionista: 17/10/2017 US 62/573,633.

(71) Depositante(es): CRISPR THERAPEUTICS AG; BAYER HEALTHCARE LLC.
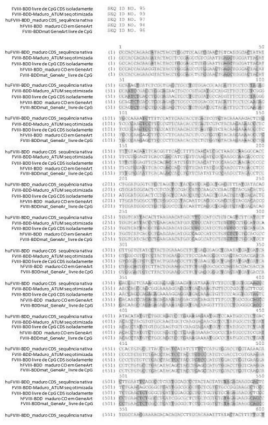
(72) Inventor(es): ALAN RICHARD BROOKS.

(86) Pedido PCT: PCT US2018056390 de 17/10/2018

(87) Publicação PCT: WO WO/2019/079527 de 25/04/2019

(85) Data da Fase Nacional: 15/04/2020

(57) Resumo: Fornecidos estão incluídos materiais e métodos para o tratamento da Hemofilia A em um sujeito ex vivo ou in vivo. Também fornecidos estão incluídos materiais e métodos para knocking em um gene que codifica FVIII em um genoma, em particular o lócus do gene da albumina.



Relatório Descritivo da Patente de Invenção para **"COMPOSIÇÕES E MÉTODOS PARA EDIÇÃO GÊNICA PARA HEMOFILIA A".**

REFERÊNCIA CRUZADA A PEDIDOS RELACIONADOS

[001] Este pedido reivindica o benefício de prioridade ao Pedido de Patente Provisório dos EUA No. 62/573.633, depositado em 17 de outubro de 2017, cuja invenção é incorporada no presente documento na sua totalidade por referência.

ÁREA

[002] As descrições fornecidas no presente documento referem-se a materiais e métodos para o tratamento de um paciente com Hemofilia A, ex vivo e in vivo. Além disso, as presentes invenções fornecem materiais e métodos para edição para modular a expressão, função ou atividade de uma proteína de coagulação do sangue, como o Fator VIII (FVIII) em uma célula por edição do genoma.

ANTECEDENTES

[003] A Hemofilia A (HemA) é causada por um defeito genético no gene do Fator VIII (FVIII) que resulta em níveis baixos ou indetectáveis da proteína FVIII no sangue. Isto resulta na formação ineficaz de coágulos nos locais de lesão do tecido, levando a sangramento descontrolado que pode ser fatal se não for tratado. A substituição da proteína FVIII ausente é um tratamento eficaz para pacientes com HemA e é o padrão atual de cuidados. No entanto, a terapia de reposição da proteína requer injeção intravenosa frequente da proteína FVIII o que é inconveniente em adultos, problemática em crianças, de custo proibitivo (> $200.000/ano) e pode resultar em interrupção por eventos hemorrágicos se o regime de tratamento não for seguido de perto.

[004] O gene FVIII é expresso principalmente nas células endoteliais sinusoidais que estão presentes no fígado, bem como em

outros locais do corpo. O FVIII exógeno pode ser expresso e secretado pelos hepatócitos do fígado, gerando FVIII na circulação e afetando assim a cura da doença. Foram desenvolvidos métodos de administração de genes direcionados aos hepatócitos do fígado e estes foram usados para fornecer um gene FVIII como tratamento para HemA em modelos animais e em pacientes em ensaios clínicos.

[005] Uma cura permanente para a Hemofilia A é altamente desejável. Embora a terapia gênica tradicional baseada em vírus usando Vírus Adeno-Associado (AAV) possa mostrar-se promissora em modelos animais pré-clínicos e em pacientes, ela tem várias desvantagens. A terapia gênica baseada em AAV usa um gene FVIII dirigido por um promotor específico do fígado que é encapsulado dentro de um capsídeo do vírus AAV (normalmente usando os sorotipos AAV5, AAV8 ou AAV9 ou AAVrh10, entre outros). Todos os vírus AAV usados para terapia gênica fornecem o cassete do gene empacotado no núcleo das células transduzidas, onde o cassete do gene permanece quase exclusivamente episssomal e são as cópias episssomais do gene terapêutico que dão origem à proteína terapêutica. O AAV não possui um mecanismo para integrar o seu DNA encapsulado no genoma das células hospedeiras, mas ao invés disso é mantido como um episssoma que assim não é replicado quando a célula hospedeira se divide. O DNA episssomal também pode estar indivíduo a degradação ao longo do tempo. Foi demonstrado que quando as células hepáticas contendo episssomas de AAV são induzidas a se dividir, o genoma do AAV não é replicado, mas ao invés disso é diluído. Como resultado, não se espera que a terapia gênica baseada em AAV seja eficaz quando administrada a crianças cujos fígados ainda não atingiram o tamanho adulto. Além disso, atualmente não se sabe quanto tempo uma terapia gênica baseada em AAV persistirá quando administrada a seres humanos adultos, embora dados em animais demonstrem apenas pequenas

perdas no efeito terapêutico por períodos de até 10 anos. Existe assim uma necessidade crítica para o desenvolvimento de novos tratamentos efetivos e permanentes para a HemA.

SUMÁRIO

[006] Em um aspecto, é no presente documento fornecida uma sequência de RNA guia (gRNA) tendo uma sequência que é complementar a uma sequência genômica dentro ou próxima de um locus endógeno de albumina.

[007] Em algumas modalidades, o gRNA compreende uma sequência espaçadora selecionada das listadas na Tabela 3 e suas variantes tendo pelo menos 85% de homologia com qualquer uma das listadas na Tabela 3.

[008] Em outro aspecto, é no presente documento fornecida uma composição que possui qualquer um dos gRNAs mencionados acima.

[009] Em algumas modalidades, o gRNA da composição compreende uma sequência espaçadora selecionada das listadas na Tabela 3 e suas variantes tendo pelo menos 85% de homologia com qualquer uma das listadas na Tabela 3.

[0010] Em algumas modalidades, a composição compreende adicionalmente um ou mais dos seguintes: uma endonuclease de ácido desoxirribonucleico (DNA) ou um ácido nucleico que codifica a endonuclease de DNA; e um molde doador tendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional.

[0011] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14,

Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional.

[0012] Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[0013] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon.

[0014] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon.

[0015] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[0016] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um ácido ribonucleico (RNA).

[0017] Em algumas modalidades, o RNA que codifica a endonuclease de DNA está ligado ao gRNA através de uma ligação covalente.

[0018] Em algumas modalidades, a composição compreende adicionalmente um lipossomo ou nanopartícula lipídica.

[0019] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[0020] Em algumas modalidades, a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[0021] Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[0022] Em algumas modalidades, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[0023] Em outro aspecto, é no presente documento proporcionado

um kit tendo qualquer uma das composições descritas acima e tendo adicionalmente instruções de uso.

[0024] Em outro aspecto, é no presente documento proporcionado um sistema compreendendo uma endonuclease de ácido desoxirribonucleico (DNA) ou ácido nucleico que codifica a referida endonuclease de DNA; um RNA guia (gRNA) compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104; e um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional.

[0025] Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[0026] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[0027] Em algumas modalidades, o ácido nucleico que codifica a

referida endonuclease de DNA é otimizado quanto a códon para expressão em uma célula hospedeira. Em algumas modalidades, a célula hospedeira é uma célula humana.

[0028] Em algumas modalidades, o ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizado quanto a códon para expressão em uma célula hospedeira. Em algumas modalidades, a célula hospedeira é uma célula humana.

[0029] Em algumas modalidades, o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[0030] Em algumas modalidades, o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA). Em algumas modalidades, o RNA que codifica a referida endonuclease de DNA é um mRNA.

[0031] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[0032] Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para um gRNA no sistema. Em algumas modalidades, o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA genômico para um gRNA no sistema.

[0033] Em algumas modalidades, a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[0034] Em algumas modalidades, o sistema compreende a

endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[0035] Em outro aspecto, é no presente documento proporcionado um método de edição de um genoma em uma célula, o método compreendendo fornecer o seguinte para a célula: (a) qualquer um dos gRNAs descritos acima; (b) uma endonuclease de ácido desoxirribonucleico (DNA) ou um ácido nucleico que codifica a endonuclease de DNA; e (c) um molde doador tendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

[0036] Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[0037] Em algumas modalidades, o gRNA tem uma sequência espaçadora selecionada das listadas na Tabela 3 e suas variantes tendo pelo menos 85% de homologia com qualquer uma das listadas na Tabela 3.

[0038] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14,

Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1; ou um seu derivado funcional.

[0039] Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[0040] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon para expressão na célula.

[0041] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizado quanto a códon para expressão na célula.

[0042] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[0043] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um ácido ribonucleico (RNA).

[0044] Em algumas modalidades, o RNA que codifica a referida endonuclease de DNA é um mRNA.

[0045] Em algumas modalidades, o RNA que codifica a endonuclease de DNA está ligado ao gRNA através de uma ligação covalente.

[0046] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[0047] Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para o gRNA de (a). Em algumas

modalidades, o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA no genoma da célula para o gRNA de (a). Em algumas modalidades,

[0048] Em algumas modalidades, um ou mais de (a), (b) e (c) são formulados em um lipossomo ou nanopartícula lipídica.

[0049] Em algumas modalidades, a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[0050] Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[0051] Em algumas modalidades, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP), antes do fornecimento para a célula.

[0052] Em algumas modalidades, (a) e (b) são fornecidos à célula após (c) ser fornecido à célula.

[0053] Em algumas modalidades, (a) e (b) são fornecidos à célula cerca de 1 a 14 dias após (c) ser fornecido à célula.

[0054] Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula.

[0055] Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após (c) ser fornecido à célula.

[0056] Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b).

[0057] Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

[0058] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é inserida em uma sequência genômica da célula.

[0059] Em algumas modalidades, a inserção está no, dentro ou próxima do gene da albumina ou dos elementos reguladores do gene da albumina no genoma da célula.

[0060] Em algumas modalidades, a inserção é no primeiro íntron do gene da albumina.

[0061] Em algumas modalidades, a inserção é de pelo menos 37 pb a jusante do final do primeiro éxon do gene da albumina humana no genoma e de pelo menos 330 pb a montante do início do segundo éxon do gene da albumina humana no genoma.

[0062] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[0063] Em algumas modalidades, a célula é um hepatócito.

[0064] Em outro aspecto, é no presente documento fornecida uma célula geneticamente modificada na qual o genoma da célula é editado por qualquer um dos métodos descritos acima.

[0065] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é

inserida em uma sequência genômica da célula.

[0066] Em algumas modalidades, a inserção está no, dentro ou próxima do gene da albumina ou dos elementos reguladores do gene da albumina no genoma da célula.

[0067] Em algumas modalidades, a inserção é no primeiro íntron do gene da albumina.

[0068] Em algumas modalidades, a inserção é de pelo menos 37 pb a jusante do final do primeiro éxon do gene da albumina humana no genoma e de pelo menos 330 pb a montante do início do segundo éxon do gene da albumina humana no genoma.

[0069] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[0070] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon.

[0071] Em algumas modalidades, a célula é um hepatócito.

[0072] Em outro aspecto, é no presente documento fornecido um método de tratamento da Hemofilia A em um indivíduo, o método compreendendo fornecer o seguinte a uma célula no indivíduo: (a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104; (b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

[0073] Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA

compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[0074] Em algumas modalidades, o indivíduo é um paciente com ou é suspeito de ter Hemofilia A.

[0075] Em algumas modalidades, o indivíduo é diagnosticado com um risco de Hemofilia A.

[0076] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1; ou um seu derivado funcional.

[0077] Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[0078] Em algumas modalidades, o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão na célula.

[0079] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizado quanto a códon para expressão na célula.

[0080] Em algumas modalidades, o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[0081] Em algumas modalidades, o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA). Em

algumas modalidades, o RNA que codifica a referida endonuclease de DNA é um mRNA.

[0082] Em algumas modalidades, um ou mais do gRNA de (a), a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) e o molde doador de (c) são formulados em um lipossomo ou nanopartícula lipídica.

[0083] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[0084] Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para o gRNA de (a). Em algumas modalidades, o local alvo de gRNA do molde doador é o complemento reverso do local alvo de gRNA no genoma da célula para o gRNA de (a).

[0085] Em algumas modalidades, fornecer o molde doador à célula compreende administrar o molde doador ao indivíduo. Em algumas modalidades, a administração é por via intravenosa.

[0086] Em algumas modalidades, endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[0087] Em algumas modalidades, fornecer o gRNA e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA à célula compreende administrar o lipossomo ou nanopartícula lipídica ao indivíduo. Em algumas modalidades, a administração é por via intravenosa.

[0088] Em algumas modalidades, o método compreende fornecer à célula a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[0089] Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após o molde doador de (c) ser fornecido à célula.

[0090] Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b). Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

[0091] Em algumas modalidades, fornecer o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) à célula compreende administrar ao indivíduo uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA.

[0092] Em algumas modalidades, fornecer o molde doador de (c) à

célula compreende administrar ao indivíduo o molde doador codificado em um vetor AAV.

[0093] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[0094] Em algumas modalidades, a célula é um hepatócito.

[0095] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa no fígado do indivíduo.

[0096] Em outro aspecto, é no presente documento fornecido um método de tratamento da Hemofilia A em um indivíduo. O método compreende a administração de qualquer uma das células geneticamente modificadas acima mencionadas ao indivíduo.

[0097] Em algumas modalidades, o indivíduo é um paciente com ou é suspeito de ter Hemofilia A.

[0098] Em algumas modalidades, o indivíduo é diagnosticado com um risco de Hemofilia A.

[0099] Em algumas modalidades, a célula geneticamente modificada é autóloga.

[00100] Em algumas modalidades, a célula é um hepatócito.

[00101] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é inserida em uma sequência genômica da célula.

[00102] Em algumas modalidades, a inserção está no, dentro ou próxima do gene da albumina ou dos elementos reguladores do gene da albumina no genoma da célula.

[00103] Em algumas modalidades, a inserção é no primeiro íntron do gene da albumina.

[00104] Em algumas modalidades, a inserção é de pelo menos 37 pb a jusante do final do primeiro éxon do gene da albumina humana no

genoma e de pelo menos 330 pb a montante do início do segundo éxon do gene da albumina humana no genoma.

[00105] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00106] Em algumas modalidades, o método compreende ainda obter uma amostra biológica do indivíduo, em que a amostra biológica compreende uma célula de hepatócito e editar o genoma da célula de hepatócito por inserção de uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional em uma sequência genômica da célula, produzindo assim a célula geneticamente modificada.

[00107] Em outro aspecto, é no presente documento fornecido um método de tratamento da Hemofilia A em um indivíduo. O método compreende a obtenção de uma amostra biológica do indivíduo em que a amostra biológica compreende uma célula de hepatócito, fornecendo o seguinte à célula de hepatócito: (a) qualquer um dos gRNA descritos acima; (b) uma endonuclease de ácido desoxirribonucleico (DNA) ou um ácido nucleico que codifica a endonuclease de DNA; e (c) um molde doador tendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, produzindo assim uma célula geneticamente modificada e administrar a célula geneticamente modificada ao indivíduo.

[00108] Em algumas modalidades, o indivíduo é um paciente com ou é suspeito de ter Hemofilia A.

[00109] Em algumas modalidades, o indivíduo é diagnosticado com um risco de Hemofilia A.

[00110] Em algumas modalidades, a célula geneticamente modificada é autóloga.

[00111] Em algumas modalidades, o gRNA compreende uma

sequência selecionada das listadas na Tabela 3 e suas variantes tendo pelo menos 85% de homologia com qualquer uma das listadas na Tabela 3.

[00112] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional.

[00113] Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00114] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon.

[00115] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon.

[00116] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é uma sequência de ácido desoxirribonucleico (DNA).

[00117] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é uma sequência de ácido ribonucleico (RNA).

[00118] Em algumas modalidades, a sequência de RNA que codifica a endonuclease de DNA está ligada ao gRNA através de uma ligação covalente.

[00119] Em algumas modalidades, um ou mais de (a), (b) e (c) são formulados em um lipossomo ou nanopartícula lipídica.

[00120] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[00121] Em algumas modalidades, a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[00122] Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[00123] Em algumas modalidades, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP), antes do fornecimento para a célula.

[00124] Em algumas modalidades, (a) e (b) são fornecidos à célula após (c) ser fornecido à célula.

[00125] Em algumas modalidades, (a) e (b) são fornecidos à célula cerca de 1 a 14 dias após (c) ser fornecido à célula.

[00126] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é inserida em uma sequência genômica da célula.

[00127] Em algumas modalidades, a inserção está no, dentro ou próxima do gene da albumina ou dos elementos reguladores do gene da albumina no genoma da célula.

[00128] Em algumas modalidades, a inserção é no primeiro íntron do gene da albumina.

[00129] Em algumas modalidades, a inserção é de pelo menos 37 pb a jusante do final do primeiro éxon do gene da albumina humana no genoma e de pelo menos 330 pb a montante do início do segundo éxon do gene da albumina humana no genoma.

[00130] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00131] Em algumas modalidades, a célula é um hepatócito.

[00132] Em outro aspecto, é no presente documento fornecido um

método de tratamento da Hemofilia A em um indivíduo. O método compreende fornecer o seguinte a uma célula no indivíduo: (a) qualquer um dos gRNA descritos acima; (b) uma endonuclease de ácido desoxirribonucleico (DNA) ou um ácido nucleico que codifica a endonuclease de DNA; e (c) um molde doador tendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

[00133] Em algumas modalidades, o indivíduo é um paciente com ou é suspeito de ter Hemofilia A.

[00134] Em algumas modalidades, o indivíduo é diagnosticado com um risco de Hemofilia A.

[00135] Em algumas modalidades, o gRNA compreende uma sequência selecionada das listadas na Tabela 3 e suas variantes tendo pelo menos 85% de homologia com qualquer uma das listadas na Tabela 3.

[00136] Em algumas modalidades, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1; ou um seu derivado funcional.

[00137] Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00138] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon.

[00139] Em algumas modalidades, a sequência de ácido nucleico

que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon.

[00140] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é uma sequência de ácido desoxirribonucleico (DNA).

[00141] Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é uma sequência de ácido ribonucleico (RNA).

[00142] Em algumas modalidades, a sequência de RNA que codifica a endonuclease de DNA está ligada ao gRNA através de uma ligação covalente.

[00143] Em algumas modalidades, um ou mais de (a), (b) e (c) são formulados em um lipossomo ou nanopartícula lipídica.

[00144] Em algumas modalidades, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[00145] Em algumas modalidades, a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[00146] Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA.

[00147] Em algumas modalidades, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP), antes do fornecimento para a célula.

[00148] Em algumas modalidades, (a) e (b) são fornecidos à célula após (c) ser fornecido à célula.

[00149] Em algumas modalidades, (a) e (b) são fornecidos à célula cerca de 1 a 14 dias após (c) ser fornecido à célula.

[00150] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é inserida em uma sequência genômica da célula.

[00151] Em algumas modalidades, a inserção está no, dentro ou próxima do gene da albumina ou dos elementos reguladores do gene

da albumina no genoma da célula.

[00152] Em algumas modalidades, a inserção é no primeiro íntron do gene da albumina no genoma da célula.

[00153] Em algumas modalidades, a inserção é de pelo menos 37 pb a jusante do final do primeiro éxon do gene da albumina humana no genoma e de pelo menos 330 pb a montante do início do segundo éxon do gene da albumina humana no genoma.

[00154] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00155] Em algumas modalidades, a célula é um hepatócito.

[00156] Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa no fígado do indivíduo.

[00157] Em outro aspecto, é no presente documento proporcionado um kit compreendendo um ou mais elementos de um sistema descrito acima e compreendendo adicionalmente instruções de uso.

BREVE DESCRIÇÃO DOS DESENHOS

[00158] Uma compreensão de determinadas características e vantagens da presente invenção será obtida por referência à seguinte descrição detalhada que apresenta modalidades ilustrativas, nas quais os princípios da invenção são usados, e os desenhos acompanhantes dos quais:

[00159] FIG. 1 mostra o alinhamento múltiplo de sequências de codificação de FVIII-BDD otimizadas quanto a códon. Apenas a sequência de codificação madura é mostrada (a região do peptídeo sinal é eliminada). O algoritmo ClustalW foi usado.

[00160] FIG. 2 mostra desenhos exemplificativos, não limitativos, do molde de DNA doador.

[00161] FIG. 3 mostra os resultados da análise TIDE da eficiência de

corte do gRNA-T1 de mAlb em células Hepa1-6.

[00162] FIG. 4 mostra os resultados de frequências INDEL no fígado e baço de camundongos 3 dias após a dosagem com nanopartículas lipídicas (LNP) encapsulando mRNA de Cas9 e gRNA_T1 de mAlb em doses diferentes ou controle de PBS. N = 5 camundongos por grupo, valores médios são representados graficamente.

[00163] FIG. 5 mostra desenhos de moldes doadores de DNA para integração direcionada ao íntron 1 da albumina usados no Exemplo 4. SA; sequência aceitadora de splice, LHA; Braço de homologia esquerdo; RHA; braço de homologia direito, pA; sinal de poliadenilação, local de gRNA; local alvo para o gRNA que medeia o corte pela nuclease Cas9 direcionada ao gRNA, delta furina; deleção do local da furina em FVIII, FVIII-BDD; sequência de codificação para FVIII humano com deleção do domínio B (BDD) na qual o domínio B é substituído pelo peptídeo de ligação SQ.

[00164] FIG. 6 mostra frequências INDEL de 8 gRNA candidatos direcionados ao íntron 1 da albumina humana em hepatócitos humanos primários de 4 doadores. O gRNA direcionado ao locus AAVS1 e o gene humano não relacionado (C3) são incluídos como controles.

[00165] FIG. 7 mostra frequências INDEL em hepatócitos primários não humanos de primatas (Macaco) transfectados com diferentes RNAs guia de albumina e mRNA de spCas9.

[00166] FIG. 8 mostra um esquema de um cassete doador de AAV-mSEAP exemplificativo.

[00167] FIG. 9 mostra um esquema de um cassete doador de FVIII exemplificativo usado para empacotamento no AAV.

[00168] FIG. 10 mostra os níveis de FVIII no sangue de camundongos com hemofilia A ao longo do tempo após a injeção de AAV8-pCB056, seguida pela LNP que encapsula o mRNA spCas9 e o RNA guia mAlbT1.

[00169] FIG. 11 mostra os níveis de FVIII nos camundongos com Hemofilia A no dia 10 e no dia 17 após a LNP que encapsula o mRNA spCas9 e o gRNA ser injetada. A LNP foi dosada 17 dias ou 4 dias após AAV8-pCB056.

[00170] FIG. 12 mostra um esquema de doadores de plasmídeo exemplificativos contendo o gene FVIII humano e diferentes sequências de sinais de poliadenilação.

[00171] FIG. 13 mostra a atividade do FVIII e as razões de atividade de FVIII/integração direcionada em camundongos após injeção hidrodinâmica de doadores plasmídicos com 3 sinais de poliA diferentes seguidos por mRNA de Cas9 e gRNA de mAlbT1 encapsulados em LNP. Os grupos 2, 3 e 4 foram doseados com pCB065, pCB076 e pCB077, respectivamente. A tabela contém os valores para a atividade do FVIII no dia 10, a frequência de integração direcionada e a atividade do FVIII/razão de TI (Razão) para cada camundongo individual.

[00172] FIG. 14 mostra um esquema de cassetes doadores AAV exemplificativos usados para avaliar a integração direcionada em hepatócitos humanos primários.

[00173] FIG. 15 mostra a atividade de SEAP no meio de hepatócitos humanos primários transduzidos com o vírus AAV-DJ-SEAP com ou sem lipofecção do mRNA de spCas9 e gRNA de hALb4. Dois doadores de células foram testados (HJK, ONR) indicados pelas barras preto e branco. Os 3 pares de barras à esquerda representam a atividade de SEAP nas condições de controle das células transfectadas apenas com Cas9 e gRNA (primeiro par de barras), AAV-DJ-pCB0107 (vírus SEAP) a 100.000 MOI isoladamente (segundo par de barras) ou AAV-DJ-pCB0156 (vírus FVIII) a 100.000 MOI isoladamente (terceiro par de barras). Os 4 pares de barras à direita representam a atividade de SEAP em poços de células transduzidas com o AAV-DJ-pCB0107 (vírus SEAP) a vários MOI e transfectadas com mRNA de Cas9 e o gRNA de

hAlb T4.

[00174] FIG. 16 mostra a atividade de FVIII no meio de hepatócitos humanos primários transduzidos com o vírus AAV-DJ-FVIII com ou sem lipofecção do mRNA de spCas9 e gRNA de hALb4. Dois doadores de células foram testados (HJK, ONR) indicados pelas barras preto e branco. Os 2 pares de barras à esquerda representam a atividade do FVIII nas condições de controle das células transduzidas com AAV-DJ-pCB0107 (vírus SEAP) a 100.000 MOI isoladamente (primeiro par de barras) ou AAV-DJ-pCB0156 (vírus FVIII) a 100.000 MOI isoladamente (segundo par de barras). Os 4 pares de barra à direita representam a atividade do FVIII nos meios de poços de células transduzidas com o AAV-DJ-pCB0156 (vírus FVIII) a vários MOI e transfectadas com mRNA de Cas9 e o gRNA de hAlb T4.

DESCRIÇÃO DETALHADA

[00175] As invenções fornecem, *inter alia*, composições e métodos para edição para modular a expressão, função ou atividade de uma proteína de coagulação do sangue, como o Fator VIII (FVIII) em uma célula por edição do genoma. As invenções também fornecem, *inter alia*, composições e métodos para o tratamento de um paciente com Hemofilia A, *ex vivo* e *in vivo*.

DEFINIÇÕES

[00176] A menos que definido de outro modo, todos os termos técnicos e científicos usados no presente documento têm o mesmo significado como é comumente entendido por um perito na técnica aos quais o assunto reivindicado pertence. Deve ser entendido que as descrições detalhadas são apenas exemplificativas e explicativas e não são restritivas a qualquer assunto reivindicado. Em este pedido, o uso do singular inclui o plural a não ser que especificamente afirmado de outro modo. Deve ser salientado que, como usadas na descrição, as formas singulares "um", "uma" e "o/a" incluem as referências plurais a

menos que o contexto dite claramente o contrário. Em este pedido, o uso de "ou" significa "e/ou" a não ser que apresentado de outro modo. Além do mais, o uso do termo "incluindo" bem como outras formas, tais como "inclui" e "incluído", não é limitante.

[00177] Embora várias características das invenções possam ser descritas no contexto de uma única modalidade, as características também podem ser fornecidas separadamente ou em qualquer combinação adequada. Por outro lado, embora as invenções possam ser descritas no presente documento no contexto de modalidades separadas para maior clareza, as invenções também podem ser implementadas em uma única modalidade. Qualquer uma das pedidos de patentes publicados e qualquer uma das outras referências, documentos, manuscritos e literatura científica publicados citados no presente documento são incorporados no presente documento por referência para qualquer propósito. Em caso de conflito, a presente descrição, incluindo definições, prevalecerá. Para além disso, os materiais, métodos e exemplos são apenas ilustrativos e não pretendem ser limitativos.

[00178] Como no presente documento usado, intervalos e quantidades podem ser expressos como "cerca de" um valor ou intervalo específico. Cerca de também inclui a quantidade exata. Portanto, "cerca de 5 µL" significa "cerca de 5 µL" e também "5 µL". Geralmente, o termo "cerca de" inclui uma quantidade que seria esperada dentro de um erro experimental, tal como ± 10%.

[00179] Quando um intervalo de valores numéricos é apresentado no presente documento, é contemplado que cada valor intermediário entre o limite inferior e superior do intervalo, os valores que são os limites superior e inferior do intervalo e todos os valores declarados com o intervalo sejam abrangidos pela invenção. Todos os subintervalos possíveis dentro dos limites inferior e superior do intervalo também são

contemplados pela invenção.

[00180] Os termos "polipeptídeo", "sequência polipeptídica", "peptídeo", "sequência peptídica", "proteína", "sequência de proteína" e "sequência de aminoácidos" são usados no presente documento indistintamente para designar uma série linear de resíduos de aminoácidos conectados um ao outro por ligações peptídicas, cuja série pode incluir proteínas, polipeptídeos, oligopeptídeos, peptídeos e seus fragmentos. A proteína pode ser composta de aminoácidos de ocorrência natural e/ou aminoácidos sintéticos (por exemplo, modificados ou de ocorrência não natural). Assim, "aminoácido" ou "resíduo peptídico", como no presente documento usado, significa aminoácidos de ocorrência natural e sintéticos. Os termos "polipeptídeo", "peptídeo" e "proteína" incluem proteínas de fusão, incluindo, mas não limitado a proteínas de fusão com uma sequência de aminoácidos heteróloga, fusões com sequências líder heterólogas e homólogas, com ou sem resíduos de metionina no terminal N; proteínas marcadas imunologicamente; proteínas de fusão com parceiros de fusão detectáveis, por exemplo, proteínas de fusão incluindo como parceiro de fusão uma proteína fluorescente, β-galactosidase, luciferase e similares. Além disso, deve ser notado que um traço no início ou no final de uma sequência de aminoácidos indica uma ligação peptídica a uma sequência adicional de um ou mais resíduos de aminoácidos ou uma ligação covalente a um grupo terminal carboxila ou hidroxila. No entanto, a ausência de um traço não deve ser entendida como significando que essa ligação peptídica ou ligação covalente a um grupo terminal carboxila ou hidroxila não está presente, pois é convencional na representação de sequências de aminoácidos omitir tal.

[00181] O termo "polinucleotídeo", "sequência polinucleotídica", "oligonucleotídeo", "sequência de oligonucleotídeo", "oligômero", "oligo", "sequência de ácido nucleico" ou "sequência de nucleotídeo" no

presente documento usado indistintamente, refere-se a uma forma polimérica de nucleotídeos de qualquer comprimento, ribonucleotídeos ou desoxirribonucleotídeos. Assim, este termo inclui, mas não está limitado a, DNA ou RNA de fita simples, dupla ou múltipla, DNA genômico, cDNA, híbridos de DNA-RNA ou um polímero com bases de purina e pirimidina ou outras bases de ocorrência natural, bases nucleotídicas quimicamente ou bioquimicamente modificadas, de ocorrência não natural ou derivatizadas.

[00182] Os termos "derivado" e "variante" referem-se, sem limitação, a qualquer composto, como ácido nucleico ou proteína, que tem uma estrutura ou sequência derivada dos compostos no presente documento descritos e cuja estrutura ou sequência é suficientemente similar às no presente documento descritas, de modo que possui atividades e utilidades iguais ou similares ou, com base nessa similaridade, seria esperado por um perito na técnica exibir atividades e utilidades iguais ou similares como as dos compostos referenciados, deste modo também indistintamente referidos como "funcionalmente equivalente" ou como "equivalentes funcionais." As modificações para obter "derivados" ou "variantes" podem incluir, por exemplo, adição, deleção e/ou substituição de um ou mais dos ácidos nucleicos ou resíduos de aminoácidos.

[00183] O equivalente funcional ou fragmento do equivalente funcional, no contexto de uma proteína, pode ter uma ou mais substituições de aminoácidos conservativas. O termo "substituição de aminoácidos conservativa" refere-se à substituição de um aminoácido por outro aminoácido que possui propriedades similares ao aminoácido original. Os grupos de aminoácidos conservativos são os seguintes:

Grupo	Nome dos aminoácidos
Alifático	Gly, Ala, Val, Leu, Ile
Contendo Hidroxila ou	Ser, Cys, Thr, Met

Sulfidrila/Selênio	
Cíclico	Pro
Aromático	Phe, Tyr, Trp
Básico	His, Lys, Arg
Ácido e sua Amida	Asp, Glu, Asn, Gln

[00184] As substituições conservativas podem ser introduzidas em qualquer posição de um peptídeo predeterminado preferido ou seu fragmento. No entanto, também pode ser desejável introduzir substituições não conservativas, particularmente, mas não se limitando a, uma substituição não conservativa em qualquer uma ou mais posições. Uma substituição não conservativa que leva à formação de um fragmento funcionalmente equivalente do peptídeo poderá, por exemplo, diferir substancialmente na polaridade, na carga elétrica e/ou no volume estérico, embora mantendo a funcionalidade do fragmento derivado ou variante.

[00185] "Porcentagem de identidade da sequência" é determinada comparando duas sequências otimamente alinhadas em uma janela de comparação, em que a porção da sequência polinucleotídica ou polipeptídica na janela de comparação pode ter adições ou deleções (isto é, intervalos) em comparação com a sequência de referência (que não possui adições ou deleções) para o alinhamento ótimo das duas sequências. Em alguns casos, a porcentagem pode ser calculada determinando o número de posições nas quais a base de ácido nucleico idêntica ou resíduo de aminoácido ocorre em ambas as sequências para originar o número de posições correspondentes, dividindo o número de posições correspondentes pelo número total de posições na janela de comparação e multiplicando o resultado por 100 para originar a porcentagem da identidade de sequências.

[00186] Os termos "idêntico" ou porcentagem de "identidade" no contexto de duas ou mais sequências de ácido nucleico ou polipeptídeo

referem-se a duas ou mais sequências ou subsequências iguais ou que têm uma porcentagem especificada de resíduos de aminoácidos ou nucleotídeos iguais (por exemplo, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% ou 99% de identidade em uma região especificada, por exemplo, todas as sequências polipeptídicas ou domínios individuais do polipeptídeos), quando comparados e alinhados para obter a máxima correspondência sobre uma janela de comparação ou região designada, como medido usando um dos seguintes algoritmos de comparação de sequência ou por alinhamento manual e inspeção visual. Tais sequências são ditas ser "substancialmente idênticas". Esta definição também refere-se ao complemento de uma sequência de teste.

[00187] O termo "complementar" ou "substancialmente complementar", no presente documento usado indistintamente, significa que um ácido nucleico (por exemplo, DNA ou RNA) tem uma sequência de nucleotídeos que permite a ligação não covalente, isto é, forma pares de bases Watson-Crick e/ou pares de bases G/U com outro ácido nucleico de forma específica de sequência, antiparalela (isto é, um ácido nucleico se liga especificamente a um ácido nucleico complementar). Como é conhecido na técnica, o emparelhamento de bases padrão Watson-Crick inclui: emparelhamento de adenina (A) com timidina (T), emparelhamento de adenina (A) com uracil (U) e emparelhamento de guanina (G) com citosina (C).

[00188] Uma sequência de DNA que "codifica" um RNA específico é uma sequência de ácido nucleico de DNA que é transcrita no RNA. Um polinucleotídeo de DNA pode codificar um RNA (mRNA) que é traduzido em proteína, ou um polinucleotídeo de DNA pode codificar um RNA que não é traduzido em proteína (por exemplo, tRNA, rRNA ou um RNA guia; também chamado de RNA "não codificante" ou "ncRNA"). Uma sequência "codificadora de proteína" ou uma sequência que codifica uma proteína ou polipeptídeo específico, é uma sequência de ácido

nucleico que é transcrita em mRNA (no caso de DNA) e é traduzida (no caso de mRNA) em um polipeptídeo *in vitro* ou *in vivo* quando colocada sob o controle de sequências reguladoras apropriadas.

[00189] Como no presente documento usado, "códon" refere-se a uma sequência de três nucleotídeos que juntos formam uma unidade de código genético em uma molécula de DNA ou RNA. Como no presente documento usado, o termo "degenerescência de códons" refere-se à natureza no código genético que permite variação da sequência nucleotídica sem afetar a sequência de aminoácidos de um polipeptídeo codificado.

[00190] O termo "otimizado quanto a códon" ou "otimização de códon" refere-se a genes ou regiões de codificação de moléculas de ácido nucleico para transformação de vários hospedeiros, refere-se à alteração de códons no gene ou em regiões de codificação de moléculas de ácido nucleico para refletir o uso típico de códon do organismo hospedeiro sem alterar o polipeptídeo codificado pelo DNA. Essa otimização inclui a substituição de pelo menos um, ou mais de um, ou um número significativo de códons por um ou mais códons que são mais frequentemente usados nos genes desse organismo. As tabelas de uso de códons estão facilmente disponíveis, por exemplo, na "Base de dados de Uso de Códon" disponível em www.kazusa.or.jp/codon/ (visitado em 20 de março de 2008). Usando o conhecimento sobre o uso ou preferência de códons em cada organismo, um vulgar perito na técnica pode aplicar as frequências a qualquer sequência polipeptídica especificada e produzir um fragmento de ácido nucleico de uma região de codificação otimizada quanto a códons que codifica o polipeptídeo, mas que usa códons ideais para uma dada espécie. As regiões de codificação otimizadas quanto a códon podem ser desenhadas por vários métodos conhecidos dos peritos na técnica.

[00191] O termo "recombinante" ou "manipulado" quando usado com

referência, por exemplo, a uma célula, um ácido nucleico, uma proteína ou um vetor, indica que a célula, ácido nucleico, proteína ou vetor foi modificado por ou é o resultado de métodos laboratoriais. Assim, por exemplo, proteínas recombinantes ou manipuladas incluem proteínas produzidas por métodos laboratoriais. As proteínas recombinantes ou manipuladas podem incluir resíduos de aminoácidos não encontrados na forma nativa (não recombinante ou de tipo selvagem) da proteína ou podem incluir resíduos de aminoácidos que foram modificados, por exemplo, marcados. O termo pode incluir qualquer uma das modificações na sequência de peptídeo, proteína ou ácido nucleico. Tais modificações podem incluir o seguinte: qualquer uma das modificações químicas da sequência do peptídeo, proteína ou ácido nucleico, incluindo um ou mais aminoácidos, desoxirribonucleotídeos ou ribonucleotídeos; adição, deleção e/ou substituição de um ou mais aminoácidos no peptídeo ou proteína; e adição, deleção e/ou substituição de um ou mais ácidos nucleicos na sequência de ácidos nucleicos.

[00192] O termo "DNA genômico" ou "sequência genômica" refere-se ao DNA de um genoma de um organismo incluindo, mas não limitado ao, DNA do genoma de uma bactéria, fungo, arquea, planta ou animal.

[00193] Como usado no presente documento, "transgene", "gene exógeno" ou "sequência exógena", no contexto de ácido nucleico, refere-se a uma sequência de ácido nucleico ou gene que não estava presente no genoma de uma célula, mas introduzido artificialmente no genoma, por exemplo, através de edição do genoma.

[00194] Como no presente documento usado, "gene endógeno" ou "sequência endógena", no contexto de ácido nucleico, refere-se a uma sequência de ácido nucleico ou gene que está naturalmente presente no genoma de uma célula, sem ser introduzido através de qualquer meio artificial.

[00195] O termo "vetor" ou "vetor de expressão" significa um replicon, tal como plasmídeo, fago, vírus ou cosmídeo, ao qual outro segmento de DNA, isto é, um "inserto", pode ser ligado de modo a provocar a replicação do segmento ligado em uma célula.

[00196] O termo "cassete de expressão" refere-se a um vetor que tem uma sequência de codificação de DNA operacionalmente ligada a um promotor. "Operacionalmente ligado" refere-se a uma justaposição em que os componentes descritos no presente documento estão em uma relação permitindo que funcionem do seu modo pretendido. Por exemplo, um promotor está operacionalmente ligado a uma sequência de codificação se o promotor afeta sua transcrição ou expressão. Os termos "vetor de expressão recombinante" ou "construto de DNA" são no presente documento usados indistintamente para se referir a uma molécula de DNA que tem um vetor e pelo menos um inserto. Os vetores de expressão recombinantes são geralmente gerados com a finalidade de expressar e/ou propagar o(s) inserto(s) ou para a construção de outras sequências de nucleotídeos recombinantes. O(s) ácido(s) nucleico(s) pode(m) ou não estar operacionalmente ligado(s) a uma sequência promotora e pode(m) ou não estar operacionalmente ligado(s) a sequências reguladoras de DNA.

[00197] O termo "operacionalmente ligado" significa que a sequência nucleotídica de interesse está ligada à(s) sequência(s) reguladora(s) de uma maneira que permita a expressão da sequência nucleotídica. O termo "sequência reguladora" pretende incluir, por exemplo, promotores, potenciadores e outros elementos de controle de expressão (por exemplo, sinais de poliadenilação). Tais sequências reguladoras são bem conhecidas na técnica e são descritas, por exemplo, em Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Sequências reguladoras incluem aquelas que direcionam a expressão constitutiva

de uma sequência nucleotídica em muitos tipos de células hospedeiras e aquelas que direcionam a expressão da sequência nucleotídica apenas em certas células hospedeiras (por exemplo, sequências reguladoras específicas de tecido). Será apreciado pelos peritos na técnica que o desenho do vetor de expressão pode depender de fatores como a escolha da célula alvo, o nível de expressão desejado e similares.

[00198] Uma célula foi "geneticamente modificada" ou "transformada" ou "transfectada" por DNA exógeno, por exemplo, um vetor de expressão recombinante, quando esse DNA foi introduzido dentro da célula. A presença do DNA exógeno resulta em alteração gênica permanente ou transiente. O DNA transformador pode ou não ser integrado (ligado covalentemente) no genoma da célula. As células geneticamente modificadas (ou transformadas ou transfectadas) que possuem atividade terapêutica, por exemplo, tratamento da Hemofilia A, podem ser usadas e referidas como células terapêuticas.

[00199] O termo "concentração" usado no contexto de uma molécula, tal como um fragmento peptídico, refere-se a uma quantidade da molécula, por exemplo, o número de moles da molécula, presente em um determinado volume de solução.

[00200] Os termos "indivíduo", "sujeito" e "hospedeiro" são usados no presente documento indistintamente e referem-se a qualquer indivíduo para quem o diagnóstico, tratamento ou terapia é desejado. Em alguns aspectos, o indivíduo é um mamífero. Em alguns aspectos, o indivíduo é um ser humano. Em alguns aspectos, o indivíduo é um paciente humano. Em alguns aspectos, o indivíduo pode ter ou é suspeito de ter Hemofilia A e/ou tem um ou mais sintomas da Hemofilia A. Em alguns aspectos, o indivíduo é um ser humano diagnosticado com um risco de Hemofilia A no momento do diagnóstico ou mais tarde. Em alguns casos, o diagnóstico com um risco de Hemofilia A pode ser

determinado com base na presença de uma ou mais mutações no gene endógeno do Fator VIII (FVIII) ou na sequência genômica próxima ao gene do Fator VIII (FVIII) no genoma que pode afetar a expressão do gene FVIII.

[00201] O termo "tratamento" usado referente a uma doença ou afecção significa que é alcançada pelo menos uma melhoria dos sintomas associados à afecção que afeta um indivíduo, onde a melhoria é usada em sentido amplo para se referir a pelo menos uma redução na magnitude de um parâmetro, por exemplo, um sintoma, associado à afecção (por exemplo, Hemofilia A) sendo tratada. Como tal, o tratamento também inclui situações em que a condição patológica, ou pelo menos os sintomas associados a ela, são completamente inibidos, por exemplo, impedidos de acontecer, ou eliminados inteiramente, de modo que o hospedeiro não sofra mais da afecção ou pelo menos os sintomas que caracterizam a afecção. Assim, o tratamento inclui: (i) prevenção, isto é, redução do risco de desenvolvimento de sintomas clínicos, inclusive fazendo com que os sintomas clínicos não se desenvolvam, por exemplo, impedindo a progressão da doença; (ii) inibição, isto é, interrupção do desenvolvimento ou desenvolvimento adicional de sintomas clínicos, por exemplo, mitigar ou inibir completamente uma doença ativa.

[00202] Os termos "quantidade eficaz", "quantidade farmaceuticamente eficaz" ou "quantidade terapeuticamente eficaz", como no presente documento usados, significam uma quantidade suficiente da composição para fornecer a utilidade desejada quando administrada a um indivíduo com uma afecção específica. No contexto do tratamento *ex vivo* da Hemofilia A, o termo "quantidade efetiva" refere-se à quantidade de uma população de células terapêuticas ou sua progênie necessária para prevenir ou aliviar pelo menos um ou mais sinais ou sintomas da Hemofilia A e refere-se a uma quantidade

suficiente de uma composição possuindo as células terapêuticas ou sua progênie para fornecer o efeito desejado, por exemplo, tratar os sintomas da Hemofilia A de um indivíduo. O termo "quantidade terapeuticamente eficaz" refere-se portanto a uma quantidade de células terapêuticas ou a uma composição tendo células terapêuticas que é suficiente para promover um efeito específico quando administrada a um indivíduo que necessita de tratamento, como alguém que tem ou está em risco para Hemofilia A. Uma quantidade eficaz também incluiria uma quantidade suficiente para impedir ou atrasar o desenvolvimento de um sintoma da doença, alterar o curso de um sintoma da doença (por exemplo, mas não limitado a, retardar a progressão de um sintoma da doença) ou reverter um sintoma da doença. No contexto do tratamento *in vivo* da Hemofilia A em um indivíduo (por exemplo, paciente) ou edição do genoma feita em uma célula cultivada *in vitro*, uma quantidade eficaz refere-se a uma quantidade de componentes usados para a edição do genoma, tal como gRNA, molde doador e/ou um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA) necessário para editar o genoma da célula no indivíduo ou a célula cultivada *in vitro*. É entendido que para qualquer caso, uma "quantidade eficaz" apropriada pode ser determinada por um vulgar perito na técnica usando experimentação de rotina.

[00203] O termo "excipiente farmaceuticamente aceitável", tal como no presente documento usado, refere-se a qualquer substância adequada que forneça um transportador, aditivo ou diluente farmaceuticamente aceitável para administração de um composto(s) de interesse a um indivíduo. "Excipiente farmaceuticamente aceitável" pode abranger substâncias referidas como diluentes farmaceuticamente aceitáveis, aditivos farmaceuticamente aceitáveis e transportadores farmaceuticamente aceitáveis.

ÁCIDOS NUCLEICOS

Ácido Nucleico ou RNA Guia Direcionado ao Genoma

[00204] A presente invenção fornece um ácido nucleico direcionado ao genoma que pode direcionar as atividades de um polipeptídeo associado (por exemplo, um polipeptídeo direcionado ao local ou endonuclease de DNA) para uma sequência alvo específica dentro de um ácido nucleico alvo. Em algumas modalidades, o ácido nucleico direcionado ao genoma é um RNA. Um RNA direcionado ao genoma é referido no presente documento como "RNA guia" ou "gRNA". Um RNA guia tem pelo menos uma sequência espaçadora que hibrida com uma sequência de ácido nucleico alvo de interesse e uma sequência de repetição CRISPR. Nos sistemas do Tipo II, o gRNA também possui um segundo RNA chamado sequência de tracrRNA. No RNA guia (gRNA) do Tipo II, a sequência de repetição CRISPR e a sequência de tracrRNA hibridam entre si para formar um dúplex. No RNA guia (gRNA) do Tipo V, o crRNA forma um dúplex. Em ambos os sistemas, o dúplex liga um polipeptídeo direcionado ao local, de modo que o RNA guia e o polipeptídeo direcionado ao local formam um complexo. O ácido nucleico direcionado ao genoma fornece especificidade para o alvo ao complexo em virtude da sua associação com o polipeptídeo direcionado ao local. O ácido nucleico direcionado ao genoma direciona assim a atividade do polipeptídeo direcionado ao local.

[00205] Em algumas modalidades, o ácido nucleico direcionado ao genoma é um RNA guia de molécula dupla. Em algumas modalidades, o ácido nucleico direcionado ao genoma é um RNA guia de molécula única. Um RNA guia de molécula dupla possui duas fitas de RNA. A primeira fita possui na direção 5' para 3' uma sequência espaçadora aumentada opcional, uma sequência espaçadora e uma sequência de repetição CRISPR mínima. A segunda fita possui uma sequência de tracrRNA mínima (complementar à sequência de repetição CRISPR

mínima), uma sequência de tracrRNA 3' e uma sequência de extensão do tracrRNA opcional. Um RNA guia de molécula única (sgRNA) em um sistema do Tipo II tem, na direção 5' para 3', uma sequência espaçadora aumentada opcional, uma sequência espaçadora, uma sequência de repetição CRISPR mínima, um ligante guia de molécula única, uma sequência de tracrRNA mínima, uma sequência de tracrRNA 3' e uma sequência de extensão do tracrRNA opcional. A extensão do tracrRNA opcional pode ter elementos que contribuem com funcionalidade adicional (por exemplo, estabilidade) para o RNA guia. O ligante guia de molécula única liga a repetição CRISPR mínima e a sequência de tracrRNA mínima para formar uma estrutura em grampo. A extensão do tracrRNA opcional possui um ou mais grampos. Um RNA guia de molécula única (sgRNA) em um sistema do Tipo V tem, na direção 5' para 3', uma sequência de repetição CRISPR mínima e uma sequência espaçadora.

[00206] A título ilustrativo, os RNAs guias usados no sistema CRISPR/Cas/Cpf1 ou outros RNAs menores podem ser facilmente sintetizados por meios químicos, como ilustrado abaixo e descrito na técnica. Embora os procedimentos químicos sintéticos estejam em constante expansão, a purificação desses RNAs por procedimentos como a cromatografia líquida de alto desempenho (HPLC, que evita o uso de géis como PAGE) tende a se tornar mais desafiadora, pois os comprimentos dos polinucleotídeos aumentam significativamente além de cem nucleotídeos. Uma abordagem usada para gerar RNAs de maior comprimento é produzir duas ou mais moléculas que são ligadas entre si. RNAs muito mais longos, tais como os que codificam uma endonuclease Cas9 ou Cpf1, são mais facilmente gerados enzimaticamente. Vários tipos de modificações de RNA podem ser introduzidos durante ou após a síntese química e/ou produção enzimática de RNAs, por exemplo, modificações que melhoram a

estabilidade, reduzem a probabilidade ou o grau de resposta imune inata e/ou melhoram outros atributos, como descrito na técnica.

Sequência Espaçadora Aumentada

[00207] Em algumas modalidades de ácidos nucleicos direcionados ao genoma, uma sequência espaçadora aumentada pode modificar a atividade, fornecer estabilidade e/ou fornecer um local para modificações de um ácido nucleico direcionado ao genoma. Uma sequência espaçadora aumentada pode modificar a atividade ou especificidade dentro ou fora do alvo. Em algumas modalidades, é fornecida uma sequência espaçadora aumentada. Uma sequência espaçadora aumentada pode ter um comprimento superior a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000 ou 7000 ou mais nucleotídeos. Uma sequência espaçadora aumentada pode ter um comprimento de cerca de 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000 ou 7000 ou mais nucleotídeos. Uma sequência espaçadora aumentada pode ter um comprimento inferior a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 1000, 2000, 3000, 4000, 5000, 6000, 7000 ou mais nucleotídeos. Em algumas modalidades, uma sequência espaçadora aumentada tem menos de 10 nucleotídeos de comprimento. Em algumas modalidades, uma sequência espaçadora aumentada tem entre 10-30 nucleotídeos de comprimento. Em algumas modalidades, uma sequência espaçadora aumentada tem entre 30-70 nucleotídeos de comprimento.

[00208] Em algumas modalidades, a sequência espaçadora aumentada tem outra fração (por exemplo, uma sequência de controle de estabilidade, uma sequência de ligação a endorribonucleases, uma

ribozima). Em algumas modalidades, a fração diminui ou aumenta a estabilidade de um ácido nucleico direcionando ácido nucleico. Em algumas modalidades, a fração é um segmento terminador da transcrição (isto é, uma sequência de terminação da transcrição). Em algumas modalidades, a fração funciona em uma célula eucariótica. Em algumas modalidades, a fração funciona em uma célula procariótica. Em algumas modalidades, a fração funciona em células eucarióticas e procarióticas. Exemplos não limitativos de frações funcionais adequados incluem: uma extremidade 5' (por exemplo, uma extremidade 7-metilguanilato (m7 G)), uma sequência ribocomutadora (por exemplo, para permitir estabilidade regulada e/ou acessibilidade regulada por proteínas e complexos de proteínas), uma sequência que forma um dúplex de dsRNA (isto é, um grampo), uma sequência que direciona o RNA para uma localização subcelular (por exemplo, núcleo, mitocôndrias, cloroplastos e similares), uma modificação ou sequência que fornece monitorização (por exemplo, conjugação direta a uma molécula fluorescente, conjugação a uma fração que facilita a detecção fluorescente, uma sequência que permite a detecção fluorescente, etc.) e/ou uma modificação ou sequência que fornece um local de ligação para proteínas (por exemplo, proteínas que atuam no DNA, incluindo ativadores transcricionais, repressores transcricionais, DNA metiltransferases, desmetilases de DNA, histonas acetiltransferases, histona deacetilases e similares).

Sequência Espaçadora

[00209] A sequência espaçadora hibrida com uma sequência em um ácido nucleico alvo de interesse. O espaçador de um ácido nucleico direcionado ao genoma interage com um ácido nucleico alvo de um modo específico de sequência através hibridação (isto é, emparelhamento de bases). A sequência nucleotídica do espaçador varia assim dependendo da sequência do ácido nucleico alvo de

interesse.

[00210] Em um sistema CRISPR/Cas do presente documento, a sequência espaçadora é projetada para hibridar com um ácido nucleico alvo que está localizado a 5' de um PAM da enzima Cas9 usada no sistema. O espaçador pode perfeitamente emparelhar com a sequência alvo ou pode ter incompatibilidades. Cada enzima Cas9 possui uma sequência do PAM específica que reconhece em um DNA alvo. Por exemplo, *S. pyogenes* reconhece em um ácido nucleico alvo um PAM que possui a sequência 5'-NRG-3', onde R tem A ou G, onde N é qualquer nucleotídeo e N está imediatamente a 3' da sequência de ácido nucleico alvo direcionada pela sequência espaçadora.

[00211] Em algumas modalidades, a sequência de ácido nucleico alvo tem 20 nucleotídeos. Em algumas modalidades, o ácido nucleico alvo tem menos de 20 nucleotídeos. Em algumas modalidades, o ácido nucleico alvo tem mais de 20 nucleotídeos. Em algumas modalidades, o ácido nucleico alvo tem pelo menos: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 ou mais nucleotídeos. Em algumas modalidades, o ácido nucleico alvo tem no máximo: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 ou mais nucleotídeos. Em algumas modalidades, a sequência de ácido nucleico alvo tem 20 bases imediatamente 5' do primeiro nucleotídeo de PAM. Por exemplo, em uma sequência tendo 5'-NNNNNNNNNNNNNNNNNNNNNRG-3' (SEQ ID NO: 100), o ácido nucleico alvo tem a sequência que corresponde aos Ns, em que N é qualquer nucleotídeo, e a sequência NRG sublinhada (R é G ou A) é o PAM de Cas9 de *Streptococcus pyogenes*. Em algumas modalidades, a sequência do PAM usada nas composições e métodos da presente invenção como uma sequência reconhecida por Cas9 de *S.p*. é NGG.

[00212] Em algumas modalidades, a sequência espaçadora que hibrida com o ácido nucleico alvo tem um comprimento de pelo menos cerca de 6 nucleotídeos (nt). A sequência espaçadora pode ser pelo

menos cerca de 6 nt, cerca de 10 nt, cerca de 15 nt, cerca de 18 nt, cerca de 19 nt, cerca de 20 nt, cerca de 25 nt, cerca de 30 nt, cerca de 35 nt ou cerca de 40 nt, de cerca de 6 nt a cerca de 80 nt, de cerca de 6 nt a cerca de 50 nt, de cerca de 6 nt a cerca de 45 nt, de cerca de 6 nt a cerca de 40 nt, de cerca de 6 nt a cerca de 35 nt, de cerca de 6 nt a cerca de 30 nt, de cerca de 6 nt a cerca de 25 nt, de cerca de 6 nt a cerca de 20 nt, de cerca de 6 nt a cerca de 19 nt, de cerca de 10 nt a cerca de 50 nt, de cerca de 10 nt a cerca de 45 nt, de cerca de 10 nt a cerca de 40 nt, de cerca de 10 nt a cerca de 35 nt, de cerca de 10 nt a cerca de 30 nt, de cerca de 10 nt a cerca de 25 nt, de cerca de 10 nt a cerca de 20 nt, de cerca de 10 nt a cerca de 19 nt, de cerca de 19 nt a cerca de 25 nt, de cerca de 19 nt a cerca de 30 nt, de cerca de 19 nt a cerca de 35 nt, de cerca de 19 nt a cerca de 40 nt, de cerca de 19 nt a cerca de 45 nt, de cerca de 19 nt a cerca de 50 nt, de cerca de 19 nt a cerca de 60 nt, de cerca de 20 nt a cerca de 25 nt, de cerca de 20 nt a cerca de 30 nt, de cerca de 20 nt a cerca de 35 nt, de cerca de 20 nt a cerca de 40 nt, de cerca de 20 nt a cerca de 45 nt, de cerca de 20 nt a cerca de 50 nt ou de cerca de 20 nt a cerca de 60 nt. Em algumas modalidades, a sequência espaçadora tem 20 nucleotídeos. Em algumas modalidades, o espaçador tem 19 nucleotídeos. Em algumas modalidades, o espaçador tem 18 nucleotídeos. Em algumas modalidades, o espaçador tem 17 nucleotídeos. Em algumas modalidades, o espaçador tem 16 nucleotídeos. Em algumas modalidades, o espaçador tem 15 nucleotídeos.

[00213] Em algumas modalidades, a porcentagem de complementaridade entre a sequência espaçadora e o ácido nucleico alvo é de pelo menos cerca de 30%, pelo menos cerca de 40%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 65%, pelo menos cerca de 70%, pelo menos cerca de 75%, pelo menos cerca de 80%, pelo menos cerca de 85%, pelo menos cerca de 90%,

pelo menos cerca de 95%, pelo menos cerca de 97%, pelo menos cerca de 98%, pelo menos cerca de 99% ou 100%. Em algumas modalidades, a porcentagem de complementaridade entre a sequência espaçadora e o ácido nucleico alvo é no máximo cerca de 30%, no máximo cerca de 40%, no máximo cerca de 50%, no máximo cerca de 60%, no máximo cerca de 65%, no máximo cerca de 70%, no máximo cerca de 75%, no máximo cerca de 80%, no máximo cerca de 85%, no máximo cerca de 90%, no máximo cerca de 95%, no máximo cerca de 97%, no máximo cerca de 98%, no máximo cerca de 99% ou 100%. Em algumas modalidades, a porcentagem de complementaridade entre a sequência espaçadora e o ácido nucleico alvo é 100% ao longo da maioria dos seis nucleotídeos 5' contíguos da sequência alvo da fita complementar do ácido nucleico alvo. Em algumas modalidades, a porcentagem de complementaridade entre a sequência espaçadora e o ácido nucleico alvo é de pelo menos 60% ao longo de cerca de 20 nucleotídeos contíguos. Em algumas modalidades, o comprimento da sequência espaçadora e do ácido nucleico alvo pode diferir em 1 a 6 nucleotídeos, o que pode ser considerado como uma protuberância ou protuberâncias.

[00214] Em algumas modalidades, a sequência espaçadora é desenhada ou escolhida usando um programa de computador. O programa de computador pode usar variáveis, tal como temperatura prevista de fusão, formação de estrutura secundária, temperatura prevista de emparelhamento, identidade de sequência, contexto genômico, acessibilidade da cromatina, % de GC, frequência de ocorrência genômica (por exemplo, sequências que são idênticas ou são similares, mas variam em um ou mais pontos como resultado de não correspondências, inserção ou deleção), nível de metilação, presença de SNPs e similares.

Sequência de Repetição CRISPR Mínima

[00215] Em algumas modalidades, uma sequência de repetição CRISPR mínima é uma sequência com pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% de identidade de sequência com uma sequência de repetição CRISPR de referência (por exemplo, crRNA de *S. pyogenes*).

[00216] Em algumas modalidades, uma sequência de repetição CRISPR mínima possui nucleotídeos que podem hibridar com uma sequência de tracrRNA mínima em uma célula. A sequência de repetição CRISPR mínima e uma sequência de tracrRNA mínima formam um dúplex, isto é, uma estrutura de fita dupla de base emparelhada. Juntas, a sequência de repetição CRISPR mínima e a sequência de tracrRNA mínima se ligam ao polipeptídeo direcionado ao local. Pelo menos uma parte da sequência de repetição CRISPR mínima hibrida com a sequência de tracrRNA mínima. Em algumas modalidades, pelo menos uma parte da sequência de repetição CRISPR mínima tem pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% de complementaridade com a sequência de tracrRNA mínima. Em algumas modalidades, pelo menos uma parte da sequência de repetição CRISPR mínima tem no máximo cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% de complementaridade com a sequência de tracrRNA mínima.

[00217] A sequência de repetição CRISPR mínima pode ter um comprimento de cerca de 7 nucleotídeos a cerca de 100 nucleotídeos. Por exemplo, o comprimento da sequência de repetição CRISPR mínima é de cerca de 7 nucleotídeos (nt) a cerca de 50 nt, de cerca de 7 nt a cerca de 40 nt, de cerca de 7 nt a cerca de 30 nt, de cerca de 7

nt a cerca de 25 nt, de cerca de 7 nt a cerca de 20 nt, de cerca de 7 nt a cerca de 15 nt, de cerca de 8 nt a cerca de 40 nt, de cerca de 8 nt a cerca de 30 nt, de cerca de 8 nt a cerca de 25 nt, de cerca de 8 nt a cerca de 20 nt, de cerca de 8 nt a cerca de 15 nt, de cerca de 15 nt a cerca de 100 nt, de cerca de 15 nt a cerca de 80 nt, de cerca de 15 nt a cerca de 50 nt, de cerca de 15 nt a cerca de 40 nt, de cerca de 15 nt a cerca de 30 nt ou de cerca de 15 nt a cerca de 25 nt. Em algumas modalidades, a sequência de repetição CRISPR mínima é aproximadamente 9 nucleotídeos em comprimento. Em algumas modalidades, a sequência de repetição CRISPR mínima é aproximadamente 12 nucleotídeos em comprimento.

[00218] Em algumas modalidades, a sequência de repetição CRISPR mínima é pelo menos cerca de 60% idêntica a uma sequência de repetição CRISPR mínima de referência (por exemplo, crRNA de tipo selvagem de *S. pyogenes*) ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos. Por exemplo, a sequência de repetição CRISPR mínima é pelo menos cerca de 65% idêntica, pelo menos cerca de 70% idêntica, pelo menos cerca de 75% idêntica, pelo menos cerca de 80% idêntica, pelo menos cerca de 85% idêntica, pelo menos cerca de 90% idêntica, pelo menos cerca de 95% idêntica, pelo menos cerca de 98% idêntica, pelo menos cerca de 99% idêntica ou 100% idêntica a uma sequência de repetição CRISPR mínima de referência ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos.

Sequência de tracrRNA Mínima

[00219] Em algumas modalidades, uma sequência de tracrRNA mínima é uma sequência com pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% de identidade de sequência com uma sequência de tracrRNA de referência (por exemplo, tracrRNA de tipo selvagem de *S. pyogenes*).

[00220] Em algumas modalidades, uma sequência de tracrRNA mínima possui nucleotídeos que hibridam com uma sequência de repetição CRISPR mínima em uma célula. Uma sequência de tracrRNA mínima e uma sequência de repetição CRISPR mínima formam um dúplex, isto é, uma estrutura de fita dupla de base emparelhada. Juntas, a sequência de tracrRNA mínima e a repetição CRISPR mínima se ligam a um polipeptídeo direcionado ao local. Pelo menos uma parte da sequência de tracrRNA mínima pode hibridar com a sequência de repetição CRISPR mínima. Em algumas modalidades, a sequência de tracrRNA mínima é pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% complementar à sequência de repetição CRISPR mínima.

[00221] A sequência de tracrRNA mínima pode ter um comprimento de cerca de 7 nucleotídeos a cerca de 100 nucleotídeos. Por exemplo, a sequência de tracrRNA mínima pode ser de cerca de 7 nucleotídeos (nt) a cerca de 50 nt, de cerca de 7 nt a cerca de 40 nt, de cerca de 7 nt a cerca de 30 nt, de cerca de 7 nt a cerca de 25 nt, de cerca de 7 nt a cerca de 20 nt, de cerca de 7 nt a cerca de 15 nt, de cerca de 8 nt a cerca de 40 nt, de cerca de 8 nt a cerca de 30 nt, de cerca de 8 nt a cerca de 25 nt, de cerca de 8 nt a cerca de 20 nt, de cerca de 8 nt a cerca de 15 nt, de cerca de 15 nt a cerca de 100 nt, de cerca de 15 nt a cerca de 80 nt, de cerca de 15 nt a cerca de 50 nt, de cerca de 15 nt a cerca de 40 nt, de cerca de 15 nt a cerca de 30 nt ou de cerca de 15 nt a cerca de 25 nt de comprimento. Em algumas modalidades, a sequência de tracrRNA mínima é aproximadamente 9 nucleotídeos em comprimento. Em algumas modalidades, a sequência de tracrRNA mínima é aproximadamente 12 nucleotídeos. Em algumas modalidades, o tracrRNA mínimo consiste em tracrRNA nt 23-48 descrito em Jinek *et al. Science, 337*(6096):816-821 (2012).

[00222] Em algumas modalidades, a sequência de tracrRNA mínima é pelo menos cerca de 60% idêntica a uma sequência de tracrRNA mínima de referência (por exemplo, tracrRNA de tipo selvagem de *S. pyogenes*) ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos. Por exemplo, a sequência de tracrRNA mínima é pelo menos cerca de 65% idêntica, cerca de 70% idêntica, cerca de 75% idêntica, cerca de 80% idêntica, cerca de 85% idêntica, cerca de 90% idêntica, cerca de 95% idêntica, cerca de 98% idêntica, cerca de 99% idêntica ou 100% idêntica a uma sequência de tracrRNA mínima de referência ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos.

[00223] Em algumas modalidades, o dúplex entre o RNA de CRISPR mínimo e o tracrRNA mínimo tem uma hélice dupla. Em algumas modalidades, o dúplex entre o RNA de CRISPR mínimo e o tracrRNA mínimo tem pelo menos cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos. Em algumas modalidades, o dúplex entre o RNA de CRISPR mínimo e o tracrRNA mínimo tem no máximo cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos.

[00224] Em algumas modalidades, o dúplex tem uma não correspondência (isto é, as duas fitas do dúplex não são 100% complementares). Em algumas modalidades, o dúplex tem pelo menos cerca de 1, 2, 3, 4 ou 5 ou mais não correspondências. Em algumas modalidades, o dúplex tem no máximo cerca de 1, 2, 3, 4 ou 5 ou mais não correspondências. Em algumas modalidades, o dúplex não tem mais que 2 não correspondências.

Protuberâncias

[00225] Em algumas modalidades, existe uma "protuberância" no dúplex entre o RNA de CRISPR mínimo e o tracrRNA mínimo. A protuberância é uma região não emparelhada de nucleotídeos dentro do dúplex. Em algumas modalidades, a protuberância contribui para a

ligação do dúplex ao polipeptídeo direcionado ao local. Uma protuberância possui, de um lado do dúplex, um 5'-XXXY-3' não emparelhado, onde X é qualquer purina e Y tem um nucleotídeo que pode formar um par alternado com um nucleotídeo na fita oposta, e uma região nucleotídica não emparelhada no outro lado do dúplex. O número de nucleotídeos não emparelhados nos dois lados do dúplex pode ser diferente.

[00226] Em um exemplo, a protuberância tem uma purina não emparelhada (por exemplo, adenina) na sequência de repetição CRISPR mínima da protuberância. Em algumas modalidades, uma protuberância tem um 5'-AAGY-3' não emparelhado da fita da sequência de tracrRNA mínima da protuberância, em que Y tem um nucleotídeo que pode formar um emparelhamento alternado com um nucleotídeo na sequência de repetição CRISPR mínima.

[00227] Em algumas modalidades, uma protuberância no lado da repetição CRISPR mínima do dúplex tem pelo menos 1, 2, 3, 4 ou 5 ou mais nucleotídeos não emparelhados. Em algumas modalidades, uma protuberância no lado da repetição CRISPR mínima do dúplex tem no máximo 1, 2, 3, 4 ou 5 ou mais nucleotídeos não emparelhados. Em algumas modalidades, uma protuberância no lado da repetição CRISPR mínima do dúplex tem 1 nucleotídeo não emparelhado.

[00228] Em algumas modalidades, uma protuberância no lado da sequência de tracrRNA mínima do dúplex tem pelo menos 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos não emparelhados. Em algumas modalidades, uma protuberância no lado da sequência de tracrRNA mínima do dúplex tem no máximo 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos não emparelhados. Em algumas modalidades, uma protuberância em um segundo lado do dúplex (por exemplo, o lado da sequência de tracrRNA mínima do dúplex) tem 4 nucleotídeos não emparelhados.

[00229] Em algumas modalidades, uma protuberância tem pelo menos um emparelhamento alternado. Em algumas modalidades, uma protuberância tem no máximo um emparelhamento alternado. Em algumas modalidades, uma protuberância tem pelo menos um nucleotídeo de purina. Em algumas modalidades, uma protuberância tem pelo menos 3 nucleotídeos de purina. Em algumas modalidades, uma sequência da protuberância tem pelo menos 5 nucleotídeos de purina. Em algumas modalidades, uma sequência da protuberância tem pelo menos um nucleotídeo de guanina. Em algumas modalidades, uma sequência da protuberância tem pelo menos um nucleotídeo de adenina.

Grampos

[00230] Em várias modalidades, um ou mais grampos estão localizados a 3' do tracrRNA mínimo na sequência de tracrRNA 3'.

[00231] Em algumas modalidades, o grampo começa pelo menos cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ou 20 ou mais nucleotídeos a 3' do último nucleotídeo emparelhado no dúplex da repetição CRISPR mínima e sequência de tracrRNA mínima. Em algumas modalidades, o grampo pode começar no máximo cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos a 3' do último nucleotídeo emparelhado no dúplex da repetição CRISPR mínima e sequência de tracrRNA mínima.

[00232] Em algumas modalidades, um grampo tem pelo menos cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ou 20 ou mais nucleotídeos consecutivos. Em algumas modalidades, um grampo tem no máximo cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ou mais nucleotídeos consecutivos.

[00233] Em algumas modalidades, um grampo tem um dinucleotídeo CC (isto é, dois nucleotídeos de citosina consecutivos).

[00234] Em algumas modalidades, um grampo tem nucleotídeos duplexados (por exemplo, nucleotídeos em um grampo, hibridados

conjuntamente). Por exemplo, um grampo tem um dinucleotídeo CC que é hibridado com um dinucleotídeo GG em um dúplex de grampo da sequência de tracrRNA 3'.

[00235] Um ou mais grampos podem interagir com as regiões que interagem com o RNA guia de um polipeptídeo direcionado ao local.

[00236] Em algumas modalidades, existem dois ou mais grampos e, em algumas modalidades, existem três ou mais grampos.

Sequência de tracrRNA 3'

[00237] Em algumas modalidades, uma sequência de tracrRNA 3' tem uma sequência com pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100% de identidade de sequência com uma sequência de tracrRNA de referência (por exemplo, um tracrRNA de *S. pyogenes*).

[00238] Em algumas modalidades, a sequência de tracrRNA 3' tem um comprimento de cerca de 6 nucleotídeos a cerca de 100 nucleotídeos. Por exemplo, a sequência de tracrRNA 3' pode ter um comprimento de cerca de 6 nucleotídeos (nt) a cerca de 50 nt, de cerca de 6 nt a cerca de 40 nt, de cerca de 6 nt a cerca de 30 nt, de cerca de 6 nt a cerca de 25 nt, de cerca de 6 nt a cerca de 20 nt, de cerca de 6 nt a cerca de 15 nt, de cerca de 8 nt a cerca de 40 nt, de cerca de 8 nt a cerca de 30 nt, de cerca de 8 nt a cerca de 25 nt, de cerca de 8 nt a cerca de 20 nt, de cerca de 8 nt a cerca de 15 nt, de cerca de 15 nt a cerca de 100 nt, de cerca de 15 nt a cerca de 80 nt, de cerca de 15 nt a cerca de 50 nt, de cerca de 15 nt a cerca de 40 nt, de cerca de 15 nt a cerca de 30 nt ou de cerca de 15 nt a cerca de 25 nt. Em algumas modalidades, a sequência de tracrRNA 3' tem um comprimento de aproximadamente 14 nucleotídeos.

[00239] Em algumas modalidades, a sequência de tracrRNA 3' é pelo menos cerca de 60% idêntica a uma sequência de tracrRNA 3' de

referência (por exemplo, sequência de tracrRNA 3' de tipo selvagem de *S. pyogenes*) ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos. Por exemplo, a sequência de tracrRNA 3' é pelo menos cerca de 60% idêntica, cerca de 65% idêntica, cerca de 70% idêntica, cerca de 75% idêntica, cerca de 80% idêntica, cerca de 85% idêntica, cerca de 90% idêntica, cerca de 95% idêntica, cerca de 98% idêntica, cerca de 99% idêntica ou 100% idêntica a uma sequência de tracrRNA 3' de referência (por exemplo, sequência de tracrRNA 3' de tipo selvagem de *S. pyogenes*) ao longo de uma extensão de pelo menos 6, 7 ou 8 nucleotídeos contíguos.

[00240] Em algumas modalidades, uma sequência de tracrRNA 3' tem mais do que uma região dúplex (por exemplo, grampo, região hibridada). Em algumas modalidades, uma sequência de tracrRNA 3' tem duas regiões dúplex.

[00241] Em algumas modalidades, a sequência de tracrRNA 3' tem uma estrutura de alça de haste. Em algumas modalidades, uma estrutura de alça de haste no tracrRNA 3' tem pelo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ou 20 ou mais nucleotídeos. Em algumas modalidades, a estrutura de alça de haste no tracrRNA 3' tem no máximo 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais nucleotídeos. Em algumas modalidades, a estrutura de alça de haste tem uma fração funcional. Por exemplo, a estrutura de alça de haste pode ter um aptâmero, uma ribozima, um grampo que interage com a proteína, uma matriz CRISPR, um íntron ou um éxon. Em algumas modalidades, a estrutura de alça de haste tem pelo menos cerca de 1, 2, 3, 4 ou 5 ou mais frações funcionais. Em algumas modalidades, a estrutura de alça de haste tem no máximo cerca de 1, 2, 3, 4 ou 5 ou mais frações funcionais.

[00242] Em algumas modalidades, o grampo na sequência de tracrRNA 3' tem um domínio P. Em algumas modalidades, o domínio P tem uma região de fita dupla no grampo.

Sequência de Extensão do tracrRNA

[00243] Em algumas modalidades, uma sequência de extensão do tracrRNA pode ser fornecida se o tracrRNA estiver no contexto de guias de molécula única ou guias de molécula dupla. Em algumas modalidades, uma sequência de extensão do tracrRNA tem um comprimento de cerca de 1 nucleotídeo a cerca de 400 nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA tem um comprimento superior a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380 ou 400 nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA tem um comprimento de cerca de 20 a cerca de 5000 ou mais nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA tem um comprimento de mais de 1000 nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA tem um comprimento inferior a 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400 ou mais nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA pode ter um comprimento inferior a 1000 nucleotídeos. Em algumas modalidades, uma sequência de extensão do tracrRNA tem menos de 10 nucleotídeos de comprimento. Em algumas modalidades, uma sequência de extensão do tracrRNA tem 10 a 30 nucleotídeos de comprimento. Em algumas modalidades, a sequência de extensão do tracrRNA tem 30-70 nucleotídeos de comprimento.

[00244] Em algumas modalidades, a sequência de extensão do tracrRNA tem uma fração funcional (por exemplo, uma sequência de controle de estabilidade, ribozima, sequência de ligação à endorribonuclease). Em algumas modalidades, a fração funcional tem um segmento terminador da transcrição (isto é, uma sequência de terminação da transcrição). Em algumas modalidades, a fração

funcional tem um comprimento total de cerca de 10 nucleotídeos (nt) a cerca de 100 nucleotídeos, de cerca de 10 nt a cerca de 20 nt, de cerca de 20 nt a cerca de 30 nt, de cerca de 30 nt a cerca de 40 nt, de cerca de 40 nt a cerca de 50 nt, de cerca de 50 nt a cerca de 60 nt, de cerca de 60 nt a cerca de 70 nt, de cerca de 70 nt a cerca de 80 nt, de cerca de 80 nt a cerca de 90 nt, ou de cerca de 90 nt a cerca de 100 nt, de cerca de 15 nt a cerca de 80 nt, de cerca de 15 nt a cerca de 50 nt, de cerca de 15 nt a cerca de 40 nt, de cerca de 15 nt a cerca de 30 nt ou de cerca de 15 nt a cerca de 25 nt. Em algumas modalidades, a fração funcional funciona em uma célula eucariótica. Em algumas modalidades, a fração funcional funciona em uma célula procariótica. Em algumas modalidades, a fração funcional funciona em células eucarióticas e procarióticas.

[00245] Exemplos não limitativos de frações funcionais de extensão do tracrRNA adequados incluem uma cauda poliadenilada 3', uma sequência ribocomutadora (por exemplo, para permitir estabilidade regulada e/ou acessibilidade regulada por proteínas e complexos de proteínas), uma sequência que forma um dúplex de dsRNA (isto é, um grampo), uma sequência que direciona o RNA para uma localização subcelular (por exemplo, núcleo, mitocôndrias, cloroplastos e similares), uma modificação ou sequência que fornece monitorização (por exemplo, conjugação direta a uma molécula fluorescente, conjugação a uma fração que facilita a detecção fluorescente, uma sequência que permite a detecção fluorescente, etc.) e/ou uma modificação ou sequência que fornece um local de ligação para proteínas (por exemplo, proteínas que atuam no DNA, incluindo ativadores transcricionais, repressores transcricionais, DNA metiltransferases, desmetilases de DNA, histonas acetiltransferases, histona deacetilases e similares). Em algumas modalidades, uma sequência de extensão do tracrRNA tem um local de ligação ao iniciador ou um índice molecular (por exemplo,

sequência de código de barras). Em algumas modalidades, a sequência de extensão do tracrRNA tem um ou mais marcadores de afinidade.

Sequência Ligadora Guia de Molécula Única

[00246] Em algumas modalidades, a sequência ligadora de um ácido nucleico guia de molécula única tem um comprimento de cerca de 3 nucleotídeos a cerca de 100 nucleotídeos. Em Jinek *et al.*, *supra*, por exemplo, foi usado um simples "tetraloop" de 4 nucleotídeos (-GAAA-), Science, 337 (6096): 816-821 (2012). Um ligante ilustrativo tem um comprimento de cerca de 3 nucleotídeos (nt) a cerca de 90 nt, de cerca de 3 nt a cerca de 80 nt, de cerca de 3 nt a cerca de 70 nt, de cerca de 3 nt a cerca de 60 nt, de cerca de 3 nt a cerca de 50 nt, de cerca de 3 nt a cerca de 40 nt, de cerca de 3 nt a cerca de 30 nt, de cerca de 3 nt a cerca de 20 nt, de cerca de 3 nt a cerca de 10 nt. Por exemplo, o ligante pode ter um comprimento de cerca de 3 nt a cerca de 5 nt, de cerca de 5 nt a cerca de 10 nt, de cerca de 10 nt a cerca de 15 nt, de cerca de 15 nt a cerca de 20 nt, de cerca de 20 nt a cerca de 25 nt, de cerca de 25 nt a cerca de 30 nt, de cerca de 30 nt a cerca de 35 nt, de cerca de 35 nt a cerca de 40 nt, de cerca de 40 nt a cerca de 50 nt, de cerca de 50 nt a cerca de 60 nt, de cerca de 60 nt a cerca de 70 nt, de cerca de 70 nt a cerca de 80 nt, de cerca de 80 nt a cerca de 90 nt ou de cerca de 90 nt a cerca de 100 nt. Em algumas modalidades, o ligante de um ácido nucleico guia de molécula única está entre 4 e 40 nucleotídeos. Em algumas modalidades, um ligante tem pelo menos cerca de 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 ou 7000 ou mais nucleotídeos. Em algumas modalidades, um ligante tem no máximo cerca de 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 ou 7000 ou mais nucleotídeos.

[00247] Os ligantes podem ter uma variedade de sequências, embora, em algumas modalidades, o ligante não tenha sequências que

tenham extensas regiões de homologia com outras porções do RNA guia, o que pode causar ligação intramolecular que pode interferir com outras regiões funcionais do guia. Em Jinek et al., supra, foi usada uma sequência simples de 4 nucleotídeos -GAAA-, Science, 337(6096):816-821 (2012), mas inúmeras outras sequências, incluindo sequências mais longas, podem igualmente ser usadas.

[00248] Em algumas modalidades, a sequência do ligante tem uma fração funcional. Por exemplo, a sequência do ligante pode ter um ou mais elementos, incluindo um aptâmero, uma ribozima, um grampo que interage com proteínas, um local de ligação a proteínas, uma matriz CRISPR, um íntron ou um éxon. Em algumas modalidades, a sequência do ligante tem pelo menos cerca de 1, 2, 3, 4 ou 5 ou mais frações funcionais. Em algumas modalidades, a sequência do ligante tem no máximo cerca de 1, 2, 3, 4 ou 5 ou mais frações funcionais.

[00249] Em algumas modalidades, uma localização genômica direcionada por gRNAs de acordo com a presente invenção pode estar no, dentro ou perto do locus endógeno da albumina em um genoma, por exemplo, genoma humano. Exemplos de RNAs de guia direcionados a esses locais incluem as sequências espaçadoras listadas nas Tabelas 3 ou 4 (por exemplo, sequências espaçadoras de qualquer uma das SEQ ID NOs: 18-44 e 104) e o local de corte associado Cas9 ou Cpf1. Por exemplo, um gRNA incluindo uma sequência espaçadora da SEQ ID NO: 18 pode incluir a sequência espaçadora UAAUUUUCUUUUGCGCACUA (SEQ ID NO: 105). Como é entendido pela pessoa de habilidade comum na técnica, cada RNA guia é desenhado para incluir uma sequência espaçadora complementar à sua sequência alvo genômica. Por exemplo, cada uma das sequências espaçadoras listadas nas Tabelas 3 ou 4 pode ser colocada em uma única quimera de RNA ou um crRNA (juntamente com um tracrRNA correspondente). Consultar Jinek *et al*., Science, 337, 816-821 (2012) e

Deltcheva *et al*., Nature, 471, 602-607 (2011).

DNA Doador ou Molde doador

[00250] Os polipeptídeos direcionados ao local, como uma endonuclease de DNA, podem introduzir quebras da fita dupla ou quebras da fita única em ácidos nucleicos, por exemplo, DNA genômico. A quebra da fita dupla pode estimular as vias endógenas de reparação de DNA de uma célula (por exemplo, reparação dependente de homologia (HDR) ou junção de extremidade não homóloga ou junção de extremidade não homóloga alternativa (A-NHEJ) ou junção de extremidade mediada por micro-homologia (MMEJ). NHEJ pode reparar o ácido nucleico alvo clivado sem a necessidade de um molde homólogo. Às vezes, isso pode resultar em pequenas deleções ou inserções (indels) no ácido nucleico alvo no local da clivagem e pode levar à interrupção ou alteração da expressão gênica. HDR, que também é conhecida como recombinação homóloga (HR), pode ocorrer quando um molde de reparação homólogo, ou doador, está disponível.

[00251] O molde doador homólogo tem sequências homólogas às sequências que flanqueiam o local de clivagem do ácido nucleico alvo. A cromátide irmã é geralmente usada pela célula como molde de reparação. No entanto, para fins de edição do genoma, o molde de reparação é frequentemente fornecido como um ácido nucleico exógeno, como um plasmídeo, oligonucleotídeo dúplex, oligonucleotídeo de fita simples, oligonucleotídeo de fita dupla ou ácido nucleico viral. Com moldes de doadores exógenos, é comum introduzir uma sequência adicional de ácido nucleico (como um transgene) ou uma modificação (como uma mudança de base única ou múltipla ou uma deleção) entre as regiões flanqueadoras de homologia, de modo que a sequência do ácido nucleico adicional ou alterado também é incorporada no local alvo. A MMEJ resulta em um efeito genético similar a NHEJ, pois pequenas deleções e inserções podem ocorrer no local da

clivagem. A MMEJ usa sequências homólogas de alguns pares de bases que flanqueiam o local de clivagem para obter um resultado preferido de reparação de DNA na junção de extremidades. Em alguns casos, pode ser possível prever resultados prováveis de reparação com base na análise de potenciais micro-homologias nas regiões alvo da nuclease.

[00252] Assim, em alguns casos, a recombinação homóloga é usada para inserir uma sequência polinucleotídica exógena no local de clivagem do ácido nucleico alvo. Uma sequência polinucleotídica exógena é denominada polinucleotídeo doador (ou doador ou sequência doadora ou molde de polinucleotídeo doador) no presente documento. Em algumas modalidades, o polinucleotídeo doador, uma porção do polinucleotídeo doador, uma cópia do polinucleotídeo doador ou uma porção de uma cópia do polinucleotídeo doador é inserido no local de clivagem do ácido nucleico alvo. Em algumas modalidades, o polinucleotídeo doador é uma sequência polinucleotídica exógena, isto é, uma sequência que não ocorre naturalmente no local de clivagem do ácido nucleico alvo.

[00253] Quando uma molécula de DNA exógena é fornecida em concentração suficiente dentro do núcleo de uma célula na qual ocorre a quebra de fita dupla, o DNA exógeno pode ser inserido na quebra de fita dupla durante o processo de reparação de NHEJ e, assim, se tornar uma adição permanente ao genoma. Estas moléculas de DNA exógenas são referidas como moldes doadores em algumas modalidades. Se o molde doador conter uma sequência de codificação para um gene de interesse, tal como um gene FVIII, opcionalmente, juntamente com sequências reguladoras relevantes, tais como promotores, potenciadores, sequências poliA e/ou sequências aceitadoras de splice (também no presente documento referidas como "cassete doador"), o gene de interesse pode ser expresso a partir da

cópia integrada no genoma, resultando em expressão permanente para a vida da célula. Além disso, a cópia integrada do molde de DNA doador pode ser transmitida para as células filhas quando a célula se divide.

[00254] Na presença de concentrações suficientes de um molde de DNA doador que contém sequências de DNA flanqueantes com homologia para a sequência de DNA de ambos os lados da quebra da fita dupla (referido como braços de homologia), o molde de DNA doador pode ser integrado através da via HDR. Os braços de homologia atuam como substratos para a recombinação homóloga entre o molde doador e as sequências de ambos os lados da quebra da fita dupla. Isto pode resultar em uma inserção livre de erros do molde doador, na qual as sequências de ambos os lados da quebra de fita dupla não são alteradas em relação aquelas no genoma não modificado.

[00255] Os doadores fornecidos para edição por HDR variam acentuadamente, mas geralmente contêm a sequência pretendida com braços de homologia flanqueantes pequenos ou grandes para permitir o emparelhamento no DNA genômico. As regiões de homologia que flanqueiam as alterações gênicas introduzidas podem ser 30 pb ou menores, ou tão grandes quanto um cassete de múltiplas quilobases que pode conter promotores, cDNAs, etc. Podem ser usados doadores de oligonucleotídeos de fita simples e dupla. Estes oligonucleotídeos variam em tamanho de menos de 100 nt a muitos kb, embora também seja possível gerar e usar ssDNA mais longo. Os doadores de fita dupla são frequentemente usados, incluindo amplicons de PCR, plasmídeos e mini-círculos. Em geral, foi verificado que um vetor AAV é um meio muito eficaz de entrega de um molde doador, embora o limite de empacotamento para doadores individuais seja < 5kb. A transcrição ativa do doador aumentou a HDR em três vezes, indicando que a inclusão do promotor pode aumentar a conversão. Por outro lado, a metilação CpG do doador pode diminuir a expressão gênica e HDR.

[00256] Em algumas modalidades, o DNA doador pode ser fornecido com a nuclease ou independentemente por uma variedade de métodos diferentes, por exemplo, por transfecção, nanopartícula, microinjeção ou transdução viral. Uma gama de opções de ligação pode ser usada para aumentar a disponibilidade dos doadores para HDR em algumas modalidades. Os exemplos incluem ligar o doador à nuclease, ligar às proteínas de ligação ao DNA que se ligam próximas ou ligar às proteínas que estão envolvidas na ligação ou reparação da extremidade do DNA.

[00257] Além da edição do genoma por NHEJ ou HDR, podem ser realizadas inserções de genes específicos do local que usam a via NHEJ e HR. Uma abordagem combinada pode ser aplicável em determinadas configurações, possivelmente incluindo fronteiras íntron/éxon. A NHEJ pode ser eficaz na ligação no íntron, enquanto a HDR livre erros pode ser mais adequada na região de codificação.

[00258] Em modalidades, uma sequência exógena que se destina a ser inserida em um genoma é um gene do Fator VIII (FVIII) ou seu derivado funcional. O gene exógeno pode incluir uma sequência nucleotídica que codifica uma proteína Fator VIII ou um seu derivado funcional. O derivado funcional de um gene FVIII pode incluir uma sequência de ácido nucleico que codifica um derivado funcional de uma proteína FVIII que possui uma atividade substancial de uma proteína FVIII de tipo selvagem, tal como a proteína FVIII humana de tipo selvagem, por exemplo, pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 95% ou cerca de 100% da atividade que a proteína FVIII de tipo selvagem exibe. Em algumas modalidades, o derivado funcional de uma proteína FVIII pode ter pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou cerca de 99% de identidade de sequência de aminoácidos

com a proteína FVIII, por exemplo, a proteína FVIII de tipo selvagem. Em algumas modalidades, um versado comum na técnica pode usar vários métodos conhecidos na área para testar a funcionalidade ou atividade de um composto, por exemplo, peptídeo ou proteína. O derivado funcional da proteína FVIII também pode incluir qualquer fragmento da proteína FVIII de tipo selvagem ou fragmento de uma proteína FVIII modificada que tenha modificação conservativa em um ou mais resíduos de aminoácidos na proteína FVIII de tipo selvagem de comprimento total. Assim, em algumas modalidades, o derivado funcional de uma sequência de ácido nucleico de um gene FVIII pode ter pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou cerca de 99% de identidade de sequência de ácido nucleico com o gene FVIII, por exemplo, o gene FVIII de tipo selvagem.

[00259] Em algumas modalidades em que a inserção de um gene do Fator VIII (FVIII) ou seu derivado funcional está envolvida, um cDNA do gene do Fator VIII ou seu derivado funcional pode ser inserido no genoma de um paciente tendo o gene FVIII defeituoso ou em suas sequências reguladoras. Nesse caso, um DNA doador ou molde doador pode ser um cassete de expressão ou construto de vetor tendo a sequência que codifica o gene do Fator VIII ou seu derivado funcional, por exemplo, sequência de cDNA. Em algumas modalidades, pode ser usado o vetor de expressão contendo uma sequência que codifica uma proteína Fator VIII modificada, tal como FVIII-BDD, que é descrita em outras partes nas descrições.

[00260] Em algumas modalidades, de acordo com qualquer um dos moldes doadores no presente documento descritos compreendendo um cassete doador, o cassete doador é flanqueado em um ou ambos os lados por um local alvo de gRNA. Por exemplo, tal molde doador pode

compreender um cassete doador com um local alvo de gRNA 5' do cassete doador e/ou um local alvo de gRNA 3' do cassete doador. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 5' do cassete doador. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 3' do cassete doador. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador, e os dois locais alvo de gRNA compreendem a mesma sequência. Em algumas modalidades, o molde doador compreende pelo menos um local alvo de gRNA, e o pelo menos um local alvo de gRNA no molde doador compreende a mesma sequência que um local alvo de gRNA em um locus alvo no qual o cassete doador do molde doador é para ser integrado. Em algumas modalidades, o molde doador compreende pelo menos um local alvo de gRNA, e o pelo menos um local alvo de gRNA no molde doador compreende o complemento reverso de um local alvo de gRNA em um locus alvo no qual o cassete doador do molde doador é para ser integrado. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador, e os dois locais alvo de gRNA no molde doador compreendem a mesma sequência que um local alvo de gRNA em um locus alvo no qual o cassete doador do molde doador é para ser integrado. Em algumas modalidades, o molde doador compreende um cassete doador com um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador, e os dois locais alvo de gRNA no molde doador compreendem o complemento reverso de um local alvo de gRNA em um locus alvo no qual o cassete

doador do molde doador é para ser integrado.

Ácido nucleico que codifica um polipeptídeo direcionado ao local ou endonuclease de DNA

[00261] Em algumas modalidades, os métodos de edição do genoma e composições podem, portanto, usar uma sequência de ácido nucleico (ou oligonucleotídeo) que codifica um polipeptídeo direcionado ao local ou endonuclease de DNA. A sequência de ácido nucleico que codifica o polipeptídeo direcionado ao local pode ser DNA ou RNA. Se a sequência de ácido nucleico que codifica o polipeptídeo direcionado ao local é RNA, ela pode ser ligada covalentemente a uma sequência de gRNA ou existir como uma sequência separada. Em algumas modalidades, uma sequência peptídica do polipeptídeo direcionado ao local ou endonuclease de DNA pode ser usada em vez da sua sequência de ácido nucleico.

Vetores

[00262] Em outro aspecto, a presente invenção fornece um ácido nucleico tendo uma sequência de nucleotídeos que codifica um ácido nucleico direcionado ao genoma da invenção, um polipeptídeo direcionado ao local da invenção e/ou qualquer ácido nucleico ou molécula proteinácea necessária para realizar as modalidades dos métodos da invenção. Em algumas modalidades, esse ácido nucleico é um vetor (por exemplo, um vetor de expressão recombinante).

[00263] Os vetores de expressão contemplados incluem, mas não estão limitados a, vetores virais baseados no vírus vaccinia, poliovírus, adenovírus, vírus adeno-associado, SV40, vírus do herpes simplex, vírus da imunodeficiência humana, retrovírus (por exemplo, Vírus da Leucemia Murina, vírus da necrose do baço e vetores derivados de retrovírus, como o Vírus do Sarcoma de Rous, Vírus do Sarcoma de Harvey, vírus da leucose aviária, um lentivírus, vírus da imunodeficiência humana, vírus do sarcoma mieloproliferativo e vírus do

tumor mamário) e outros vetores recombinantes. Outros vetores contemplados para células alvo eucarióticas incluem, mas não estão limitados aos, vetores pXT1, pSG5, pSVK3, pBPV, pMSG e pSVLSV40 (Pharmacia). Os vetores adicionais contemplados para células alvo eucarióticas incluem, mas não estão limitados aos, vetores pCTx-1, pCTx-2 e pCTx-3. Outros vetores podem ser usados desde que sejam compatíveis com a célula hospedeira.

[00264] Em algumas modalidades, um vetor tem um ou mais elementos de controle da transcrição e/ou tradução. Dependendo do sistema hospedeiro/vetor usado, qualquer de um número de elementos adequados de controle da transcrição e tradução, incluindo promotores constitutivos e induzíveis, elementos potenciadores da transcrição, terminadores de transcrição, etc., pode ser usado no vetor de expressão. Em algumas modalidades, o vetor é um vetor de autoinativação que inativa as sequências virais ou os componentes da maquinaria CRISPR ou outros elementos.

[00265] Exemplos não limitativos de promotores eucarióticos adequados (isto é, promotores funcionais em uma célula eucariótica) incluem aqueles do citomegalovírus (CMV) precoce imediato, timidina quinase do vírus herpes simplex (HSV), SV40 precoce e tardio, repetições terminais longas (LTRs) de retrovírus, promotor do fator de alongamento 1 humano (EF1), um construto híbrido tendo o potenciador do citomegalovírus (CMV) fundido ao promotor da beta-actina de galinha (CAG), promotor do vírus de célula tronco murina (MSCV), promotor do locus de fosfoglicerato quinase-1 (PGK) e metalotioneína-I de camundongo.

[00266] Para expressar pequenos RNAs, incluindo RNAs guia usados em conexão com a endonuclease Cas, vários promotores, tais como promotores de RNA polimerase III, incluindo, por exemplo, U6 e H1, podem ser vantajosos. Descrições e parâmetros para melhorar o

uso de tais promotores são conhecidos na técnica, e informações e abordagens adicionais estão sendo regularmente descritas; ver, por exemplo, Ma, H. *et al.*, *Molecular Therapy - Nucleic Acids 3*, e161 (2014) doi:10.1038/mtna.2014.12.

[00267] O vetor de expressão também pode conter um local de ligação ao ribossomo para iniciar a tradução e um terminador de transcrição. O vetor de expressão também pode incluir sequências apropriadas para amplificar a expressão. O vetor de expressão também pode incluir sequências de nucleotídeos que codificam marcadores não nativos (por exemplo, marcador de histidina, marcador de hemaglutinina, proteína fluorescente verde, etc.) que são fundidos ao polipeptídeo direcionado ao local, resultando assim em uma proteína de fusão.

[00268] Em algumas modalidades, um promotor é um promotor induzível (por exemplo, um promotor de choque térmico, promotor regulado por tetraciclina, promotor regulado por esteroides, promotor regulado por metal, promotor regulado pelo receptor de estrogênio, etc.). Em algumas modalidades, um promotor é um promotor constitutivo (por exemplo, promotor CMV, promotor UBC). Em algumas modalidades, o promotor é um promotor espacialmente restrito e/ou temporalmente restrito (por exemplo, um promotor específico de tecido, um promotor específico de tipo de célula, etc.). Em algumas modalidades, um vetor não possui um promotor para que pelo menos um gene seja expresso em uma célula hospedeira se o gene for expresso, depois de inserido em um genoma, sob um promotor endógeno presente no genoma.

POLIPEPTÍDEO DIRIGIDO AO LOCAL OU ENDONUCLEASE DE DNA

[00269] As modificações do DNA alvo devido a NHEJ e/ou HDR podem levar a, por exemplo, mutações, deleções, alterações,

integrações, correção de genes, substituição de genes, marcação de genes, inserção de transgene, deleção de nucleotídeos, interrupção de genes, translocações e/ou mutação de gene. O processo de integração do ácido nucleico não nativo no DNA genômico é um exemplo de edição do genoma.

[00270] Um polipeptídeo direcionado ao local é uma nuclease usada na edição do genoma para clivar o DNA. O local direcionado pode ser administrado a uma célula ou a um paciente como: um ou mais polipeptídeos ou um ou mais mRNAs que codificam o polipeptídeo.

[00271] No contexto de um sistema CRISPR/Cas ou CRISPR/Cpf1, o polipeptídeo direcionado ao local pode-se ligar a um RNA guia que, por sua vez, especifica o local no DNA alvo ao qual o polipeptídeo é direcionado. Em modalidades dos sistemas CRISPR/Cas ou CRISPR/Cpf1 descritas no presente documento, o polipeptídeo direcionado ao local é uma endonuclease, tal como uma endonuclease de DNA.

[00272] Em algumas modalidades, um polipeptídeo direcionado ao local tem uma pluralidade de domínios de clivagem de ácidos nucleicos (isto é, nuclease). Dois ou mais domínios de clivagem de ácidos nucleicos podem ser ligados através de um ligante. Em algumas modalidades, o ligante tem um ligante flexível. Os ligantes podem ter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40 ou mais aminoácidos de comprimento.

[00273] As enzimas Cas9 de tipo selvagem de ocorrência natural têm dois domínios de nuclease, um domínio de nuclease HNH e um domínio RuvC. no presente documento, a "Cas9" refere-se a Cas9s de ocorrência natural e recombinante. As enzimas Cas9 no presente documento contempladas têm um domínio nuclease HNH ou do tipo HNH e/ou um domínio nuclease RuvC ou do tipo RuvC.

[00274] Os domínios HNH ou do tipo HNH têm uma dobra do tipo

McrA. Os domínios HNH ou do tipo HNH têm duas fitas β antiparalelas e uma hélice α. Os domínios HNH ou do tipo HNH têm um local de ligação a metal (por exemplo, um local de ligação a cátion divalente). Os domínios HNH ou do tipo HNH podem clivar uma fita de um ácido nucleico alvo (por exemplo, a fita complementar da fita alvo do crRNA).

[00275] Os domínios RuvC ou do tipo RuvC têm uma dobra RNaseH ou do tipo RNaseH. Os domínios RuvC/RNaseH estão envolvidos em um conjunto diversificado de funções baseadas em ácidos nucleicos, incluindo a atuação em RNA e DNA. O domínio RNaseH possui 5 fitas β rodeadas por uma pluralidade de hélices α. Os domínios RuvC/RNaseH ou do tipo RuvC/RNaseH têm um local de ligação a metal (por exemplo, um local de ligação a cátion divalente). Os domínios RuvC/RNaseH ou do tipo RuvC/RNaseH podem clivar uma fita de um ácido nucleico alvo (por exemplo, a fita não complementar de um DNA alvo de fita dupla).

[00276] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos tendo pelo menos 10%, pelo menos 15%, pelo menos 20%, pelo menos 30%, pelo menos 40%, pelo menos 50%, pelo menos 60%, pelo menos 70%, pelo menos 75%, pelo menos 80%, pelo menos 85%, pelo menos 90%, pelo menos 95%, pelo menos 99% ou 100% de identidade de sequência de aminoácidos com um local exemplificativo de tipo selvagem do polipeptídeo direcionado [por exemplo, Cas9 de *S. pyogenes*, US2014/0068797 Sequência ID No. 8 ou Sapranauskas *et al., Nucleic Acids Res, 39*(21): 9275-9282 (2011)] e vários outros polipeptídeos direcionados ao local).

[00277] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos tendo pelo menos 10%, pelo menos 15%, pelo menos 20%, pelo menos 30%, pelo menos 40%, pelo menos 50%, pelo menos 60%, pelo menos 70%, pelo menos 75%, pelo menos 80%, pelo menos 85%, pelo menos 90%, pelo menos 95%, pelo

menos 99% ou 100% de identidade de sequência de aminoácidos com o domínio de nuclease de um local exemplificativo de tipo selvagem do polipeptídeo direcionado [por exemplo, Cas9 de *S. pyogenes*, *supra*).

[00278] Em algumas modalidades, um polipeptídeo direcionado ao local tem pelo menos 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo de mais de 10 aminoácidos contíguos. Em algumas modalidades, um polipeptídeo direcionado ao local tem no máximo: 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo de mais de 10 aminoácidos contíguos. Em algumas modalidades, um polipeptídeo direcionado ao local tem pelo menos: 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo de 10 aminoácidos contíguos em um domínio de nuclease HNH do polipeptídeo direcionado ao local. Em algumas modalidades, um polipeptídeo direcionado ao local tem no máximo: 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo de 10 aminoácidos contíguos em um domínio de nuclease HNH do polipeptídeo direcionado ao local. Em algumas modalidades, um polipeptídeo direcionado ao local tem pelo menos: 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo de 10 aminoácidos contíguos em um domínio de nuclease RuvC do polipeptídeo direcionado ao local. Em algumas modalidades, um polipeptídeo direcionado ao local tem no máximo: 70, 75, 80, 85, 90, 95, 97, 99 ou 100% de identidade com um polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*) ao longo

de 10 aminoácidos contíguos em um domínio de nuclease RuvC do polipeptídeo direcionado ao local.

[00279] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma forma modificada de um polipeptídeo direcionado ao local exemplificativo do tipo selvagem. A forma modificada do polipeptídeo direcionado ao local exemplificativo do tipo selvagem tem uma mutação que reduz a atividade de clivagem do ácido nucleico do polipeptídeo direcionado ao local. Em algumas modalidades, a forma modificada do polipeptídeo direcionado ao local exemplificativo do tipo selvagem tem menos do que 90%, menos do que 80%, menos do que 70%, menos do que 60%, menos do que 50%, menos do que 40%, menos do que 30%, menos do que 20%, menos do que 10%, menos do que 5% ou menos do que 1% da atividade de clivagem do ácido nucleico do polipeptídeo direcionado ao local exemplificativo do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*). A forma modificada do polipeptídeo direcionado ao local não pode ter atividade substancial de clivagem de ácido nucleico. Quando um polipeptídeo direcionado ao local é uma forma modificada que não possui atividade substancial de clivagem de ácido nucleico, é no presente documento referido como "enzimaticamente inativo".

[00280] Em algumas modalidades, a forma modificada do polipeptídeo direcionado ao local tem uma mutação tal que pode induzir uma quebra de fita simples (SSB) em um ácido nucleico alvo (por exemplo, cortando apenas uma das estruturas principais de açúcar-fosfato de uma fita dupla do ácido nucleico alvo). Em algumas modalidades, a mutação resulta em menos do que 90%, menos do que 80%, menos do que 70%, menos do que 60%, menos do que 50%, menos do que 40%, menos do que 30%, menos do que 20%, menos do que 10%, menos do que 5% ou menos do que 1% da atividade de clivagem do ácido nucleico em um ou mais da pluralidade de domínios

de clivagem de ácido nucleico do polipeptídeo direcionado ao local do tipo selvagem (por exemplo, Cas9 de *S. pyogenes*, *supra*). Em algumas modalidades, a mutação resulta em um ou mais da pluralidade de domínios de clivagem do ácido nucleico, mantendo a capacidade de clivar a fita complementar do ácido nucleico alvo, mas reduzindo a sua capacidade de clivar a fita não complementar do ácido nucleico alvo. Em algumas modalidades, a mutação resulta em um ou mais da pluralidade de domínios de clivagem do ácido nucleico, mantendo a capacidade de clivar a fita não complementar do ácido nucleico alvo, mas reduzindo a sua capacidade de clivar a fita complementar do ácido nucleico alvo. Por exemplo, resíduos no polipeptídeo exemplificativo Cas9 de *S. pyogenes* do tipo selvagem, como Asp10, His840, Asn854 e Asn856, são mutados para inativar um ou mais da pluralidade de domínios de clivagem de ácidos nucleicos (por exemplo, domínios de nuclease). Em algumas modalidades, os resíduos a serem mutados correspondem aos resíduos Asp10, His840, Asn854 e Asn856 no polipeptídeo exemplificativo Cas9 de *S. pyogenes* do tipo selvagem (por exemplo, como determinado pela sequência e/ou alinhamento estrutural). Exemplos não limitativos de mutações incluem D10A, H840A, N854A ou N856A. Um perito na técnica irá reconhecer que outras mutações além das substituições de alanina são adequadas.

[00281] Em algumas modalidades, a mutação D10A é combinada com uma ou mais mutações H840A, N854A ou N856A para produzir um polipeptídeo direcionado ao local, substancialmente sem atividade de clivagem de DNA. Em algumas modalidades, a mutação H840A é combinada com uma ou mais mutações D10A, N854A ou N856A para produzir um polipeptídeo direcionado ao local, substancialmente sem atividade de clivagem de DNA. Em algumas modalidades, a mutação N854A é combinada com uma ou mais mutações H840A, D10A ou N856A para produzir um polipeptídeo direcionado ao local,

substancialmente sem atividade de clivagem de DNA. Em algumas modalidades, a mutação N856A é combinada com uma ou mais mutações H840A, N854A ou D10A para produzir um polipeptídeo direcionado ao local, substancialmente sem atividade de clivagem de DNA. Os polipeptídeos direcionados ao local que possuem um domínio de nuclease substancialmente inativo são referidos como "nickases".

[00282] Em algumas modalidades, variantes de endonucleases guiadas por RNA, por exemplo, Cas9, podem ser usadas para aumentar a especificidade da edição do genoma mediada por CRISPR. A Cas9 de tipo selvagem é tipicamente guiada por um único RNA guia projetado para hibridar com uma sequência especificada de ~ 20 nucleotídeos na sequência alvo (tal como um locus genômico endógeno). No entanto, várias não correspondências podem ser toleradas entre o RNA guia e o locus alvo, reduzindo efetivamente o comprimento da homologia necessária no local alvo para, por exemplo, apenas 13 nt de homologia, e assim resultando em potencial elevado para ligação e clivagem de ácido nucleico de fita dupla pelo complexo CRISPR/Cas9 em outras partes do genoma alvo – também conhecida como clivagem fora do alvo. Como as variantes de nickase da Cas9 cortam apenas uma fita, de modo a criar uma quebra de fita dupla é necessário que um par de nickases se ligue nas proximidades e em fitas opostas do ácido nucleico alvo, criando assim um par de cortes, que é o equivalente a uma quebra de fita dupla. Isto requer que dois RNAs guia separados - um para cada nickase - se liguem muito próximos e em fitas opostas do ácido nucleico alvo. Este requisito basicamente duplica o comprimento mínimo de homologia necessário para que a quebra da fita dupla ocorra, reduzindo assim a probabilidade de um evento de clivagem de fita dupla ocorrer em outro local do genoma, onde os dois locais de RNA guia - se existirem - são improváveis estar suficientemente próximos um do outro para permitir a quebra da fita dupla. Como descrito na técnica, as

nickases também podem ser usadas para promover HDR versus NHEJ. A HDR pode ser usada para introduzir alterações selecionadas nos locais alvo no genoma através do uso de sequências específicas de doadores que mediam efetivamente as alterações desejadas. Descrições de vários sistemas CRISPR/Cas para uso na edição de genes podem ser encontradas, por exemplo, na publicação de pedido de patente internacional número WO2013/176772 e na *Nature Biotechnology 32*, 347–355 (2014), e referências no presente documento citadas.

[00283] Em algumas modalidades, o polipeptídeo direcionado ao local (por exemplo, variante, mutado, enzimaticamente inativo e/ou polipeptídeo direcionado ao local condicionalmente enzimaticamente inativo) tem como alvo o ácido nucleico. Em algumas modalidades, o polipeptídeo direcionado ao local (por exemplo, variante, mutado, enzimaticamente inativo e/ou endorribonuclease condicionalmente enzimaticamente inativa) tem como alvo o DNA. Em algumas modalidades, o polipeptídeo direcionado ao local (por exemplo, variante, mutado, enzimaticamente inativo e/ou endorribonuclease condicionalmente enzimaticamente inativa) tem como alvo o RNA.

[00284] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma ou mais sequências não nativas (por exemplo, o polipeptídeo direcionado ao local é uma proteína de fusão).

[00285] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*), um domínio de ligação a ácido nucleico e dois domínios de clivagem de ácido nucleico (isto é, um domínio HNH e um domínio RuvC).

[00286] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de

identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*) e dois domínios de clivagem de ácido nucleico (isto é, um domínio HNH e um domínio RuvC).

[00287] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*) e dois domínios de clivagem de ácido nucleico, em que um ou ambos dos domínios de clivagem de ácido nucleico têm pelo menos 50% de identidade de aminoácidos com um domínio nuclease de Cas9 de uma bactéria (por exemplo, *S. pyogenes*).

[00288] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*), dois domínios de clivagem de ácido nucleico (isto é, um domínio HNH e um domínio RuvC) e sequência não nativa (por exemplo, um sinal de localização nuclear) ou um ligante que liga o polipeptídeo direcionado ao local a uma sequência não nativa.

[00289] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*), dois domínios de clivagem de ácido nucleico (isto é, um domínio HNH e um domínio RuvC), em que o polipeptídeo direcionado ao local tem uma mutação em um ou em ambos os domínios de clivagem de ácido nucleico que reduz a atividade de clivagem dos domínios de nuclease em pelo menos 50%.

[00290] Em algumas modalidades, o polipeptídeo direcionado ao local tem uma sequência de aminoácidos com pelo menos 15% de identidade de aminoácidos em relação a uma Cas9 de uma bactéria (por exemplo, *S. pyogenes*), e dois domínios de clivagem de ácido nucleico (isto é, um domínio HNH e um domínio RuvC), em que um dos domínios

da nuclease tem mutação do ácido aspártico 10 e/ou em que um dos domínios da nuclease tem mutação da histidina 840, e em que a mutação reduz a atividade de clivagem do(s) domínio(s) da nuclease em pelo menos 50%.

[00291] Em algumas modalidades, o um ou mais polipeptídeos direcionados ao local, por exemplo, endonucleases de DNA, incluem duas nickases que juntas efetuam uma quebra da fita dupla em um locus específico no genoma, ou quatro nickases que juntas efetuam duas quebras da fita dupla em loci específicos no genoma. Em alternativa, um polipeptídeo direcionado ao local, por exemplo, endonuclease de DNA, afeta uma quebra de fita dupla em um locus específico do genoma.

[00292] Em algumas modalidades, um polinucleotídeo que codifica um polipeptídeo direcionado ao local pode ser usado para editar o genoma. Em algumas dessas modalidades, o polinucleotídeo que codifica um polipeptídeo direcionado ao local é otimizado quanto a códon de acordo com métodos padrão na técnica para expressão na célula que contém o DNA alvo de interesse. Por exemplo, se o ácido nucleico alvo pretendido estiver em uma célula humana, um polinucleotídeo humano otimizado quanto a códon que codifica Cas9 é contemplado para uso na produção do polipeptídeo Cas9.

[00293] A seguir, são apresentados alguns exemplos de polipeptídeos direcionados ao local que podem ser usados em várias modalidades das invenções.

Sistema de Endonuclease CRISPR

[00294] Um locus genômico CRISPR (Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas) pode ser encontrado nos genomas de muitos procariontes (por exemplo, bactérias e arqueobactérias). Nos procariontes, o locus CRISPR codifica produtos que funcionam como um tipo de sistema imunológico

para ajudar a defender os procariontes contra invasores externos, como vírus e fagos. Existem três estágios da função do locus CRISPR: integração de novas sequências no locus CRISPR, expressão do RNA CRISPR (crRNA) e silenciamento do ácido nucleico invasor externo. Foram identificados cinco tipos de sistemas CRISPR (por exemplo, Tipo I, Tipo II, Tipo III, Tipo U e Tipo V).

[00295] Um locus CRISPR inclui um número de sequências curtas de repetição referidas como "repetições". Quando expressas, as repetições podem formar estruturas secundárias em grampo (por exemplo, grampos) e/ou ter sequências não estruturadas de fita simples. As repetições geralmente ocorrem em grupos e frequentemente divergem entre espécies. As repetições são regularmente espaçadas com sequências intermediárias únicas chamadas de "espaçadores", resultando em uma arquitetura de locus de repetição-espaçador-repetição. Os espaçadores são idênticos ou possuem alta homologia com sequências invasoras externas conhecidas. Uma unidade de repetição do espaçador codifica um crisprRNA (crRNA), que é processado em uma forma madura da unidade de repetição do espaçador. Um crRNA tem uma "semente" ou sequência espaçadora que está envolvida no direcionamento de um ácido nucleico alvo (na forma que ocorre naturalmente nos procariontes, a sequência espaçadora tem como alvo o ácido nucleico do invasor externo). Uma sequência espaçadora está localizada na extremidade 5' ou 3' do crRNA.

[00296] Um locus CRISPR também possui sequências polinucleotídicas que codificam genes Associados a CRISPR (Cas). Os genes Cas codificam endonucleases envolvidas na biogênese e nos estágios de interferência da função de crRNA em procariontes. Alguns genes Cas têm estruturas secundárias e/ou terciárias homólogas.

Sistemas CRISPR Tipo II

[00297] A biogênese do crRNA em um sistema CRISPR Tipo II na natureza requer um RNA transativador de CRISPR (tracrRNA). O tracrRNA é modificado pela RNaseIII endógena e depois hibrida com uma repetição de crRNA na matriz pré-crRNA. A RNaseIII endógena é recrutada para clivar o pré-crRNA. Os crRNAs clivados são submetidos a aparamento de exoribonuclease para produzir a forma madura de crRNA (por exemplo, aparamento 5'). O tracrRNA permanece hibridizado com o crRNA, e o tracrRNA e o crRNA se associam a um polipeptídeo direcionado ao local (por exemplo, Cas9). O crRNA do complexo crRNA-tracrRNA-Cas9 guia o complexo para um ácido nucleico alvo ao qual o crRNA pode hibridar. A hibridação do crRNA no ácido nucleico alvo ativa Cas9 para a clivagem direcionada de ácidos nucleicos. O ácido nucleico alvo em um sistema CRISPR Tipo II é referido como motivo adjacente ao protoespaçador (PAM). Na natureza, o PAM é essencial para facilitar a ligação de um polipeptídeo direcionado ao local (por exemplo, Cas9) ao ácido nucleico alvo. Os sistemas tipo II (também chamados de Nmeni ou CASS4) são subdivididos em Tipo II-A (CASS4) e II-B (CASS4a). Jinek *et al.*, *Science, 337*(6096):816-821 (2012) mostraram que o sistema CRISPR/Cas9 é útil para edição programável do genoma por RNA, e o número da publicação internacional do pedido de patente WO 2013/176772 fornece numerosos exemplos e aplicações do sistema de endonuclease CRISPR/Cas para edição de genes específicos de local.

Sistemas CRISPR Tipo V

[00298] Os sistemas CRISPR Tipo V apresentam várias diferenças importantes em relação aos sistemas Tipo II. Por exemplo, Cpf1 é uma endonuclease guiada por RNA simples que, em contraste com os sistemas Tipo II, não possui tracrRNA. De fato, as matrizes CRISPR associadas a Cpf1 são processadas em crRNAs maduros sem a necessidade de um tracrRNA adicional de transativação. A matriz

CRISPR Tipo V é processada em crRNAs curtos maduros de 42-44 nucleotídeos de comprimento, com cada crRNA maduro começando com 19 nucleotídeos da repetição direta seguidos por 23-25 nucleotídeos da sequência espaçadora. Em contraste, os crRNAs maduros nos sistemas Tipo II começam com 20-24 nucleotídeos da sequência espaçadora, seguidos por cerca de 22 nucleotídeos da repetição direta. Além disso, a Cpf1 usa um motivo adjacente ao protoespaçador rico em T, de modo que os complexos Cpf1-crRNA clivam com eficiência o DNA alvo precedido por um PAM rico em T curto, que contrasta com o PAM rico em G após o DNA alvo para sistemas Tipo II. Assim, os sistemas Tipo V clivam em um ponto que é distante do PAM, enquanto os sistemas Tipo II clivam em um ponto que é adjacente ao PAM. Além disso, em contraste com os sistemas Tipo II, a Cpf1 cliva o DNA através de uma quebra progressiva de fita dupla de DNA com uma sobreposição 5' de 4 ou 5 nucleotídeos. Os sistemas Tipo II clivam através de quebra da fita dupla abrupta. Similar aos sistemas Tipo II, a Cpf1 contém um domínio previsto de endonuclease similar a RuvC, mas não possui um segundo domínio de endonuclease HNH, o que contrasta com os sistemas Tipo II.

Genes/Polipeptídeos Cas e Motivos Adjacentes ao Protoespaçador

[00299] Polipeptídeos CRISPR/Cas exemplificativos incluem os polipeptídeos Cas9 na Fig. 1 de Fonfara *et al.*, *Nucleic Acids Research, 42*: 2577-2590 (2014). O sistema de nomeação de genes CRISPR/Cas passou por uma extensa rescrita desde a descoberta dos genes Cas. A Fig. 5 de Fonfara, *supra*, fornece sequências de PAM para os polipeptídeos Cas9 de várias espécies.

Complexos de um Ácido Nucleico Direcionado ao Genoma e de um Polipeptídeo Direcionado ao Local

[00300] Um ácido nucleico direcionado ao genoma interage com um polipeptídeo direcionado ao local (por exemplo, uma nuclease guiada

por ácido nucleico como Cas9), formando assim um complexo. O ácido nucleico direcionado ao genoma (por exemplo, gRNA) guia o polipeptídeo direcionado ao local para um ácido nucleico alvo.

[00301] Como afirmado precedentemente, em algumas modalidades, o polipeptídeo direcionado ao local e o ácido nucleico direcionado ao genoma podem ser administrados separadamente a uma célula ou a um paciente. Por outro lado, em algumas outras modalidades, o polipeptídeo direcionado ao local pode ser pré-complexado com um ou mais RNAs guia, ou um ou mais crRNA em conjunto com um tracrRNA. O material pré-complexado pode então ser administrado a uma célula ou a um paciente. Tal material pré-complexado é conhecido como partícula de ribonucleoproteína (RNP).

SISTEMAS PARA EDIÇÃO DO GENOMA

[00302] São no presente documento fornecidos sistemas para edição do genoma, em particular, para inserir um gene do Fator VIII (FVIII) ou seu derivado funcional no genoma de uma célula. Estes sistemas podem ser usados nos métodos no presente documento descritos, tais como para editar o genoma de uma célula e para tratar um indivíduo, por exemplo, um paciente de Hemofilia A.

[00303] Em algumas modalidades, é no presente documento fornecido um sistema compreendendo (a) uma endonuclease de ácido desoxirribonucleico (DNA) ou ácido nucleico que codifica a referida endonuclease de DNA; (b) um RNA guia (gRNA) direcionado ao locus da albumina no genoma de uma célula; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional. Em algumas modalidades, o gRNA tem como alvo o íntron 1 do gene da albumina. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104.

[00304] Em algumas modalidades, é no presente documento

proporcionado um sistema compreendendo (a) uma endonuclease de ácido desoxirribonucleico (DNA) ou ácido nucleico que codifica a referida endonuclease de DNA; (b) um RNA guia (gRNA) compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 21, 22, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[00305] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00306] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão em uma célula

hospedeira. Em algumas modalidades, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizado quanto a códon para expressão em uma célula humana.

[00307] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, o sistema compreende um ácido nucleico que codifica a endonuclease de DNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon para expressão em uma célula hospedeira. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon para expressão em uma célula humana. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é DNA, tal como um plasmídeo de DNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um RNA, tal como mRNA.

[00308] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV). Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para um gRNA no sistema. Em algumas modalidades, o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA do genoma da célula para um gRNA no sistema.

[00309] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um

lipossomo ou nanopartícula lipídica. Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA. Em algumas modalidades, o lipossomo ou nanopartícula lipídica é uma nanopartícula lipídica. Em algumas modalidades, o sistema compreende uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA.

[00310] Em algumas modalidades, de acordo com qualquer um dos sistemas no presente documento descritos, a endonuclease de DNA é complexada com o gRNA, formando um complexo de ribonucleoproteína (RNP).

MÉTODOS DE EDIÇÃO DO GENOMA

[00311] É no presente documento fornecido um método para edição do genoma, em particular, inserir um gene do Fator VIII (FVIII) ou seu derivado funcional no genoma de uma célula. Este método pode ser usado para tratar um indivíduo, por exemplo, um paciente de Hemofilia A e, nesse caso, uma célula pode ser isolada do paciente ou de um doador separado. Em seguida, o DNA cromossômico da célula é editado usando os materiais e métodos no presente documento descritos.

[00312] Em algumas modalidades, uma estratégia de modificar ("knock-in") envolve a modificação ("knocking-in") de uma sequência de codificação de FVIII, por exemplo, um gene FVIII de tipo selvagem (por exemplo, o gene FVIII humano de tipo selvagem), um cDNA de FVIII, um minigene (com potenciador e promotor natural ou sintético, um ou mais éxons e íntrons naturais ou sintéticos e sinal de 3'UTR e poliadenilação natural ou sintético) ou um gene FVIII modificado, em uma sequência genômica. Em algumas modalidades, a sequência genômica em que a sequência que codifica FVIII é inserida está no, dentro ou perto do locus da albumina.

[00313] São no presente documento fornecidos métodos para modificar ("knock-in") um gene FVIII ou seu derivado funcional em um genoma. Em um aspecto, a presente invenção fornece a inserção de uma sequência de ácido nucleico de um gene FVIII, isto é, uma sequência de ácido nucleico que codifica uma proteína FVIII ou seu derivado funcional em um genoma de uma célula. Nas modalidades, o gene FVIII pode codificar uma proteína FVIII do tipo selvagem. O derivado funcional de uma proteína FVIII pode incluir um peptídeo que tem uma atividade substancial da proteína FVIII de tipo selvagem, por exemplo, pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 95% ou cerca de 100% da atividade que a proteína FVIII de tipo selvagem exibe. Em algumas modalidades, um versado comum na técnica pode usar vários métodos conhecidos na área para testar a funcionalidade ou atividade de um composto, por exemplo, peptídeo ou proteína. Em algumas modalidades, o derivado funcional da proteína FVIII também pode incluir qualquer fragmento da proteína FVIII de tipo selvagem ou fragmento de uma proteína FVIII modificada que tenha modificação conservativa em um ou mais resíduos de aminoácidos na proteína FVIII de tipo selvagem de comprimento total. Em algumas modalidades, o derivado funcional da proteína FVIII também pode incluir qualquer modificação (ou qualquer uma das modificações), por exemplo, deleção, inserção e/ou mutação de um ou mais aminoácidos que não afetam substancialmente negativamente a funcionalidade da proteína FVIII de tipo selvagem. Assim, em algumas modalidades, o derivado funcional de uma sequência de ácido nucleico de um gene FVIII pode ter pelo menos cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95%, cerca de 96%, cerca de 97%, cerca de 98% ou cerca de 99% de identidade de sequência de ácido nucleico com o gene FVIII.

[00314] Em algumas modalidades, um gene FVIII ou seu derivado funcional é inserido em uma sequência genômica em uma célula. Em algumas modalidades, o local de inserção está no ou dentro do locus da albumina no genoma da célula. O método de inserção usa um ou mais gRNAs direcionando o primeiro íntron (ou íntron 1) do gene da albumina. Em algumas modalidades, o DNA doador é um DNA de fita simples ou dupla tendo um gene FVIII ou seu derivado funcional.

[00315] Em algumas modalidades, os métodos de edição do genoma usam uma endonuclease de DNA, como um sistema CRISPR/Cas, para introduzir geneticamente (modificar, "knock-in") um gene FVIII ou seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, um seu homólogo, recombinação da molécula de ocorrência natural, sua versão otimizada quanto a códon ou modificada, e combinações de qualquer uma das precedentes. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00316] Em algumas modalidades, a célula sujeita à edição do genoma possui uma ou mais mutações no genoma, o que resulta na redução da expressão do gene endógeno FVIII em comparação com a expressão em um normal que não possui essa(s) mutação(ões). A célula normal pode ser uma célula saudável ou de controle que é originada (ou isolada) de um indivíduo diferente que não possui defeitos no gene FVIII. Em algumas modalidades, a célula sujeita à edição do

genoma pode ser originada (ou isolada) de um indivíduo que está com necessidade do tratamento da afecção ou distúrbio relacionado ao gene FVIII, por exemplo, Hemofilia A. Portanto, em algumas modalidades, a expressão do gene endógeno FVIII nessa célula é de cerca de 10%, cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90% ou cerca de 100% reduzida em comparação com a expressão do gene FVIII endógeno na célula normal.

[00317] Em algumas modalidades, os métodos de edição do genoma conduzem a integração direcionada na região não codificante do genoma de um gene FVIII funcional, por exemplo, uma sequência de codificação de FVIII que está operacionalmente ligada a um promotor fornecido de modo a produzir estavelmente a proteína FVIII *in vivo*. Em algumas modalidades, a integração direcionada de uma sequência de codificação de FVIII ocorre em um íntron do gene da albumina que é altamente expresso no tipo de célula de interesse, por exemplo, hepatócitos ou células endoteliais sinusoidais. Em algumas modalidades, a sequência de codificação de FVIII a ser inserida pode ser uma sequência de codificação do FVIII de tipo selvagem, por exemplo, a sequência de codificação de FVIII humano de tipo selvagem. Em algumas modalidades, a sequência de codificação de FVIII pode ser um derivado funcional de uma sequência de codificação do FVIII de tipo selvagem, como a sequência de codificação de FVIII humano de tipo selvagem.

[00318] Em um aspecto, a presente invenção propõe a inserção de uma sequência de ácido nucleico de um gene FVIII ou seu derivado funcional em um genoma de uma célula. Nas modalidades, a sequência de codificação de FVIII a ser inserida é uma sequência de codificação de FVIII modificada. Em algumas modalidades, na sequência de codificação de FVIII modificada, o domínio B da sequência de

codificação de FVIII de tipo selvagem é deletado e substituído por um peptídeo ligante chamado "ligante SQ" (sequência de aminoácidos SFSQNPPVLKRHQR – SEQ ID NO: 1). Este FVIII com o domínio B deletado (FVIII-BDD) é bem conhecido na técnica e possui atividade biológica equivalente ao FVIII de comprimento total. Em algumas modalidades, um FVIII com domínio B deletado é preferível a um FVIII de comprimento total devido ao seu tamanho menor (4371 pb *versus* 7053 pb). Assim, em algumas modalidades, a sequência de codificação de FVIII-BDD sem o peptídeo sinal de FVIII e contendo uma sequência aceitadora de splice na extremidade 5' (Terminal N da sequência de codificação de FVIII) é integrada especificamente no íntron 1 do gene da albumina nos hepatócitos de mamíferos, incluindo humanos. A transcrição desta sequência codificadora de FVIII modificada do promotor da albumina pode resultar em um pré-mRNA que contém o éxon 1 da albumina, parte do íntron 1 e a sequência do gene FVIII-BDD integrada. Quando esse pré-mRNA passa pelo processo de splicing natural para remover os íntrons, a maquinaria de splicing pode unir o doador de splice no lado 3' do éxon 1 da albumina ao próximo aceitador de splice disponível, que será o aceitador de splice na extremidade 5' da sequência de codificação de FVIII-BDD do DNA doador inserido. Isto pode resultar em um mRNA maduro contendo o éxon 1 da albumina fundido com a sequência de codificação madura para FVIII-BDD. O éxon 1 da albumina codifica o peptídeo sinal mais 2 aminoácidos adicionais e 1/3 de um códon que em humanos codifica normalmente a sequência de proteína DAH no terminal N da albumina. Portanto, em algumas modalidades após a clivagem prevista do peptídeo sinal da albumina durante a secreção da célula, uma proteína FVIII-BDD pode ser gerada com 3 resíduos de aminoácidos adicionais adicionados ao terminal N, resultando na sequência de aminoácidos –DAHATRRYY (SEQ ID NO: 98)- no terminal N da proteína FVIII-BDD. Como o 3°

desses 3 aminoácidos (sublinhado) é codificado parcialmente no final do éxon 1 e parcialmente no molde de DNA doador FVIII-BDD, é possível selecionar a identidade do 3° resíduo de aminoácido adicional para ser Leu, Pro, His, Gln ou Arg. Entre essas opções, Leu é preferível em algumas modalidades, uma vez que Leu é o menos complexo molecularmente e, portanto, tem menor probabilidade de formar um novo epítopo de célula T, resultando na sequência de aminoácidos – DALATRRYY- no terminal N da proteína FVIII-BDD. Em alternativa, o molde de DNA doador pode ser projetado para deletar o 3° resíduo, resultando na sequência de aminoácidos DALTRRYY no terminal N da proteína FVIII-BDD. Em alguns casos, a adição de aminoácidos adicionais à sequência de uma proteína nativa pode aumentar o risco de imunogenicidade. Portanto, em algumas modalidades em que uma análise in silico para prever a potencial imunogenicidade das 2 opções possíveis para o terminal N do FVIII-BDD demonstra que a exclusão de 1 resíduo (DALTRRYY) possui uma pontuação mais baixa de imunogenicidade, esse pode ser um modelo preferido pelo menos em algumas modalidades.

[00319] Em algumas modalidades, uma sequência de DNA que codifica FVIII-BDD na qual o uso de códons foi otimizado pode ser usada para melhorar a expressão em células de mamíferos (a chamada otimização de códons). Diferentes algoritmos de computador também estão disponíveis na área para realizar a otimização de códons e geram sequências distintas de DNA. Exemplos de algoritmos de otimização de códon disponíveis comercialmente são aqueles empregues pelas empresas ATUM e GeneArt (parte da Thermo Fisher Scientific). A otimização de códons da sequência de codificação do FVIII demonstrou melhorar significativamente a expressão do FVIII após a entrega baseada em genes a camundongos (Nathwani AC, Gray JT, Ng CY, et al. Blood. 2006; 107(7):2653-2661; Ward NJ, Buckley SM, Waddington

SN, et al. Blood. 2011; 117(3):798-807; Radcliffe PA, Sion CJ, Wilkes FJ, et al. Gene Ther. 2008;15(4):289-297).

[00320] Em algumas modalidades, a homologia ou identidade de sequência entre a sequência de codificação de FVIII-BDD que foi otimizada quanto a códons por algoritmos diferentes e a sequência nativa de FVIII (como presente no genoma humano) pode variar de cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90%, cerca de 95% ou 100%. Em algumas modalidades, a sequência de codificação de FVIII-BDD otimizada quanto a códon tem entre cerca de 75% a cerca de 79% de homologia ou identidade de sequência com a sequência nativa de FVIII. Em algumas modalidades, a sequência de codificação de FVIII-BDD otimizada quanto a códon tem cerca de 70%, cerca de 71%, cerca de 72%, cerca de 73%, cerca de 74%, cerca de 75%, cerca de 76%, cerca de 77%, cerca de 78%, cerca de 79% ou cerca de 80% de homologia ou identidade de sequência com a sequência nativa de FVIII.

[00321] Em algumas modalidades, um molde doador ou construto doador é preparado para conter uma sequência de DNA que codifica FVIII-BDD. Em algumas modalidades, um molde de DNA doador é projetado para conter uma sequência de codificação de FVIII-BDD humana otimizada quanto a códon. Em algumas modalidades, a otimização de códon é feita de tal maneira que a sequência na extremidade 5' que codifica o peptídeo sinal de FVIII foi excluída e substituída por uma sequência aceitadora de splice e, além disso, um sinal de poliadenilação é adicionado à extremidade 3' após o códon de terminação do FVIII (MAB8A – SEQ ID NO: 87). A sequência aceitadora de splice pode ser selecionada dentre as sequências aceitadoras de splice conhecidas de genes conhecidos ou pode ser usada uma sequência aceitadora de splice de consenso que é derivada de um

alinhamento de muitas sequências aceitadoras de splice conhecidas na área. Em algumas modalidades, uma sequência aceitadora de splice de genes altamente expressos é usada, uma vez que se pensa que essas sequências fornecem ótima eficiência de splicing. Em algumas modalidades, a sequência aceitadora de splice de consenso é composta por um local de ramificação com a sequência de consenso T/CNC/TT/CA/GAC/T (SEQ ID NO: 99) seguido dentro de 20 pb com um trato de polipirimidina (C ou T) de 10 a 12 bases seguido por AG>G/A, em que > é a localização do limite íntron/éxon. Em uma modalidade preferida, uma sequência sintética aceitadora de splice (ctgacctcttctcttcctcccacag – SEQ ID NO: 2) é usada. Em outra modalidade preferida, a sequência aceitadora de splice nativa do gene da albumina do limite íntron 1/éxon 2 humano (TTAACAATCCTTTTTTTTCTTCCCTTGCC**CAG**– SEQ ID NO: 3) ou camundongo (ttaaatatgttgtgtggtttttctctccctgtttccacag– SEQ ID NO: 4) é usada.

[00322] A sequência de poliadenilação fornece um sinal para a célula adicionar uma cauda de poliA que é essencial para a estabilidade do mRNA dentro da célula. Em algumas modalidades em que o molde de DNA doador será empacotado em partículas de AAV, é preferível manter o tamanho do DNA empacotado dentro dos limites de empacotamento do AAV, que são preferencialmente menores que cerca de 5 Kb e, idealmente, não mais que 4,7 Kb. Assim, em algumas modalidades, é desejável usar a menor sequência possível de poliA, por exemplo, cerca de 10-mer, cerca de 20-mer, cerca de 30-mer, cerca de 40-mer, cerca de 50-mer ou cerca de 60-mer ou qualquer número intermediário de nucleotídeos dos precedentes. Uma sequência de sinal poli A sintética de consenso foi descrita na literatura (Levitt N, Briggs D, Gil A, Proudfoot NJ. Genes Dev. 1989; 3(7):1019-1025) com a sequência

AATAAAAGATCTTTATTTTCATTAGATCTGTGTGTTGGTTTTTTGTG TG (SEQ ID NO: 5) e é comumente usada em vários vetores de expressão.

[00323] Em algumas modalidades, elementos de sequência adicionais podem ser adicionados ao molde de DNA doador para melhorar a frequência de integração. Um desses elementos são os braços de homologia que são sequências idênticas à sequência de DNA de ambos os lados da quebra de fita dupla no genoma no qual a integração é direcionada para permitir a integração por HDR. Uma sequência do lado esquerdo da quebra da fita dupla (LHA) é anexada à extremidade 5' (terminal N da sequência de codificação de FVIII) do molde de DNA doador e uma sequência do lado direito da quebra da fita dupla (RHA) é anexada à extremidade 3' (terminal C da sequência de codificação de FVIII) do molde de DNA doador, por exemplo, MAB8B (SEQ ID NO: 88).

[00324] Um desenho de molde de DNA doador alternativo que é fornecido em algumas modalidades tem uma sequência complementar à sequência de reconhecimento para o sgRNA que será usado para clivar o local genômico. MAB8C (SEQ ID NO: 89) representa um exemplo deste tipo de moldes de DNA doador. Ao incluir o local de reconhecimento do sgRNA, o molde de DNA doador irá ser clivado pelo complexo sgRNA/Cas9 dentro do núcleo da célula ao qual o molde de DNA doador e o sgRNA/Cas9 foram entregues. A clivagem do molde de DNA doador em fragmentos lineares pode aumentar a frequência de integração em uma quebra de fita dupla pelo mecanismo junção de extremidade não homóloga ou pelo mecanismo HDR. Isto pode ser particularmente benéfico no caso de entrega de moldes de DNA doador empacotados em AAV, porque após a entrega ao núcleo sabe-se que os genomas de AAV concatemerizam para formar moléculas maiores de DNA de fita dupla circular (Nakai et al JOURNAL OF VIROLOGY

2001, vol75 p. 6969–6976). Portanto, em alguns casos, os concatâmeros circulares podem ser doadores menos eficientes para integração em quebras de fita dupla, principalmente pelo mecanismo NHEJ. Foi relatado precedentemente que a eficiência da integração direcionada usando moldes de DNA doador de plasmídeo circular poderia ser aumentada incluindo locais de corte de nuclease de dedo de zinco no plasmídeo (Cristea et al. Biotechnol. Bioeng. 2013; 110: 871-880). Mais recentemente, essa abordagem também foi aplicada usando a nuclease CRISPR/Cas9 (Suzuki et al 2017, *Nature* **540**, 144–149). Embora uma sequência de reconhecimento de sgRNA esteja ativa quando presente em qualquer fita de um molde de DNA doador de fita dupla, é previsto que o uso do complemento reverso da sequência de reconhecimento de sgRNA presente no genoma favorece a integração estável, pois a integração na orientação reversa recria a sequência de reconhecimento de sgRNA que pode ser recortada libertando assim o molde de DNA doador inserido. É previsto que a integração de um tal molde de DNA doador no genoma na orientação direta pelo NHEJ não recrie a sequência de reconhecimento de sgRNA, de modo que o molde de DNA doador integrado não pode ser retirado do genoma. O benefício de incluir sequências de reconhecimento de sgRNA no doador com ou sem braços de homologia sobre a eficiência da integração do molde de DNA doador de FVIII pode ser testado e determinado, por exemplo, em camundongos usando AAV para entrega do doador e LNP para entrega dos componentes CRISPR-Cas9.

[00325] Em algumas modalidades, o molde de DNA doador compreende o gene FVIII ou seu derivado funcional em um cassete doador, de acordo com qualquer uma das modalidades descritas no presente documento, flanqueados em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o molde doador compreende um local alvo de gRNA 5' do cassete doador e/ou um local alvo de gRNA

3' do cassete doador. Em algumas modalidades, o molde doador compreende dois locais alvo de gRNA de flanqueamento e os dois locais alvo de gRNA compreendem a mesma sequência. Em algumas modalidades, o molde doador compreende pelo menos um local alvo de gRNA, e o pelo menos um local alvo de gRNA no molde doador é um local de direcionamento para pelo menos um dos um ou mais gRNAs direcionados ao primeiro íntron do gene da albumina. Em algumas modalidades, o molde doador compreende pelo menos um local alvo de gRNA, e o pelo menos um local alvo de gRNA no molde doador é o complemento reverso de um local alvo para pelo menos um dos um ou mais gRNAs no primeiro íntron do gene da albumina. Em algumas modalidades, o molde doador compreende um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador, e os dois locais alvo de gRNA no molde doador são direcionados por um ou mais gRNAs direcionados ao primeiro íntron do gene da albumina. Em algumas modalidades, o molde doador compreende um local alvo de gRNA 5' do cassete doador e um local alvo de gRNA 3' do cassete doador, e os dois locais alvo de gRNA no molde doador são o complemento reverso de um local alvo para pelo menos um dos um ou mais gRNAs no primeiro íntron do gene da albumina.

[00326] A inserção de um gene que codifica FVIII em um local alvo, isto é, um local genômico em que o gene que codifica FVIII é inserido, pode estar no locus do gene da albumina endógena ou em suas sequências vizinhas. Em algumas modalidades, o gene que codifica o FVIII é inserido de uma maneira em que a expressão do gene inserido é controlada pelo promotor endógeno do gene da albumina. Em algumas modalidades, o gene que codifica o FVIII é inserido em um dos íntrons do gene da albumina. Em algumas modalidades, o gene que codifica o FVIII é inserido em um dos éxons do gene da albumina. Em algumas modalidades, o gene que codifica o FVIII é inserido em uma

junção do íntron:éxon (ou *vice versa*). Em algumas modalidades, a inserção do gene que codifica o FVIII está no primeiro íntron (ou íntron 1) do locus da albumina. Em algumas modalidades, a inserção do gene que codifica o FVIII não afeta significativamente, por exemplo, regula positivamente ou regula negativamente a expressão do gene da albumina.

[00327] Em modalidades, o local alvo para a inserção de um gene que codifica o FVIII está no, dentro ou perto do gene da albumina endógena. Em algumas modalidades, o local alvo está em uma região intergênica que está a montante do promotor do locus do gene da albumina no genoma. Em algumas modalidades, o local alvo está dentro do locus do gene da albumina. Em algumas modalidades, o local alvo está em um dos íntrons do locus do gene da albumina. Em algumas modalidades, o local alvo está em um dos éxons do locus do gene da albumina. Em algumas modalidades, o local alvo está em uma das junções entre um íntron e um éxon (ou *vice versa*) do locus do gene da albumina. Em algumas modalidades, o local alvo está no primeiro íntron (ou íntron 1) do locus do gene da albumina. Em certas modalidades, o local alvo está pelo menos, aproximadamente ou no máximo 0, 1, 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450 ou 500 ou 550 ou 600 ou 650 pb a jusante do primeiro éxon (isto é, do último ácido nucleico do primeiro éxon) do gene da albumina. Em algumas modalidades, o local alvo está a, pelo menos, cerca de ou no máximo 0,1 kb, cerca de 0,2 kb, cerca de 0,3 kb, cerca de 0,4 kb, cerca de 0,5 kb, cerca de 1 kb, cerca de 1,5 kb, cerca de 2 kb, cerca de 2,5 kb, cerca de 3 kb, cerca de 3,5 kb, cerca de 4 kb, cerca de 4,5 kb ou cerca de 5 kb a montante do primeiro íntron do gene da albumina. Em algumas modalidades, o local alvo está em qualquer lugar entre cerca de 0 pb a cerca de 100 pb a montante, cerca de 101 pb a cerca de 200 pb a montante, cerca de 201 pb a cerca de 300 pb a montante, cerca de 301

pb a cerca de 400 pb a montante, cerca de 401 pb a cerca de 500 pb a montante, cerca de 501 pb a cerca de 600 pb a montante, cerca de 601 pb a cerca de 700 pb a montante, cerca de 701 pb a cerca de 800 pb a montante, cerca de 801 pb a cerca de 900 pb a montante, cerca de 901 pb a cerca de 1000 pb a montante, cerca de 1001 pb a cerca de 1500 pb a montante, cerca de 1501 pb a cerca de 2000 pb a montante, cerca de 2001 pb a cerca de 2500 pb a montante, cerca de 2501 pb a cerca de 3000 pb a montante, cerca de 3001 pb a cerca de 3500 pb a montante, cerca de 3501 pb a cerca de 4000 pb a montante, cerca de 4001 pb a cerca de 4500 pb a montante ou cerca de 4501 pb a cerca de 5000 pb a montante do segundo éxon do gene da albumina. Em algumas modalidades, o local alvo está pelo menos 37 pb a jusante da extremidade (isto é, a extremidade 3') do primeiro éxon do gene da albumina humana no genoma. Em algumas modalidades, o local alvo está pelo menos 330 pb a montante do início (isto é, o início 5') do segundo éxon do gene da albumina humana no genoma.

[00328] Em algumas modalidades, é no presente documento proporcionado um método de edição de um genoma em uma célula, o método compreendendo fornecer o seguinte para a célula: (a) um RNA guia (gRNA) direcionado ao locus da albumina no genoma da célula; (b) uma endonuclease de DNA ou um ácido nucleico que codifica a referida endonuclease de DNA; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional. Em algumas modalidades, o gRNA tem como alvo o íntron 1 do gene da albumina. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104.

[00329] Em algumas modalidades, é no presente documento proporcionado um método de edição de um genoma em uma célula, o método compreendendo fornecer o seguinte para a célula: (a) um gRNA

compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104; (b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 21, 22, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30. Em algumas modalidades, a célula é uma célula humana, por exemplo, uma célula de hepatócito humana.

[00330] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00331] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína

Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão em uma célula. Em algumas modalidades, a célula é uma célula humana.

[00332] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, o método usa um ácido nucleico que codifica a endonuclease de DNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é otimizado quanto a códon para expressão na célula. Em algumas modalidades, a célula é uma célula humana, por exemplo, uma célula de hepatócito humana. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um DNA, tal como um plasmídeo de DNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um RNA, tal como mRNA.

[00333] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV). Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para o gRNA de (a). Em algumas modalidades, o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA do genoma da célula para o gRNA de (a).

[00334] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, a endonuclease de DNA ou ácido nucleico que codifica a

endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA. Em algumas modalidades, o lipossomo ou nanopartícula lipídica é uma nanopartícula lipídica. Em algumas modalidades, o método usa uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA.

[00335] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de ribonucleoproteína (RNP).

[00336] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 17 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, (a) e (b) são fornecidos à célula como uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas

modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA. Em algumas modalidades, (c) é fornecido à célula como um vetor AAV que codifica o molde doador.

[00337] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b). Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

[00338] Em algumas modalidades, de acordo com qualquer um dos métodos de edição do genoma em uma célula no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00339] Em algumas modalidades, é no presente documento fornecido um método para inserir um gene FVIII ou seu derivado funcional no locus da albumina de um genoma celular, compreendendo introduzir na célula (a) uma endonuclease de DNA Cas (por exemplo, Cas9) ou ácido nucleico que codifica a endonuclease de DNA Cas, (b) um gRNA ou ácido nucleico que codifica o gRNA, em que o gRNA é

capaz de guiar a endonuclease de DNA Cas para clivar uma sequência polinucleotídica alvo no locus da albumina e (c) um molde doador, de acordo com qualquer uma das modalidades no presente documento descritas, compreendendo o gene FVIII ou seu derivado funcional. Em algumas modalidades, o método compreende a introdução na célula de um mRNA que codifica a endonuclease de DNA Cas. Em algumas modalidades, o método compreende a introdução na célula de uma LNP de acordo com qualquer uma das modalidades descritas no presente documento, compreendendo i) um mRNA que codifica a endonuclease de DNA Cas e ii) o gRNA. Em algumas modalidades, o molde doador é um molde doador AAV. Em algumas modalidades, o molde doador compreende um cassete doador compreendendo o gene Fator VIII (FVIII) ou seu derivado funcional, em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, os locais alvo de gRNA que flanqueiam o cassete doador são o complemento reverso do local alvo de gRNA no locus da albumina. Em algumas modalidades, a endonuclease de DNA Cas ou ácido nucleico que codifica a endonuclease de DNA Cas e o gRNA ou ácido nucleico que codifica o gRNA são introduzidos na célula após a introdução do molde doador na célula. Em algumas modalidades, a endonuclease de DNA Cas ou ácido nucleico que codifica a endonuclease de DNA Cas e o gRNA ou ácido nucleico que codifica o gRNA são introduzidos na célula um tempo suficiente após a introdução do molde doador na célula para permitir que o molde doador entre no núcleo da célula. Em algumas modalidades, a endonuclease de DNA Cas ou ácido nucleico que codifica a endonuclease de DNA Cas e o gRNA ou ácido nucleico que codifica o gRNA são introduzidos na célula um tempo suficiente após a introdução do molde doador na célula para permitir que o molde doador seja convertido de um genoma de AAV de fita simples em uma molécula de DNA de fita dupla no núcleo da célula.

Em algumas modalidades, a endonuclease de DNA Cas é Cas9.

[00340] Em algumas modalidades, de acordo com qualquer um dos métodos de inserção de um gene FVIII ou seu derivado funcional no locus da albumina de um genoma celular descrito no presente documento, a sequência polinucleotídica alvo está no íntron 1 do gene da albumina. Em algumas modalidades, o gRNA compreende uma sequência espaçadora listada na Tabela 3 ou 4. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 21, 22, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[00341] Em algumas modalidades, é no presente documento fornecido um método para inserir um gene FVIII ou seu derivado funcional no locus da albumina de um genoma celular, compreendendo introduzir na célula (a) uma LNP de acordo com qualquer uma das modalidades descritas neste documento compreendendo i) um mRNA que codifica uma endonuclease de DNA Cas9 e ii) um gRNA, em que o gRNA é capaz de guiar a endonuclease de DNA Cas9 para clivar uma sequência polinucleotídica alvo no locus da albumina e (b) um molde doador AAV, de acordo com qualquer uma das modalidades descritas no presente documento, compreendendo o gene FVIII ou seu derivado funcional. Em algumas modalidades, o molde doador compreende um cassete doador compreendendo o gene Fator VIII (FVIII) ou seu derivado funcional, em que o cassete doador é flanqueado em um ou

nos dois lados por um local alvo de gRNA. Em algumas modalidades, os locais alvo de gRNA que flanqueiam o cassete doador são o complemento reverso do local alvo de gRNA no locus da albumina. Em algumas modalidades, a LNP é introduzida na célula após a introdução do molde doador AAV na célula. Em algumas modalidades, a LNP é introduzida na célula um tempo suficiente após a introdução do molde doador AAV na célula para permitir que o molde doador entre no núcleo da célula. Em algumas modalidades, a LNP é introduzida na célula um tempo suficiente após a introdução do molde doador AAV na célula para permitir que o molde doador seja convertido de um genoma de AAV de fita simples em uma molécula de DNA de fita dupla no núcleo da célula. Em algumas modalidades, uma ou mais (tal como 2, 3, 4, 5 ou mais) introduções adicionais da LNP na célula são realizadas após a primeira introdução da LNP na célula. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 21, 22, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

SELEÇÃO DE SEQUÊNCIA ALVO

[00342] Em algumas modalidades, mudanças na localização do limite 5' e/ou do limite 3' em relação a loci de referência específicos são usadas para facilitar ou potenciar aplicações específicas de edição de genes, que dependem em parte do sistema de endonucleases selecionado para a edição, como adicionalmente descrito e ilustrado no presente documento.

[00343] Em um primeiro aspecto não limitativo dessa seleção de sequência alvo, muitos sistemas de endonucleases têm regras ou critérios que orientam a seleção inicial de locais alvo potenciais para clivagem, como a exigência de um motivo de sequência PAM em uma posição específica adjacente aos locais de clivagem de DNA no caso de endonucleases CRISPR Tipo II ou Tipo V.

[00344] Em outro aspecto não limitativo da seleção ou otimização da sequência alvo, a frequência da atividade "fora do alvo" para uma combinação específica de sequências alvo e endonuclease de edição de genes (isto é, a frequência de DSBs ocorrendo em locais que não a sequência alvo selecionada) é avaliada em relação à frequência da atividade no alvo. Em alguns casos, as células que foram editadas corretamente no locus desejado podem ter uma vantagem seletiva em relação a outras células. Exemplos ilustrativos, mas não limitativos, de uma vantagem seletiva incluem a no presente documentosição de atributos tais como taxas melhoradas de replicação, persistência, resistência a determinadas condições, taxas melhoradas de enxerto ou persistência bem-sucedidos *in vivo* após a introdução em um paciente e outros atributos associados com a manutenção ou aumento do número ou viabilidade de tais células. Em outros casos, as células que foram editadas corretamente no locus desejado podem ser selecionadas positivamente por um ou mais métodos de análise usados para identificar, classificar ou de outra forma selecionar as células que foram editadas corretamente. A vantagem seletiva e os métodos de seleção direcionada podem tirar vantagem do fenótipo associado à correção. Em algumas modalidades, as células podem ser editadas duas ou mais vezes, a fim de criar uma segunda modificação que cria um novo fenótipo que é usado para selecionar ou purificar a população pretendida de células. Esta segunda modificação pode ser criada adicionando um segundo gRNA para um marcador selecionável ou

avaliável. Em alguns casos, as células podem ser editadas corretamente no locus desejado usando um fragmento de DNA que contém o cDNA e também um marcador selecionável.

[00345] Nas modalidades, se qualquer vantagem seletiva é aplicável ou se uma seleção direcionada deve ser aplicada em um caso específico, a seleção da sequência alvo também é orientada pela consideração de frequências fora do alvo, a fim de melhorar a eficácia da aplicação e/ou reduzir o potencial para alterações indesejadas em locais diferentes do alvo desejado. Como descrito mais adiante e ilustrado no presente documento e na técnica, a ocorrência de atividade fora do alvo é influenciada por vários fatores, incluindo semelhanças e diferenças entre o local alvo e vários locais fora do alvo, bem como a endonuclease específica usada. Estão disponíveis ferramentas de bioinformática que auxiliam na previsão de atividade fora do alvo e frequentemente essas ferramentas também podem ser usadas para identificar os locais mais prováveis de atividade fora do alvo, que podem ser então avaliados em contextos experimentais para avaliar frequências relativas para atividade fora do alvo e no alvo, permitindo assim a seleção de sequências que possuem atividades relativas no alvo mais elevadas. Exemplos ilustrativos de tais técnicas são fornecidos neste documento e outros são conhecidos na técnica.

[00346] Outro aspecto da seleção da sequência alvo refere-se a eventos de recombinação homóloga. As sequências que partilham regiões da homologia podem servir como pontos focais para eventos de recombinação homóloga que resultam na deleção de sequências intervenientes. Tais eventos de recombinação ocorrem durante o curso normal da replicação de cromossomos e outras sequências de DNA e também em outros momentos em que as sequências de DNA estão sendo sintetizadas, tais como no caso de reparações de quebras de fita dupla (DSBs), que ocorrem em uma base regular durante o ciclo normal

da replicação celular, mas também pode ser melhorada pela ocorrência de vários eventos (tal como luz UV e outros indutores de quebra de DNA) ou pela presença de certos agentes (como vários indutores químicos). Muitos desses indutores fazem com que as DSBs ocorram indiscriminadamente no genoma, e as DSBs estão regularmente sendo induzidas e reparadas em células normais. Durante a reparação, a sequência original pode ser reconstruída com total fidelidade, no entanto, em alguns casos, pequenas inserções ou deleções (referidas como "indels") são introduzidas no local da DSB.

[00347] As DSBs também podem ser induzidas especificamente em locais específicos, como no caso dos sistemas de endonucleases no presente documento descritos, que podem ser usados para causar eventos de modificação gênica direcionados ou preferenciais em locais cromossômicos selecionados. A tendência de sequências homólogas estarem sujeitas a recombinação no contexto da reparação de DNA (bem como a replicação) pode ser aproveitada em várias circunstâncias e é a base para uma aplicação de sistemas de edição de genes, como CRISPR, no qual a reparação direcionado por homologia é usado para inserir uma sequência de interesse, fornecida através do uso de um polinucleotídeo "doador", em um local cromossômico desejado.

[00348] Regiões de homologia entre sequências particulares, que podem ser pequenas regiões de "micro-homologia" que podem ter até dez pares de bases ou menos, também podem ser usadas para provocar deleções desejadas. Por exemplo, uma única DSB é introduzida em um local que exibe micro-homologia com uma sequência próxima. Durante o curso normal de reparação dessa DSB, um resultado que ocorre com alta frequência é a deleção da sequência intermediária como resultado da recombinação facilitada pela DSB e pelo processo de reparação celular concomitante.

[00349] Em algumas circunstâncias, no entanto, a seleção de

sequências alvo dentro de regiões de homologia também pode dar origem a deleções muito maiores, incluindo fusões de genes (quando as deleções estão em regiões de codificação), que podem ou não ser desejadas, dadas as circunstâncias particulares.

[00350] Os exemplos no presente documento fornecidos ilustram ainda mais a seleção de várias regiões alvo para a criação de DSBs projetadas para inserir um gene que codifica FVIII, bem como a seleção de sequências alvo específicas nessas regiões projetadas para minimizar eventos fora do alvo em relação a eventos no alvo.

INTEGRAÇÃO DIRECIONADA

[00351] Em algumas modalidades, o método no presente documento fornecido é para integrar um gene que codifica FVIII ou um gene funcional de FVIII em um local específico no genoma dos hepatócitos que é referido como "integração direcionada". Em algumas modalidades, a integração direcionada é ativada usando uma nuclease específica da sequência para gerar uma quebra de fita dupla no DNA genômico.

[00352] O sistema CRISPR-Cas usado em algumas modalidades tem a vantagem de que um grande número de alvos genômicos pode ser analisado rapidamente para identificar um desenho ótimo de CRISPR-Cas. O sistema CRISPR-Cas usa uma molécula de RNA chamada RNA guia único (sgRNA) que tem como alvo uma nuclease Cas associada (por exemplo, a nuclease Cas9) a uma sequência específica no DNA. Esse direcionamento ocorre pelo emparelhamento baseado em Watson-Crick entre o sgRNA e a sequência do genoma dentro da sequência de direcionamento de aproximadamente 20 pb do sgRNA. Uma vez ligada ao local alvo, a nuclease Cas cliva as duas fitas do DNA genômico, criando uma quebra de fita dupla. O único requisito para projetar um sgRNA para direcionar uma sequência de DNA específica é que a sequência alvo deve conter uma sequência de motivo

adjacente ao protoespaçador (PAM) na extremidade 3' da sequência de sgRNA que é complementar à sequência genômica. No caso da nuclease Cas9, a sequência PAM é NRG (onde R é A ou G e N é qualquer base) ou a sequência PAM mais restrita NGG. Portanto, as moléculas de sgRNA que direcionam qualquer região do genoma podem ser projetadas *in silico*, localizando a sequência de 20 pb adjacente a todos os motivos PAM. Os motivos de PAM ocorrem em média muitos 15 pb no genoma dos eucariotas. No entanto, o sgRNA projetado por métodos in silico irá gerar quebras de fita dupla em células com eficiências diferentes e não é possível prever a eficiência de corte de uma série de moléculas de sgRNA usando métodos in silico. Como o sgRNA pode ser rapidamente sintetizado *in vitro*, isso permite a análise rápida de todas as potenciais sequências de sgRNA em uma determinada região genômica para identificar o sgRNA que resulta no corte mais eficiente. Tipicamente, quando uma série de sgRNA dentro de uma dada região genômica é testada nas células, é observada uma gama de eficiências de clivagem entre 0 e 90%. Algoritmos in silico, bem como experiências de laboratório, também podem ser usados para determinar o potencial fora do alvo de qualquer sgRNA. Embora uma combinação perfeita com a sequência de reconhecimento de 20 pb de um sgRNA ocorra principalmente apenas uma vez na maioria dos genomas eucarióticos, haverá um número de locais adicionais no genoma com 1 ou mais não correspondências de pares de bases com o sgRNA. Estes locais podem ser clivados em frequências variáveis que geralmente não são previsíveis com base no número ou localização das não correspondências. Também pode ocorrer clivagem em locais adicionais fora do alvo que não foram identificados pela análise in silico. Assim, a análise de um número de sgRNA em um tipo de célula relevante para identificar o sgRNA que possui o perfil fora do alvo mais favorável é um componente crítico da seleção de um sgRNA ideal para

uso terapêutico. Um perfil fora do alvo favorável levará em conta não apenas o número real de locais fora do alvo e a frequência de cortes nesses locais, mas também a localização no genoma desses locais. Por exemplo, locais fora do alvo próximos ou dentro de genes funcionalmente importantes, particularmente oncogenes ou anti-oncogenes, seriam considerados menos favoráveis do que locais em regiões intergênicas sem função conhecida. Assim, a identificação de um sgRNA ideal não pode ser prevista simplesmente pela análise in silico da sequência genômica de um organismo, mas requer testes experimentais. Embora a análise in silico possa ser útil para reduzir o número de guias para teste, não é possível prever guias com corte no alvo elevado ou prever guias com corte fora do alvo baixo desejável. Dados experimentais indicam que a eficiência de corte do sgRNA que cada um tem uma combinação perfeita com o genoma em uma região de interesse (como o íntron 1 da albumina) varia de nenhum corte a > 90% de corte e não é previsível por nenhum algoritmo conhecido. A capacidade de um determinado sgRNA de promover a clivagem por uma enzima Cas pode se relacionar com a acessibilidade desse local específico no DNA genômico que pode ser determinado pela estrutura da cromatina naquela região. Enquanto a maioria do DNA genômico em uma célula diferenciada quiescente, como um hepatócito, existe em heterocromatina altamente condensada, regiões que são transcritas ativamente existem em estados de cromatina mais abertos que são conhecidos por serem mais acessíveis a moléculas grandes, tal como proteínas como a proteína Cas. Mesmo dentro de genes ativamente transcritos, algumas regiões específicas do DNA são mais acessíveis que outras devido à presença ou ausência de fatores de transcrição ligados ou outras proteínas reguladoras. A previsão de locais no genoma ou dentro de um locus genômico específico ou região de um locus genômico como um íntron, e como o íntron 1 da albumina não é

possível e, portanto, precisaria ser determinado experimentalmente em um tipo de célula relevante. Depois que alguns locais são selecionados como locais potenciais para inserção, pode ser possível adicionar algumas variações a esse local, por exemplo, movendo alguns nucleotídeos a montante ou a jusante dos locais selecionados, com ou sem testes experimentais.

[00353] Em algumas modalidades, os gRNAs que podem ser usados nos métodos no presente documento descritos são um ou mais listados da Tabela 3 ou qualquer uma dos seus derivados tendo pelo menos cerca de 85% de identidade de sequência de nucleotídeos em relação aos da Tabela 3.

MODIFICAÇÕES DE ÁCIDO NUCLEICO

[00354] Em algumas modalidades, os polinucleotídeos introduzidos nas células têm uma ou mais modificações que podem ser usadas individualmente ou em combinação, por exemplo, para potenciar a atividade, estabilidade ou especificidade, alterar a entrega, reduzir as respostas imunes inatas nas células hospedeiras ou para outras melhorias, como ainda descrito no presente documento e conhecido na técnica.

[00355] Em certas modalidades, polinucleotídeos modificados são usados no sistema CRISPR/Cas9/Cpf1, caso em que os RNAs guia (guias de molécula única ou guias de molécula dupla) e/ou um DNA ou um RNA que codifica uma endonuclease Cas ou Cpf1 introduzida em uma célula pode ser modificado, como descrito e ilustrado abaixo. Tais polinucleotídeos modificados podem ser usados no sistema CRISPR/Cas9/Cpf1 para editar qualquer um ou mais loci genômicos.

[00356] Usando o sistema CRISPR/Cas9/Cpf1 para a finalidade de ilustrações não limitativas de tais usos, podem ser usadas modificações nos RNAs guia para melhorar a formação ou estabilidade do complexo de edição de genoma CRISPR/Cas9/Cpf1 com RNAs guia, que podem

ser guias de moléculas únicas ou de moléculas duplas e uma endonuclease Cas ou Cpf1. As modificações de RNAs guia podem também, ou em alternativa, ser usadas para melhorar a iniciação, estabilidade ou cinética das interações entre o complexo de edição do genoma e a sequência alvo no genoma, que pode ser usada, por exemplo, para melhorar a atividade no alvo. As modificações de RNAs guia podem também, ou em alternativa, ser usadas para aumentar a especificidade, por exemplo, as taxas relativas de edição do genoma no local no alvo, em comparação com os efeitos em outros locais (fora do alvo).

[00357] As modificações podem também, ou em alternativa, ser usadas para aumentar a estabilidade de um RNA guia, por exemplo, aumentando sua resistência à degradação por ribonucleases (RNases) presentes em uma célula, aumentando assim sua meia-vida na célula. As modificações que melhoram a meia-vida do RNA guia podem ser particularmente úteis nas modalidades em que uma endonuclease Cas ou Cpf1 é introduzida na célula a ser editada por meio de um RNA que precisa ser traduzido para gerar endonuclease, pois aumenta a meia-vida dos RNAs guia introduzidos, ao mesmo tempo que o RNA que codifica a endonuclease pode ser usado para aumentar o tempo em que os RNAs guia e a endonuclease Cas ou Cpf1 codificada coexistem na célula.

[00358] As modificações podem também, ou em alternativa, ser usadas para diminuir a probabilidade ou o grau em que os RNAs introduzidos nas células provocam respostas imunes inatas. Tais respostas, que foram bem caracterizadas no contexto de interferência de RNA (RNAi), incluindo pequenos RNAs de interferência (siRNAs), como descrito abaixo e na técnica, tendem a estarem associadas à meia-vida reduzida do RNA e/ou a indução de citocinas ou outros fatores associados às respostas imunes.

[00359] Um ou mais tipos de modificações também podem ser feitos nos RNAs que codificam uma endonuclease que são introduzidos em uma célula, incluindo, sem limitação, modificações que melhoram a estabilidade do RNA (tal como por aumento da sua degradação pelas RNAses presentes na célula), modificações que melhoram a tradução do produto resultante (isto é, a endonuclease) e/ou modificações que diminuem a probabilidade ou o grau em que os RNAs introduzidos nas células provocam respostas imunes inatas.

[00360] Combinações de modificações, como as precedentes e outras, também podem ser usadas. No caso de CRISPR/Cas9/Cpf1, por exemplo, um ou mais tipos de modificações podem ser feitas para guiar RNAs (incluindo aqueles exemplificados acima) e/ou um ou mais tipos de modificações podem ser feitas para RNAs que codificam a endonuclease Cas (incluindo os exemplificados acima).

[00361] A título ilustrativo, os RNAs guia usados no sistema CRISPR/Cas9/Cpf1 ou outros RNAs menores podem ser facilmente sintetizados por meios químicos, permitindo que várias modificações sejam prontamente incorporadas, como ilustrado abaixo e descrito na técnica. Embora os procedimentos químicos sintéticos estejam em constante expansão, a purificação desses RNAs por procedimentos como a cromatografia líquida de alto desempenho (HPLC, que evita o uso de géis como PAGE) tende a se tornar mais desafiadora, pois os comprimentos dos polinucleotídeos aumentam significativamente além de cem nucleotídeos. Uma abordagem usada para gerar RNAs quimicamente modificados de maior comprimento é produzir duas ou mais moléculas que são ligadas entre si. RNAs muito mais longos, tais como os que codificam uma endonuclease Cas9, são mais facilmente gerados enzimaticamente. Embora menos tipos de modificações estejam geralmente disponíveis para uso em RNAs produzidos enzimaticamente, ainda existem modificações que podem ser usadas

para, por exemplo, melhorar a estabilidade, reduzir a probabilidade ou o grau de resposta imune inata e/ou melhorar outros atributos, como descrito adicionalmente em seguida e na técnica; e novos tipos de modificações estão sendo desenvolvidos regularmente.

[00362] A título ilustrativo de vários tipos de modificações, especialmente aquelas usadas frequentemente com RNAs menores sintetizados quimicamente, as modificações podem ter um ou mais nucleotídeos modificados na posição 2' do açúcar, em algumas modalidades um nucleotídeo modificado 2'-O-alquila, 2'-O-alquil-O-alquila ou 2'-fluoro. Em algumas modalidades, as modificações de RNA incluem modificações de 2'-fluoro, 2'-amino ou 2'-O-metila na ribose de pirimidinas, resíduos abásicos ou uma base invertida na extremidade 3' do RNA. Tais modificações são rotineiramente incorporadas nos oligonucleotídeos e mostrou-se que esses oligonucleotídeos têm uma Tm maior (isto é, maior afinidade de ligação ao alvo) do que os 2'-desoxioligonucleotídeos contra um determinado alvo.

[00363] Foi demonstrado que várias modificações de nucleotídeos e nucleosídeos tornam o oligonucleotídeo no qual são incorporados mais resistente à digestão com nucleases do que o oligonucleotídeo nativo; esses oligos modificados sobrevivem intatos por mais tempo do que os oligonucleotídeos não modificados. Exemplos específicos de oligonucleotídeos modificados incluem aqueles que possuem estruturas principais modificadas, por exemplo, fosforotioatos, fosfotriésteres, fosfonatos de metila, ligações entre açúcares alquila de cadeia curta ou cicloalquila ou ligações heteroatômicas ou heterocíclicas entre açúcares de cadeia curta. Alguns oligonucleotídeos são oligonucleotídeos com estruturas principais de fosforotioato e aqueles com estruturas principais de heteroátomos, particularmente CH_2 -NH-O-CH_2, CH, ~N(CH_3)~O~CH_2 (conhecido como uma estrutura principal metileno(metilimino) ou MMI), CH_2 --O--N (CH_3)-CH_2, CH_2 -N (CH_3)-N

(CH_3)-CH_2 e O-N (CH_3)- CH_2 -CH2, em que a estrutura principal fosfodiéster nativa é representada como O- P-- O- CH); estrutura principal amida [ver De Mesmaeker *et al*., Ace Chem. Res., 28:366-374 (1995)]; estruturas principais morfolino (ver Summerton e Weller, Patente dos EUA No. 5.034.506); estrutura principal de ácido nucleico peptídico (PNA) (em que a estrutura principal do fosfodiéster do oligonucleotídeo é substituída por uma estrutura principal de poliamida, os nucleotídeos sendo ligados diretamente ou indiretamente aos átomos de aza nitrogênio da estrutura principal de poliamida, ver Nielsen *et al*., Science 1991, 254, 1497). As ligações contendo fósforo incluem, mas não se limitam a, fosforotioatos, fosforotioatos quirais, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, fosfonatos de metila e outros fosfonatos de alquila tendo fosfonatos 3'alquileno e fosfonatos quirais, fosfinatos, fosforamidatos tendo fosforamidato 3'-amino e aminoalquilfosforamidatos, tiono-fosforamidatos, tiono-alquilfosfonatos, tiono-alquilfosfotriésteres e boranofosfatos com ligações 3'-5' normais, seus análogos ligados 2'-5' e aqueles com polaridade invertida em que os pares adjacentes de unidades nucleosídicas estão ligados 3'-5' a 5'-3 'ou 2'-5' a 5'-2' ; ver patentes dos EUA Nos. 3.687.808; 4.469.863; 4.476.301; 5.023.243; 5.177.196; 5.188.897; 5.264.423; 5.276.019; 5.278.302; 5.286.717; 5.321.131; 5.399.676; 5.405.939; 5.453.496; 5.455.233; 5.466.677; 5.476.925; 5.519.126; 5.536.821; 5.541.306; 5.550.111; 5.563.253; 5.571.799; 5.587.361; e 5.625.050.

[00364] Os compostos oligoméricos à base de morfolino são descritos em Braasch e David Corey, Biochemistry, 41 (14): 4503-4510 (2002); Genesis, Volume 30, Edição 3, (2001); Heasman, Dev. Biol., 243: 209-214 (2002); Nasevicius *et al*., Nat. Genet., 26:216-220 (2000); Lacerra *et al*., Proc. Natl. Acad. Sci., 97: 9591-9596 (2000); e Patente dos EUA No. 5.034.506, editado Jul. 23, 1991.

[00365] Os miméticos dos oligonucleotídeos do ácido nucleico ciclo-hexenila são descritos em Wang *et al*., J. Am. Chem. Soc., 122: 8595-8602 (2000).

[00366] As estruturas principais de oligonucleotídeos modificados que não incluem um átomo de fósforo nas mesmas têm estruturas principais formadas por ligações internucleosídicas de alquila ou cicloalquila de cadeia curta, heteroátomo misto e ligações internucleosídicas de alquila ou cicloalquila, ou uma ou mais ligações internucleosídicas heteroatômicas ou heterocíclicas de cadeia curta. Estes têm aqueles tendo ligações morfolino (formadas em parte a partir da fração de açúcar de um nucleósido); estruturas principais de siloxano; estruturas principais de sulfeto, sulfóxido e sulfona; estruturas principais de formacetil e tioformacetil; estruturas principais de metileno formacetil e tioformacetil; estruturas principais contendo alqueno; estruturas principais de sulfamato; estruturas principais de metilenoimina e metileno-hidrazina; estruturas principais de sulfonato e sulfonamida; estruturas principais de amida; e outras tendo partes componentes N, O, S e CH_2 misturadas; ver patentes dos EUA Nos. 5.034.506; 5.166.315; 5.185.444; 5.214.134; 5.216.141; 5.235.033; 5.264.562; 5.264.564; 5.405.938; 5.434.257; 5.466.677; 5.470.967; 5.489.677; 5.541.307; 5.561.225; 5.596.086; 5.602.240; 5.610.289; 5.602.240; 5.608.046; 5.610.289; 5.618.704; 5.623.070; 5.663.312; 5.633.360; 5.677.437; e 5.677.439, cada uma das quais está incorporada ao presente documento a título de referência.

[00367] Uma ou mais frações de açúcar substituídas também podem ser incluídas, por exemplo, um dos seguintes na posição 2': OH, SH, SCH_3, F, OCN, OCH_3, OCH_3 $O(CH_2)_n$ CH_3, $O(CH_2)_n$ NH_2 ou $O(CH_2)_n$ CH_3, em que n é de 1 a cerca de 10; alquila inferior C1 a C10, alcoxialcóxi, alquila inferior substituída, alcarila ou aralquila; Cl; Br; CN; CF_3; OCF_3; O-, S- ou N-alquila; O-, S- ou N-alquenila; $SOCH_3$; SO_2 CH_3;

ONO_2; NO_2; N_3; NH_2; heterocicloalquila; heterocicloalcarila; aminoalquilamino; polialquilamino; sililo substituído; um grupo de clivagem de RNA; um grupo de repórteres; um intercalador; um grupo para melhorar as propriedades farmacocinéticas de um oligonucleotídeo; ou um grupo para melhorar as propriedades farmacodinâmicas de um oligonucleotídeo e outros substituintes tendo propriedades similares. Em algumas modalidades, uma modificação inclui 2'-metoxietóxi (2'-O-$CH_2CH_2OCH_3$, também conhecido como 2'-O-(2-metoxietila)) (Martin *et al*, Helv. Chim. Acta, 1995, 78, 486). Outras modificações incluem 2'-metóxi (2'-0-CH_3), 2'-propóxi (2'-OCH_2 CH_2CH_3) e 2'-fluoro (2'-F). Modificações similares também podem ser feitas em outras posições no oligonucleotídeo, particularmente na posição 3' do açúcar no nucleotídeo terminal 3' e na posição 5' do nucleotídeo terminal 5'. Os oligonucleotídeos também podem ter miméticos de açúcar, tais como ciclobutilas no lugar do grupo pentofuranosila.

[00368] Em algumas modalidades, uma ligação de açúcar e uma internucleosídica, isto é, a estrutura principal, das unidades nucleotídicas são substituídas por novos grupos. As unidades de base são mantidas para hibridação com um composto alvo de ácido nucleico apropriado. Um desses compostos oligoméricos, um oligonucleotídeo mimético que demonstrou ter excelentes propriedades de hibridação, é referido como um ácido nucleico peptídico (PNA). Nos compostos do PNA, a estrutura principal de açúcar de um oligonucleotídeo é substituída por uma estrutura principal contendo amida, por exemplo, uma estrutura principal de aminoetilglicina. As nucleobases são retidas e estão ligadas diretamente ou indiretamente aos átomos de nitrogênio aza da porção amida da estrutura principal. As patentes representativas dos Estados Unidos que ensinam a preparação de compostos de PNA têm, mas não estão limitadas a, patentes dos EUA Nos. 5.539.082; 5.714.331; e 5.719.262. Ensinamentos adicionais de compostos de PNA

podem ser encontrados em Nielsen *et al*, Science, 254: 1497-1500 (1991).

[00369] Em algumas modalidades, os RNAs guia também podem incluir, adicionalmente ou em alternativa, modificações ou substituições da nucleobase (frequentemente referida na técnica simplesmente como "base"). Como no presente documento usado, nucleobases "não modificadas" ou "naturais" incluem adenina (A), guanina (G), timina (T), citosina (C) e uracila (U). As nucleobases modificadas incluem nucleobases encontradas apenas com pouca frequência ou transitoriamente em ácidos nucleicos de ocorrência natural, por exemplo, hipoxantina, 6-metiladenina, 5-Me pirimidinas, particularmente 5-metilcitosina (também referida como 5-metil-2' desoxicitosina e frequentemente referida na técnica como 5-Me-C), 5-hidroximetilcitosina (HMC), glicosil HMC e gentobiosil HMC, bem como nucleobases sintéticas, por exemplo, 2-aminoadenina, 2-(metilamino)adenina, 2-(imidazolilalquil)adenina, 2-(aminoalquilamino)adenina ou outras alquiladeninas heterossubstituídas, 2-tiouracila, 2-tiotimina, 5-bromouracila, 5-hidroximetiluracila, 8-azaguanina, 7-desazaguanina, N6(6-amino-hexil)adenina e 2,6-diaminopurina. Kornberg, A., DNA Replication, W. H. Freeman & Co., São Francisco, pp 75-77 (1980); Gebeyehu *et al*., Nucl. Acids Res. 15:4513 (1997). Uma base "universal" conhecida na técnica, por exemplo, inosina, também pode ser incluída. Foi demonstrado que as substituições 5-Me-C aumentam a estabilidade do dúplex de ácido nucleico em 0,6-1,2 ºC. (Sanghvi, Y. S., em Crooke, S. T. e Lebleu, B., eds., Antisense Research and Applications, CRC Press, Boca Raton, 1993, pp. 276-278) e são modalidades de substituições de bases.

[00370] Em algumas modalidades, as nucleobases modificadas incluem outras nucleobases sintéticas e naturais, tais como 5-

metilcitosina (5-me-C), 5-hidroximetilcitosina, xantina, hipoxantina, 2-aminoadenina, 6-metila e outros derivados alquilados de adenina e guanina, 2-propila e outros derivados alquilados da adenina e guanina, 2-tiouracila, 2-tiotimina e 2-tiocitosina, 5-halouracila e citosina, 5-propinila uracila e citosina, 6-azo uracila, citosina e timina, 5-uracila (pseudo-uracila), 4-tiouracila, 8-halo, 8-amino, 8-tiol, 8-tioalquila, 8-hidroxila e outras adeninas e guaninas substituídas em 8, 5-halo, particularmente 5-bromo, 5-trifluorometila e outras 5-uracilas e citosinas substituídas em 7-metilquanina e 7-metiladenina, 8-azaguanina e 8-aza-adenina, 7-deazaguanina e 7-deaza-adenina e 3-deazaguanina e 3-deaza-adenina.

[00371] Além disso, as nucleobases incluem as descritas na Patente dos EUA No. 3.687.808, as descritas em 'The Concise Encyclopedia of Polymer Science And Engineering', páginas 858-859, Kroschwitz, J.I., ed. John Wiley & Sons, 1990, as descritas por Englisch *et al*., Angewandle Chemie, International Edition', 1991, 30, página 613, e as descritas por Sanghvi, Y.S., Capítulo 15, 'Antisense Research and Applications', páginas 289-302, Crooke, S.T. e Lebleu, B. ea., CRC Press, 1993. Algumas destas nucleobases são particularmente úteis para aumentar a afinidade de ligação dos compostos oligoméricos da invenção. Estes incluem pirimidinas 5-substituídas, 6-azapirimidinas e purinas substituídas N-2, N-6 e 0-6, tendo 2-aminopropiladenina, 5-propiniluracila e 5-propinilcitosina. Foi demonstrado que as substituições de 5-metilcitosina aumentam a estabilidade do dúplex de ácido nucleico em 0,6-1,2 ºC (Sanghvi, Y.S., Crooke, S.T. e Lebleu, B., eds, 'Antisense Research and Applications', CRC Press, Boca Raton, 1993, pp. 276-278) e são modalidades de substituições de bases, ainda mais particularmente quando combinadas com modificações de açúcar 2'-O-metoxietila. As nucleobases modificadas são descritas nas patentes dos EUA Nos. 3.687.808, bem como 4.845.205; 5.130.302;

5.134.066; 5.175.273; 5.367.066; 5.432.272; 5.457.187; 5.459.255; 5.484.908; 5.502.177; 5.525.711; 5.552.540; 5.587.469; 5.596.091; 5.614.617; 5.681.941; 5.750.692; 5.763.588; 5.830.653; 6.005.096; e a Publicação do Pedido de Patente dos EUA 2003/0158403.

[00372] Em algumas modalidades, os RNAs guia e/ou mRNA (ou DNA) que codifica uma endonuclease são quimicamente ligados a uma ou mais frações ou conjugados que aumentam a atividade, distribuição celular ou absorção celular do oligonucleotídeo. Tais frações incluem, mas não estão limitadas a, frações lipídicas, tal como uma fração de colesterol [Letsinger *et al.*, Proc. Natl. Acad. Sci. USA, 86: 6553-6556 (1989)]; ácido cólico [Manoharan *et al*., Bioorg. Med. Chem. Let., 4: 1053-1060 (1994)]; um tioéter, por exemplo, hexil-S-tritiltiol [Manoharan *et al*, Ann. N. Y. Acad. Sci., 660: 306-309 (1992) e Manoharan *et al*., Bioorg. Med. Chem. Let., 3: 2765-2770 (1993)]; um tiocolesterol [Oberhauser *et al*., Nucl. Acids Res., 20: 533-538 (1992)]; uma cadeia alifática, por exemplo, resíduos de dodecandiol ou undecila [Kabanov *et al*., FEBS Lett., 259: 327-330 (1990) e Svinarchuk *et al*., Biochimie, 75: 49-54 (1993)]; um fosfolipídio, por exemplo, di-hexadecil-rac-glicerol ou 1,2-di-O-hexadecil- rac-glicero-3-H-fosfonato de trietilamônio [Manoharan et al., Tetrahedron Lett., 36: 3651-3654 (1995) e Shea *et al*., Nucl. Acids Res., 18: 3777-3783 (1990)]; uma cadeia de poliamina ou polietilenoglicol [Mancharan *et al*., Nucleosides & Nucleotides, 14: 969-973 (1995)]; ácido acético de adamantano [Manoharan *et al*., Tetrahedron Lett., 36: 3651-3654 (1995)]; uma fração palmitila [(Mishra *et a*l., Biochim. Biophys. Acta, 1264: 229-237 (1995)]; ou uma fração octadecilamina ou hexilamino-carbonil-t oxicolesterol [Crooke *et al*., J. Pharmacol. Exp. Ther., 277: 923-937 (1996)]. Consultar também Patente dos EUA Nos. 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717, 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603;

5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250; 5.292.873; 5.317.098; 5.371.241, 5.391.723; 5.416.203, 5.451.463; 5.510.475; 5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371; 5.595.726; 5.597.696; 5.599.923; 5.599.928 e 5.688.941.

[00373] Em algumas modalidades, açúcares e outras frações podem ser usados para direcionar proteínas e complexos com nucleotídeos, tal como polissomos e lipossomos catiônicos, para locais específicos. Por exemplo, a transferência dirigida a células hepáticas pode ser mediada através de receptores de asialoglicoproteínas (ASGPRs); ver, por exemplo, Hu, *et al*., Protein Pept Lett. 21(10):1025-30 (2014). Outros sistemas conhecidos na técnica e regularmente desenvolvidos podem ser usados para direcionar biomoléculas de uso no presente caso e/ou seus complexos para células alvo específicas de interesse.

[00374] Em algumas modalidades, estas frações ou conjugados de direcionamento podem incluir grupos conjugados ligados covalentemente a grupos funcionais, tais como grupos hidroxila primários ou secundários. Grupos conjugados da invenção incluem intercaladores, moléculas repórteres, poliaminas, poliamidas, polietilenoglicóis, poliéteres, grupos que melhoram as propriedades farmacodinâmicas dos oligômeros e grupos que melhoram as propriedades farmacocinéticas dos oligômeros. Grupos conjugados típicos incluem colesteróis, lipídios, fosfolipídios, biotina, fenazina, folato, fenantridina, antraquinona, acridina, fluoresceínas, rodaminas, cumarinas e corantes. Grupos que melhoram as propriedades farmacodinâmicas, no contexto desta invenção, incluem grupos que melhoram a absorção, melhoram a resistência à degradação e/ou

fortalecem a hibridação específica de sequência com o ácido nucleico alvo. Grupos que melhoram as propriedades farmacocinéticas, no contexto desta invenção, incluem grupos que melhoram a absorção, distribuição, metabolismo ou excreção dos compostos da presente invenção. Grupos conjugados representativos são descritos no Pedido de Patente Internacional No. PCT/US92/09196, depositado em 23 de outubro de 1992, e Patente dos EUA No. 6.287.860, que são no presente documento incorporadas por referência. As frações conjugadas incluem, mas não estão limitadas a, frações lipídicas, como uma fração de colesterol, ácido cólico, um tioéter, por exemplo, hexil-5-tritiltiol, um tiocolesterol, uma cadeia alifática, por exemplo, resíduos de dodecandiol ou undecila, um fosfolipídio, por exemplo, di-hexadecil-rac-glicerol ou trietilamônio l,2-di-O-hexadecil-rac-glicero-3-H-fosfonato, uma cadeia de poliamina ou polietilenoglicol, ou ácido acético de adamantano, uma fração palmitilo ou uma fração octadecilamina ou hexilamino-carbonil-óxi-colesterol. Consultar, por exemplo, Patente dos EUA. Nos. 4.828.979; 4.948.882; 5.218.105; 5.525.465; 5.541.313; 5.545.730; 5.552.538; 5.578.717, 5.580.731; 5.580.731; 5.591.584; 5.109.124; 5.118.802; 5.138.045; 5.414.077; 5.486.603; 5.512.439; 5.578.718; 5.608.046; 4.587.044; 4.605.735; 4.667.025; 4.762.779; 4.789.737; 4.824.941; 4.835.263; 4.876.335; 4.904.582; 4.958.013; 5.082.830; 5.112.963; 5.214.136; 5.082.830; 5.112.963; 5.214.136; 5.245.022; 5.254.469; 5.258.506; 5.262.536; 5.272.250; 5.292.873; 5.317.098; 5.371.241, 5.391.723; 5.416.203, 5.451.463; 5.510.475; 5.512.667; 5.514.785; 5.565.552; 5.567.810; 5.574.142; 5.585.481; 5.587.371; 5.595.726; 5.597.696; 5.599.923; 5.599.928 e 5.688.941.

[00375] Polinucleotídeos mais longos que são menos passíveis de síntese química e são tipicamente produzidos por síntese enzimática também podem ser modificados por vários meios. Tais modificações podem incluir, por exemplo, a introdução de determinados análogos de

nucleotídeos, a incorporação de sequências particulares ou outras frações nas extremidades 5' ou 3' das moléculas, e outras modificações. A título de ilustração, o mRNA que codifica Cas9 tem aproximadamente 4 kb de comprimento e pode ser sintetizado por transcrição *in vitro*. Modificações no mRNA podem ser aplicadas para, por exemplo, aumentar a sua tradução ou estabilidade (tal como por aumento da sua resistência à degradação com uma célula) ou reduzir a tendência do RNA de provocar uma resposta imune inata que é frequentemente observada nas células após a introdução de RNAs exógenos, particularmente RNAs mais longos, como o que codifica a Cas9.

[00376] Várias modificações foram descritas na técnica, tais como caudas de poliA, análogos de extremidades 5' (por exemplo, Análogo de Extremidade Antirreverso (ARCA) ou m7G(5')ppp(5')G (mCAP)), regiões não traduzidas modificadas em 5' ou em 3' (UTRs), uso de bases modificadas (tais como Pseudo-UTP, 2-Tio-UTP, 5-Metilcitidina-5'-Trifosfato (5-Metil-CTP) ou N6-metil-ATP) ou tratamento com fosfatase para remover os fosfatos terminais 5'. Estas e outras modificações são conhecidas na técnica e novas modificações de RNAs estão sendo desenvolvidas regularmente.

[00377] Existem inúmeros fornecedores comerciais de RNAs modificados, incluindo, por exemplo, TriLink Biotech, AxoLabs, Bio-Synthesis Inc., Dharmacon e muitos outros. Como descrito por TriLink, por exemplo, o 5-Metil-CTP pode ser usado para conferir características desejáveis, como estabilidade da nuclease aumentada, tradução aumentada ou interação reduzida de receptores imunes inatos com RNA transcrito *in vitro*. Também foi mostrado que a 5-Metilcitidina-5'-Trifosfato (5-metil-CTP), N6-metil-ATP, bem como pseudo-UTP e 2-tio-UTP, reduzem a estimulação imune inata na cultura e *in vivo* enquanto potenciam a tradução, como ilustrado nas publicações por Kormann *et al*. and Warren *et al*. referidas abaixo.

[00378] Foi mostrado que o mRNA quimicamente modificado entregue *in vivo* pode ser usado para alcançar efeitos terapêuticos melhorados; ver, por exemplo, Kormann *et al*., Nature Biotechnology 29, 154–157 (2011). Tais modificações podem ser usadas, por exemplo, para aumentar a estabilidade da molécula de RNA e/ou reduzir sua imunogenicidade. Usando modificações químicas como Pseudo-U, N6-Metil-A, 2-Tio-U e 5-Metil-C, foi mostrado que substituindo apenas um quarto dos resíduos de uridina e citidina por 2-Tio-U e 5-Metil-C, respectivamente, resultou em uma diminuição significativa no reconhecimento do mRNA mediado pelo receptor tipo toll (TLR) em camundongos. Ao reduzir a ativação do sistema imune inato, estas modificações podem ser usadas para aumentar efetivamente a estabilidade e a longevidade do mRNA *in vivo*; ver, por exemplo, Kormann *et al*., *supra*.

[00379] Também foi mostrado que a administração repetida de RNAs mensageiros sintéticos incorporando modificações projetadas para contornar respostas antivirais inatas pode reprogramar células humanas diferenciadas para pluripotência. Consultar, por exemplo, Warren, *et al*., Cell Stem Cell, 7(5):618-30 (2010). Tais mRNAs modificados que atuam como proteínas de reprogramação primárias podem ser um meio eficiente de reprogramar vários tipos de células humanas. Tais células são denominadas células-tronco de pluripotência induzida (iPSCs), e verificou-se que o RNA sintetizado enzimaticamente incorporando 5-Metil-CTP, Pseudo-UTP e um Análogo de Extremidade Antirreverso (ARCA) poderia ser usado para evitar efetivamente a resposta antiviral da célula; ver, por exemplo, Warren *et al*., *supra*.

[00380] Outras modificações de polinucleotídeos descritas na técnica incluem, por exemplo, o uso de caudas de poliA, a adição de análogos da extremidade 5' (tal como m7G(5')ppp(5')G (mCAP)), modificações nas regiões não traduzidas 5' ou 3' (UTRs) ou tratamento com fosfatase

para remover os fosfatos terminais 5' - e novas abordagens estão sendo desenvolvidas regularmente.

[00381] Várias composições e técnicas aplicáveis à geração de RNAs modificados para uso neste documento foram desenvolvidas em conexão com a modificação da interferência de RNA (RNAi), incluindo pequenos RNAs de interferência (siRNAs). Os siRNAs apresentam desafios particulares *in vivo* porque seus efeitos no silenciamento de genes por interferência de mRNA são geralmente transientes, o que pode exigir administração repetida. Além disso, os siRNAs são RNAs de fita dupla (dsRNA) e as células de mamíferos têm respostas imunes que evoluíram para detectar e neutralizar o dsRNA, que geralmente é um subproduto da infecção viral. Assim, existem enzimas de mamíferos como a PKR (quinase responsiva ao dsRNA) e o gene I potencialmente induzível por ácido retinoico (RIG-I), que podem mediar as respostas celulares ao dsRNA, bem como receptores do tipo Toll (tais como TLR3, TLR7 e TLR8) que podem desencadear a indução de citocinas em resposta a tais moléculas; ver, por exemplo, as revisões de Angart *et al.*, Pharmaceuticals (Basel) 6(4): 440–468 (2013); Kanasty *et al.*, Molecular Therapy 20(3): 513–524 (2012); Burnett *et al.*, Biotechnol J. 6(9):1130-46 (2011); Judge e MacLachlan, Hum Gene Ther 19(2):111-24 (2008); e referências neles citadas.

[00382] Uma grande variedade de modificações foi desenvolvida e aplicada para melhorar a estabilidade do RNA, reduzir as respostas imunes inatas e/ou alcançar outros benefícios que podem ser úteis em conexão com a introdução de polinucleotídeos nas células humanas, como descrito no presente documento; ver, por exemplo, os artigos de revisão por Whitehead KA *et al.*, Annual Review of Chemical and Biomolecular Engineering, 2: 77-96 (2011); Gaglione e Messere, Mini Rev Med Chem, 10(7):578-95 (2010); Chernolovskaya *et al*, Curr Opin Mol Ther., 12(2):158-67 (2010); Deleavey *et al.*, Curr Protoc Nucleic Acid

Chem Capítulo 16:Unidade 16.3 (2009); Behlke, Oligonucleotides 18(4):305-19 (2008); Fucini *et al*., Nucleic Acid Ther 22(3): 205–210 (2012); Bremsen *et al*., Front Genet 3:154 (2012).

[00383] Como observado acima, existem vários fornecedores comerciais de RNAs modificados, muitos dos quais especializados em modificações projetadas para melhorar a eficácia dos siRNAs. Uma variedade de abordagens é oferecida com base em várias descobertas relatadas na literatura. Por exemplo, Dharmacon observa que a substituição de oxigênio sem ponte por enxofre (fosforotioato, PS) tem sido amplamente usada para melhorar a resistência à nuclease de siRNAs, como relatado por Kole, Nature Reviews Drug Discovery 11:125-140 (2012). Foram relatadas modificações da posição 2' da ribose para melhorar a resistência à nuclease da ligação fosfato internucleotídeo, enquanto aumentam a estabilidade do duplex (Tm), que também mostrou fornecer proteção contra a ativação imune. Uma combinação de modificações moderadas da estrutura principal PS com pequenas substituições 2' bem toleradas (2'-O-Metila, 2'-Fluoro, 2'-Hidro) foram associadas a siRNAs altamente estáveis para aplicações *in vivo*, como relatado por Soutschek *et al*. Nature 432:173-178 (2004); foi relatado que as modificações 2'-O-Metila são eficazes na melhoria da estabilidade, como relatado por Volkov, Oligonucleotides 19:191-202 (2009). Com relação à diminuição da indução de respostas imunes inatas, foi relatado que a modificação de sequências específicas com 2'-O-Metila, 2'-Fluoro, 2'-Hidro reduz a interação TLR7/TLR8, preservando geralmente a atividade silenciadora; ver, por exemplo, Judge *et al*., Mol. Ther. 13:494-505 (2006); e Cekaite *et al*., J. Mol. Biol. 365:90-108 (2007). Também foram mostradas modificações adicionais, tais como 2-tiouracila, pseudouracila, 5-metilcitosina, 5-metiluracila e N6-metiladenosina, para minimizar os efeitos imunes mediados por TLR3, TLR7 e TLR8; ver, por exemplo, Kariko, K. *et al*., Immunity 23:165-175

(2005).

[00384] Como também é conhecido na técnica, e disponível comercialmente, vários conjugados podem ser aplicados a polinucleotídeos, tais como RNAs, para uso no presente documento que podem melhorar sua entrega e/ou absorção pelas células, incluindo, por exemplo, colesterol, tocoferol e fólico ácido, lipídios, peptídeos, polímeros, ligantes e aptâmeros; ver, por exemplo, a revisão por Winkler, Ther. Deliv. 4:791-809 (2013) e referências neles citadas.

ENTREGA

[00385] Em algumas modalidades, qualquer uma das moléculas de ácido nucleico usadas nos métodos no presente documento fornecidos, por exemplo, um ácido nucleico que codifica um ácido nucleico direcionado ao genoma da invenção e/ou um polipeptídeo direcionado ao local é empacotado na superfície dos veículos de entrega ou na superfície dos veículos de entrega para entrega às células. Os veículos de entrega contemplados incluem, mas não estão limitados a, nanoesferas, lipossomos, pontos quânticos, nanopartículas, partículas de polietilenoglicol, hidrogéis e micelas. Como descrito na técnica, uma variedade de frações de direcionamento pode ser usada para potenciar a interação preferencial de tais veículos com os tipos ou locais de células desejados.

[00386] A introdução dos complexos, polipeptídeos e ácidos nucleicos da invenção nas células pode ocorrer por infecção viral ou bacteriófago, transfecção, conjugação, fusão de protoplastos, lipofecção, eletroporação, nucleofecção, precipitação de fosfato de cálcio, transfecção mediada por polietilenoimina (PEI), transfecção mediada por DEAE-dextrano, transfecção mediada por lipossomos, tecnologia de pistola de partículas, precipitação com fosfato de cálcio, microinjeção direta, entrega de ácido nucleico mediada por nanopartículas, e similares.

[00387] Em modalidades, os polinucleotídeos de RNA guia (RNA ou DNA) e/ou polinucleotídeo(s) de endonuclease (RNA ou DNA) podem ser entregues por veículos de entrega virais ou não virais conhecidos na técnica. Em alternativa, o(s) polipeptídeo(s) de endonuclease podem ser entregues por veículos de entrega virais ou não virais conhecidos na técnica, como eletroporação ou nanopartículas lipídicas. Em algumas modalidades, a endonuclease de DNA pode ser entregue como um ou mais polipeptídeos, isolados ou pré-complexados com um ou mais RNAs guia, ou um ou mais crRNA juntamente com um tracrRNA.

[00388] Em modalidades, os polinucleotídeos podem ser entregues por veículos de entrega não virais, incluindo, mas não limitados a, nanopartículas, lipossomos, ribonucleoproteínas, peptídeos carregados positivamente, conjugados de RNA de molécula pequena, quimeras de aptâmero-RNA e complexos de RNA-proteínas de fusão. Alguns veículos de entrega não virais exemplificativos são descritos em Peer e Lieberman, Gene Therapy, 18: 1127-1133 (2011) (que se foca em veículos de entrega não virais para siRNA que também são úteis para a entrega de outros polinucleotídeos).

[00389] Em modalidades, polinucleotídeos, tais como RNA guia, sgRNA e mRNA que codificam uma endonuclease, podem ser entregues a uma célula ou a um paciente por uma nanopartícula lipídica (LNP).

[00390] Embora vários métodos de entrega não viral de ácidos nucleicos tenham sido testados tanto em modelos animais quanto em humanos, o sistema mais bem desenvolvido são as nanopartículas lipídicas. As nanopartículas lipídicas (LNP) são geralmente compostas por um lipídio catiônico ionizável e 3 ou mais componentes adicionais, tipicamente colesterol, DOPE e um lipídio contendo PoliEtilenoGlicol (PEG), ver, por exemplo, Exemplo 2. O lipídio catiônico pode se ligar ao ácido nucleico carregado positivamente, formando um complexo denso

que protege o núcleo da degradação. Durante a passagem através de um sistema microfluídico, os componentes se automontam para formar partículas na gama de tamanho de 50 a 150 nM, nas quais o ácido nucleico é encapsulado no núcleo complexado com o lipídio catiônico e cercado por uma estrutura similar à bicamada lipídica. Após a injeção na circulação de um indivíduo estas partículas podem se ligar à apolipoproteína E (apoE). A ApoE é um ligante para o receptor de LDL e medeia a absorção para os hepatócitos do fígado através de endocitose mediada por receptor. Foi demonstrado que a LNP deste tipo fornece eficientemente mRNA e siRNA aos hepatócitos do fígado de roedores, primatas e humanos. Após a endocitose, a LNP está presente nos endossomos. O ácido nucleico encapsulado sofre um processo de escape endossomal mediado pela natureza ionizável do lipídio catiônico. Isto entrega o ácido nucleico no citoplasma, onde o mRNA pode ser traduzido para a proteína codificada. Assim, em algumas modalidades, o encapsulamento de gRNA e mRNA que codifica Cas9 em uma LNP é usado para entregar eficientemente ambos os componentes aos hepatócitos após injeção IV. Após a fuga endossomal, o mRNA de Cas9 é traduzido para a proteína Cas9 e pode formar um complexo com o gRNA. Em algumas modalidades, a inclusão de um sinal de localização nuclear na sequência da proteína Cas9 promove a translocação do complexo da proteína Cas9/gRNA para o núcleo. Em alternativa, o pequeno gRNA atravessa o complexo de poros nucleares e forma complexos com a proteína Cas9 no núcleo. Uma vez no núcleo, o complexo gRNA/Cas9 analisa o genoma em busca de locais alvo homólogos e gera quebras de fita dupla preferencialmente no local alvo desejado no genoma. A meia-vida das moléculas de RNA *in vivo* é curta, da ordem de horas a dias. Da mesma forma, a meia-vida das proteínas tende a ser curta, da ordem de horas a dias. Assim, em algumas modalidades, a entrega do gRNA e mRNA de Cas9 usando

uma LNP pode resultar em apenas expressão e atividade transiente do complexo gRNA/Cas9. Isto pode fornecer a vantagem de reduzir a frequência da clivagem fora do alvo e, assim, minimizar o risco de genotoxicidade em algumas modalidades. A LNP é geralmente menos imunogênica que as partículas virais. Enquanto muitos humanos têm imunidade preexistente para o AAV, não há imunidade preexistente para a LNP. Uma resposta imune adicional e adaptativa contra a LNP é improvável que ocorra, o que permite doses repetidas da LNP.

[00391] Vários lipídios catiônicos ionizáveis diferentes foram desenvolvidos para uso nas LNP. Estes incluem C12-200 (Love et al (2010), PNAS vol. 107, 1864–1869), MC3, LN16, MD1, entre outros. Em um tipo de LNP, uma fração GalNac é anexada à parte externa da LNP e atua como um ligante para absorção no fígado por meio do receptor da asialoglicoproteína. Qualquer um desses lipídios catiônicos é usado para formular LNP para a entrega de gRNA e mRNA de Cas9 ao fígado.

[00392] Em algumas modalidades, uma LNP refere-se a qualquer partícula com um diâmetro inferior a 1000 nm, 500 nm, 250 nm, 200 nm, 150 nm, 100 nm, 75 nm, 50 nm ou 25 nm. Em alternativa, uma nanopartícula pode variar em tamanho de 1-1000 nm, 1-500 nm, 1-250 nm, 25-200 nm, 25-100 nm, 35-75 nm ou 25-60 nm.

[00393] As LNPs podem ser feitas a partir de lipídios catiônicos, aniônicos ou neutros. Os lipídios neutros, como o DOPE fosfolipídio fusogênico ou o componente da membrana colesterol, podem ser incluídos nas LNPs como 'lipídios auxiliares' para potenciar a atividade de transfecção e a estabilidade das nanopartículas. As limitações dos lipídios catiônicos incluem baixa eficácia devido à baixa estabilidade e rápida eliminação, bem como à geração de respostas inflamatórias ou anti-inflamatórias. As LNPs também podem ter lipídios hidrofóbicos, lipídios hidrofílicos ou lipídios hidrofóbicos e hidrofílicos.

[00394] Qualquer lipídio ou combinação de lipídios que são

conhecidos na técnica pode ser usado para produzir uma LNP. Exemplos de lipídios usados para produzir LNPs são: DOTMA, DOSPA, DOTAP, DMRIE, DC-colesterol, DOTAP–colesterol, GAP-DMORIE–DPyPE e GL67A–DOPE–DMPE–polietilenoglicol (PEG). Exemplos de lipídios catiônicos são: 98N12-5, C12-200, DLin-KC2-DMA (KC2), DLin-MC3-DMA (MC3), XTC, MD1 e 7C1. Exemplos de lipídios neutros são: DPSC, DPPC, POPC, DOPE e SM. Exemplos de lipídios modificados com PEG são: PEG-DMG, PEG-CerC14 e PEG-CerC20.

[00395] Nas modalidades, os lipídios podem ser combinados em qualquer número de razões molares para produzir uma LNP. Além disso, o(s) polinucleotídeo(s) pode(m) ser combinado(s) com lipídio(s) em uma ampla gama de razões molares para produzir uma LNP.

[00396] Em modalidades, o polipeptídeo direcionado ao local e o ácido nucleico direcionado ao genoma podem ser administrados separadamente a uma célula ou a um paciente. Por outro lado, o polipeptídeo direcionado ao local pode ser pré-complexado com um ou mais RNAs guia, ou um ou mais crRNA em conjunto com um tracrRNA. O material pré-complexado pode então ser administrado a uma célula ou a um paciente. Tal material pré-complexado é conhecido como partícula de ribonucleoproteína (RNP).

[00397] O RNA é capaz de formar interações específicas com RNA ou DNA. Embora essa propriedade seja explorada em muitos processos biológicos, ela também traz o risco de interações promíscuas em um ambiente celular rico em ácidos nucleicos. Uma solução para esse problema é a formação de partículas de ribonucleoproteínas (RNPs), nas quais o RNA é pré-complexado com uma endonuclease. Outro benefício da RNP é a proteção do RNA contra a degradação.

[00398] Em algumas modalidades, a endonuclease na RNP pode ser modificada ou não modificada. Da mesma forma, o gRNA, crRNA, tracrRNA ou sgRNA podem ser modificados ou não modificados.

Numerosas modificações são conhecidas na técnica e podem ser usadas.

[00399] A endonuclease e o sgRNA podem geralmente ser combinados na razão molar de 1:1. Em alternativa, a endonuclease, crRNA e tracrRNA podem geralmente ser combinados em uma razão molar de 1:1:1. No entanto, uma ampla gama de razões molares pode ser usada para produzir uma RNP.

[00400] Em algumas modalidades, um vetor de vírus adeno-associado (AAV) recombinante pode ser usado para entrega. Técnicas para produzir partículas de rAAV, nas quais um genoma do AAV a ser empacotado que inclui o polinucleotídeo a ser entregue, genes rep e cap e funções de vírus auxiliar são fornecidas a uma célula, são padrão na técnica. A produção de rAAV requer que os seguintes componentes estejam presentes em uma única célula (aqui denominada célula de empacotamento): um genoma de rAAV, genes rep e cap de AAV separados (isto é, não dentro) do genoma do rAAV e funções de vírus auxiliar. Os genes rep e cap do AAV podem ser de qualquer sorotipo do AAV para o qual o vírus recombinante possa ser derivado e podem ser de um sorotipo AAV diferente dos ITRs do genoma do rAAV, incluindo, sem limitação, os sorotipos AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAV-13 e AAV rh.74. A produção de rAAV pseudotipado é descrita em, por exemplo, número de publicação do pedido de patente internacional WO 01/83692. Consultar a Tabela 1.

Tabela 1. Sorotipo de AAV e No. de acesso ao Genbank de alguns AAVs selecionados.

Sorotipo de AAV	**No. de Acesso ao Genbank**
AAV-1	NC_002077,1
AAV-2	NC_001401,2
AAV-3	NC_001729,1
AAV-3B	AF028705,1

AAV-4	NC_001829,1
AAV-5	NC_006152,1
AAV-6	AF028704,1
AAV-7	NC_006260,1
AAV-8	NC_006261,1
AAV-9	AX753250,1
AAV-10	AY631965,1
AAV-11	AY631966,1
AAV-12	DQ813647,1
AAV-13	EU285562,1

[00401] Em algumas modalidades, um método para gerar uma célula de empacotamento envolve a criação de uma linha celular que expressa de maneira estável todos os componentes necessários para a produção de partículas de AAV. Por exemplo, um plasmídeo (ou vários plasmídeos) com um genoma de rAAV sem os genes rep e cap de AAV, os genes rep e cap de AAV separados do genoma de rAAV e um marcador selecionável, tal como um gene de resistência à neomicina, são integrados ao genoma de uma célula. Os genomas de AAV foram introduzidos em plasmídeos bacterianos através de procedimentos como cauda de GC (Samulski *et al*., 1982, Proc. Natl. Acad. S6. USA, 79:2077-2081), adição de ligantes sintéticos contendo locais de clivagem por endonucleases de restrição (Laughlin *et al*., 1983, Gene, 23:65-73) ou por ligação direta com extremidades abruptas (Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666). A linha celular do empacotamento é então infectada com um vírus auxiliar, como o adenovírus. As vantagens deste método são que as células são selecionáveis e são adequadas para a produção em larga escala de rAAV. Outros exemplos de métodos adequados empregam adenovírus ou baculovírus, em vez de plasmídeos, para introduzir genomas de rAAV e/ou genes rep e cap em células de empacotamento.

[00402] Os princípios gerais da produção de rAAV são revistos em, por exemplo, Carter, 1992, Current Opinions in Biotechnology, 1533-

539; e Muzyczka, 1992, Curr. Topics in Microbial. and Immunol., 158:97-129. Várias abordagens são descritas em Ratschin *et al*., Mol. Cell. Biol. 4:2072 (1984); Hermonat *et al*., Proc. Natl. Acad. Sci. USA, 81:6466 (1984); Tratschin *et al*., Mo1. Cell. Biol. 5:3251 (1985); McLaughlin *et al*., J. Virol., 62:1963 (1988); e Lebkowski *et al*., 1988 Mol. Cell. Biol., 7:349 (1988). Samulski *et al*. (1989, J. Virol., 63:3822-3828); Patente dos EUA No. 5.173.414; WO 95/13365 e Patente dos EUA correspondente No. 5.658.776; WO 95/13392; WO 96/17947; PCT/US98/18600; WO 97/09441 (PCT/US96/14423); WO 97/08298 (PCT/US96/13872); WO 97/21825 (PCT/US96/20777); WO 97/06243 (PCT/FR96/01064); WO 99/11764; Perrin *et al*. (1995) Vaccine 13:1244-1250; Paul *et al*. (1993) Human Gene Therapy 4:609-615; Clark *et al*. (1996) Gene Therapy 3:1124-1132; Patente dos EUA No. 5.786.211; Patente dos EUA No. 5.871.982; e Patente dos EUA No. 6.258.595.

[00403] Os sorotipos do vetor AAV podem ser correspondidos aos tipos de células alvo. Por exemplo, os seguintes tipos de células exemplificativas podem ser transduzidos pelos sorotipos AAV indicados, entre outros. Por exemplo, os sorotipos de vetores de AAV adequados ao tecido hepático/tipo de célula incluem, mas não se limitam a, AAV3, AAV5, AAV8 e AAV9.

[00404] Além dos vetores virais adeno-associados, outros vetores virais podem ser usados. Tais vetores virais incluem, mas não estão limitados a, lentivírus, alfavírus, enterovírus, pestivírus, baculovírus, herpesvírus, vírus Epstein Barr, papovavírus, poxvírus, vírus vaccinia e vírus herpes simplex.

[00405] Em algumas modalidades, o mRNA de Cas9, o sgRNA direcionado a um ou dois loci nos genes da albumina e o DNA doador são formulados separadamente em nanopartículas lipídicas, ou são coformulados em uma nanopartícula lipídica, ou coformulados em duas ou mais nanopartículas lipídicas.

[00406] Em algumas modalidades, o mRNA de Cas9 é formulado em uma nanopartícula lipídica, enquanto o sgRNA e o DNA doador são entregues em um vetor AAV. Em algumas modalidades, o mRNA de Cas9 e o sgRNA são co-formulados em uma nanopartícula lipídica, enquanto o DNA doador é entregue em um vetor AAV.

[00407] Estão disponíveis opções para entrega da nuclease Cas9 como um plasmídeo de DNA, como mRNA ou como uma proteína. O RNA guia pode ser expresso a partir do mesmo DNA ou também pode ser entregue como um RNA. O RNA pode ser quimicamente modificado para alterar ou melhorar sua meia-vida ou diminuir a probabilidade ou o grau de resposta imune. A proteína endonuclease pode ser complexada com o gRNA antes da entrega. Os vetores virais permitem entrega eficiente; versões divididas da Cas9 e ortólogos menores da Cas9 podem ser empacotadas no AAV, assim como doadores para HDR. Também existe uma variedade de métodos de entrega não virais que podem entregar cada um desses componentes ou métodos não virais e virais podem ser empregues em conjunto. Por exemplo, nanopartículas podem ser usadas para entregar a proteína e o RNA guia, enquanto o AAV pode ser usado para entregar um DNA doador.

[00408] Em algumas modalidades que estão relacionadas à entrega de componentes de edição do genoma para tratamentos terapêuticos, pelo menos dois componentes são entregues no núcleo de uma célula a ser transformada, por exemplo, hepatócitos; uma nuclease específica de sequência e um molde de DNA doador. Em algumas modalidades, o molde de DNA doador é empacotado em um Vírus Adeno-Associado (AAV) com tropismo para o fígado. Em algumas modalidades, o AAV é selecionado entre os sorotipos AAV8, AAV9, AAVrh10, AAV5, AAV6 ou AAV-DJ. Em algumas modalidades, o molde de DNA doador empacotado em AAV é administrado a um indivíduo, por exemplo, um paciente, primeiro por injeção IV periférica seguida pela nuclease

específica de sequência. A vantagem de entregar primeiro um molde de DNA doador empacotado em AAV é que o molde de DNA doador entregue será mantido de forma estável no núcleo dos hepatócitos transduzidos, o que permite a administração subsequente da nuclease específica da sequência que criará uma quebra de fita dupla no genoma com subsequente integração do DNA doador por HDR ou NHEJ. É desejável em algumas modalidades que a nuclease específica de sequência permaneça ativa na célula alvo apenas pelo tempo necessário para promover a integração direcionada do transgene em níveis suficientes para o efeito terapêutico desejado. Se a nuclease específica de sequência permanecer ativa na célula por um período prolongado, isto resultará em uma frequência aumentada de quebras de fita dupla em locais fora do alvo. Especificamente, a frequência da clivagem fora do alvo é uma função da eficiência de corte fora do alvo multiplicada pelo tempo durante o qual a nuclease está ativa. A entrega de uma nuclease específica de sequência na forma de um mRNA resulta em uma curta duração da atividade da nuclease no intervalo de horas a alguns dias porque o mRNA e a proteína traduzida têm vida curta na célula. Assim, se espera que a entrega da nuclease específica de sequência nas células que já contêm o molde doador resulte na maior razão possível de integração direcionada em relação à integração fora do alvo. Além disso, a entrega mediada por AAV de um molde de DNA doador no núcleo dos hepatócitos após a injeção periférica IV leva tempo, geralmente da ordem de 1 a 14 dias devido à necessidade de o vírus infectar a célula, escapar dos endossomos e depois transitar ao núcleo e conversão do genoma do AAV de fita simples em uma molécula de DNA de fita dupla pelos componentes do hospedeiro. Assim, é preferível, pelo menos em algumas modalidades, permitir que o processo de entrega do molde de DNA doador ao núcleo seja concluído antes de fornecer os componentes CRISPR-Cas9, uma vez

que estes componentes de nuclease só estarão ativos por cerca de 1 a 3 dias.

[00409] Em algumas modalidades, a nuclease específica de sequência é CRISPR-Cas9 que é composta por um sgRNA direcionado para uma sequência de DNA dentro do íntron 1 do gene da albumina juntamente com uma nuclease Cas9. Em algumas modalidades, a nuclease Cas9 é entregue como um mRNA que codifica a proteína Cas9 operacionalmente fundida com um ou mais sinais de localização nuclear (NLS). Em algumas modalidades, o sgRNA e o mRNA de Cas9 são entregues aos hepatócitos por empacotamento em uma nanopartícula lipídica. Em algumas modalidades, a nanopartícula lipídica contém o lipídio C12-200 (Love et al 2010, PNAS vol 107 1864–1869). Em algumas modalidades, a razão do sgRNA para o mRNA de Cas9 que é empacotado na LNP é de 1:1 (razão de massa) para resultar na clivagem máxima do DNA *in vivo* em camundongos. Em modalidades alternativas, diferentes razões de massa do sgRNA para o mRNA de Cas9 empacotado na LNP podem ser usadas, por exemplo, 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1, 3:1 ou 2:1 ou proporções inversas. Em algumas modalidades, o mRNA de Cas9 e o sgRNA são empacotados em formulações separadas de LNP e a LNP contendo o mRNA de Cas9 é entregue ao paciente cerca de 1 a cerca de 8 horas antes da LNP que contém o sgRNA para permitir o tempo ideal para a tradução do mRNA de Cas9 antes da entrega do sgRNA.

[00410] Em algumas modalidades, uma formulação de LNP encapsulando um gRNA e um mRNA de Cas9 ("a formulação de LNP-nuclease") é administrada a um indivíduo, por exemplo, um paciente, que precedentemente recebeu um molde de DNA doador empacotado em um AAV. Em algumas modalidades, a formulação de LNP-nuclease é administrada ao indivíduo dentro de 1 dia a 28 dias ou dentro de 7 dias a 28 dias ou dentro de 7 dias a 14 dias após a administração do AAV-

molde de DNA doador. O momento ideal de entrega da formulação de LNP-nuclease em relação ao AAV-molde de DNA doador pode ser determinado usando as técnicas conhecidas na técnica, por exemplo, estudos realizados em modelos animais, incluindo camundongos e macacos.

[00411] Em algumas modalidades, um molde de DNA doador é entregue aos hepatócitos de um indivíduo, por exemplo, um paciente, usando um método de entrega não viral. Embora alguns pacientes (tipicamente 30%) possuam anticorpos neutralizantes preexistentes direcionados aos sorotipos de AAV mais comumente usados que impedem a entrega eficaz de genes pelo referido AAV, todos os pacientes serão tratáveis com um método de entrega não viral. Várias metodologias de entrega não-virais são conhecidas na área. Em particular, sabe-se que as nanopartículas lipídicas (LNP) entregam eficientemente a sua carga encapsulada ao citoplasma de hepatócitos após injeção intravenosa em animais e humanos. Essas LNP são ativamente absorvidas pelo fígado através de um processo de endocitose mediada por receptores, resultando em absorção preferencial no fígado.

[00412] Em algumas modalidades, a fim de promover a localização nuclear de um molde doador, a sequência de DNA que pode promover a localização nuclear de plasmídeos, por exemplo, uma região de 366 pb da origem da replicação do vírus símio 40 (SV40) e o promotor inicial podem ser adicionados ao molde doador. Outras sequências de DNA que se ligam às proteínas celulares também podem ser usadas para melhorar a entrada nuclear de DNA.

[00413] Em algumas modalidades, um nível de expressão ou atividade do gene FVIII introduzido é medido no sangue de um indivíduo, por exemplo, um paciente, após a primeira administração de uma formulação de LNP-nuclease, por exemplo, contendo gRNA e

nuclease Cas9 ou mRNA que codifica a nuclease Cas9, após o AAV-molde de DNA doador. Se o nível de FVIII não for suficiente para curar a doença, como definido, por exemplo, com os níveis de FVIII de pelo menos 5 a 50%, em particular 5 a 20% dos níveis normais, uma segunda ou terceira administração da formulação de LNP-nuclease pode ser dado para promover uma integração direcionada adicional no local do íntron 1 da albumina. A viabilidade de usar doses múltiplas da formulação de LNP-nuclease para obter os níveis terapêuticos desejados de FVIII pode ser testada e otimizada usando as técnicas conhecidas na área, por exemplo, testes usando modelos animais, incluindo o camundongo e o macaco.

[00414] Em algumas modalidades, de acordo com qualquer um dos métodos descritos no presente documento, compreendendo a administração de i) um AAV-molde de DNA doador compreendendo um cassete doador e ii) uma formulação de LNP-nuclease a um indivíduo, é administrada uma dose inicial da formulação de LNP-nuclease ao indivíduo dentro de 1 dia a 28 dias após a administração do AAV-molde de DNA doador ao indivíduo. Em algumas modalidades, a dose inicial da formulação de LNP-nuclease é administrada ao indivíduo após um tempo suficiente para permitir a entrega do molde de DNA doador ao núcleo de uma célula alvo. Em algumas modalidades, a dose inicial da formulação de LNP-nuclease é administrada ao indivíduo após um tempo suficiente para permitir a conversão do genoma de AAV de fita simples em uma molécula de DNA de fita dupla no núcleo de uma célula alvo. Em algumas modalidades, uma ou mais (tal como 2, 3, 4, 5 ou mais) doses adicionais da formulação de LNP-nuclease são administradas ao indivíduo após a administração da dose inicial. Em algumas modalidades, uma ou mais doses da formulação de LNP-nuclease são administradas ao indivíduo até que seja alcançado um nível alvo de integração direcionada do cassete doador e/ou um nível

alvo de expressão do cassete doador. Em algumas modalidades, o método compreende ainda medir o nível de integração direcionada do cassete doador e/ou o nível de expressão do cassete doador após cada administração da formulação de LNP-nuclease e administrar uma dose adicional da formulação de LNP-nuclease se o nível alvo de integração direcionada do cassete dador e/ou o nível alvo de expressão do cassete dador não for atingido. Em algumas modalidades, a quantidade de pelo menos uma das uma ou mais doses adicionais da formulação de LNP-nuclease é a mesma que a dose inicial. Em algumas modalidades, a quantidade de pelo menos uma das uma ou mais doses adicionais da formulação de LNP-nuclease é menor que a dose inicial. Em algumas modalidades, a quantidade de pelo menos uma das uma ou mais doses adicionais da formulação de LNP-nuclease é mais do que a dose inicial.

CÉLULAS E POPULAÇÕES DE CÉLULAS GENETICAMENTE MODIFICADAS

[00415] Em um aspecto, as invenções a seguir fornecem um método de edição de um genoma em uma célula, criando assim uma célula geneticamente modificada. Em alguns aspectos, é fornecida uma população de células geneticamente modificadas. A célula geneticamente modificada refere-se, portanto, a uma célula que possui pelo menos uma modificação gênica introduzida pela edição do genoma (por exemplo, usando o sistema CRISPR/Cas9/Cpf1). Em algumas modalidades, a célula geneticamente modificada é uma célula de hepatócito geneticamente modificada. Uma célula geneticamente modificada com um ácido nucleico exógeno direcionado ao genoma e/ou um ácido nucleico exógeno que codifica um ácido nucleico direcionado ao genoma é contemplada no presente documento.

[00416] Em algumas modalidades, o genoma de uma célula pode ser editado inserindo uma sequência de ácido nucleico de um gene FVIII ou seu derivado funcional em uma sequência genômica da célula. Em

algumas modalidades, a célula sujeita à edição do genoma possui uma ou mais mutações no genoma, o que resulta na redução da expressão do gene endógeno FVIII em comparação com a expressão em um normal que não possui essa(s) mutação(ões). A célula normal pode ser uma célula saudável ou de controle que é originada (ou isolada) de um indivíduo diferente que não possui defeitos no gene FVIII. Em algumas modalidades, a célula sujeita à edição do genoma pode ser originada (ou isolada) de um indivíduo que precisa de tratamento da afecção ou distúrbio relacionado ao gene FVIII. Portanto, em algumas modalidades, a expressão do gene FVIII endógeno nessa célula é de cerca de 10%, cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90% ou cerca de 100% reduzido em comparação com a expressão da expressão endógena do gene FVIII na célula normal.

[00417] Após inserção bem-sucedida do transgene, por exemplo, um ácido nucleico que codifica um gene FVIII ou seu fragmento funcional, a expressão do gene FVIII introduzido ou seu derivado funcional na célula pode ser de pelo menos cerca de 10%, cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 100%, cerca de 200%, cerca de 300%, cerca de 400%, cerca de 500%, cerca de 600%, cerca de 700%, cerca de 800%, cerca de 900 %, cerca de 1.000%, cerca de 2.000%, cerca de 3.000%, cerca de 5.000%, cerca de 10.000% ou mais em comparação com a expressão do gene endógeno FVIII da célula. Em algumas modalidades, a atividade dos produtos do gene FVIII introduzidos, incluindo o fragmento funcional do FVIII na célula editada pelo genoma, pode ser de pelo menos cerca de 10%, cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90%, cerca de 100%, cerca de 200%, cerca de 300%, cerca de 400%, cerca de 500%, cerca de 600%, cerca de 700%, cerca

de 800%, cerca de 900%, cerca de 1.000%, cerca de 2.000%, cerca de 3.000%, cerca de 5.000%, cerca de 10.000% ou mais em comparação com a expressão do gene endógeno FVIII da célula. Em algumas modalidades, a expressão do gene FVIII introduzido ou seu derivado funcional na célula é pelo menos cerca de 2 vezes, cerca de 3 vezes, cerca de 4 vezes, cerca de 5 vezes, cerca de 6 vezes, cerca de 7 vezes, cerca de 8 vezes, cerca de 9 vezes, cerca de 10 vezes, cerca de 15 vezes, cerca de 20 vezes, cerca de 30 vezes, cerca de 50 vezes, cerca de 100 vezes, cerca de 1000 vezes ou mais da expressão do gene endógeno FVIII da célula. Além disso, em algumas modalidades, a atividade dos produtos do gene FVIII introduzidos, incluindo o fragmento funcional do FVIII na célula editada pelo genoma, pode ser comparável à, ou mais do que à atividade dos produtos do gene FVIII em uma célula normal e saudável.

[00418] Nas modalidades em que o tratamento ou a melhoria da Hemofilia A está em causa, os principais alvos para a edição de genes são as células humanas. Por exemplo, nos métodos *ex vivo* e nos métodos *in vivo*, as células humanas são hepatócitos. Em algumas modalidades, pela realização da edição de genes em células autólogas derivadas e, portanto, já completamente compatíveis com o paciente em necessidade, é possível gerar células que podem ser reintroduzidas com segurança no paciente e efetivamente dar origem a uma população de células que serão eficazes para melhorar uma ou mais condições clínicas associadas à doença do paciente. Em algumas modalidades de tais tratamentos, as células de hepatócitos podem ser isoladas de acordo com qualquer método conhecido na técnica e usadas para criar células terapeuticamente eficazes geneticamente modificadas. Em uma modalidade, as células-tronco hepáticas são geneticamente modificadas ex vivo e depois reintroduzidas no paciente onde elas darão origem a hepatócitos geneticamente modificados ou células endoteliais

sinusoidais que expressam o gene FVIII inserido.

ABORDAGEM TERAPÊUTICA

[00419] Em um aspecto, é no presente documento fornecida uma abordagem de terapia gênica para o tratamento da Hemofilia A em um paciente por edição do genoma do paciente. Em algumas modalidades, a abordagem de terapia gênica integra um gene FVIII funcional no genoma de um tipo de célula relevante em pacientes e isso pode fornecer uma cura permanente para a Hemofilia A. Em algumas modalidades, um tipo de célula indivíduo à abordagem de terapia gênica na qual se integra o gene FVIII é o hepatócito porque essas células expressam e secretam eficientemente muitas proteínas no sangue. Além disso, esta abordagem de integração usando hepatócitos pode ser considerada em pacientes pediátricos cujos fígados não estão totalmente crescidos porque o gene integrado seria transmitido às células filhas à medida que os hepatócitos se dividem.

[00420] Em outro aspecto, são fornecidos neste documento métodos celulares *ex vivo* e *in vivo* para o uso de ferramentas de manipulação do genoma para criar alterações permanentes no genoma ao modificar um gene que codifica o FVIII ou seu derivado funcional em um locus gênico em um genoma e restaurar a atividade proteica do FVIII. Tais métodos usam endonucleases, tais como nucleases associadas a CRISPR(CRISPR/Cas9, Cpf1 e similares), para permanentemente deletar, inserir, editar, corrigir ou substituir qualquer sequência de um genoma ou inserir uma sequência exógena, por exemplo, um gene que codifica o FVIII em um locus genômico. Desta maneira, os exemplos apresentados na presente invenção restauram a atividade do gene FVIII com um único tratamento (em vez de administrar terapias potenciais durante toda a vida do paciente).

[00421] Em algumas modalidades, uma terapia baseada em células *ex vivo* é feita usando um hepatócito que é isolado de um paciente. Em

seguida, o DNA cromossômico destas células é editado usando os materiais e métodos no presente documento descritos. Finalmente, as células editadas são implantadas no paciente.

[00422] Uma vantagem de uma abordagem de terapia celular *ex vivo* é a capacidade de realizar uma análise abrangente da terapêutica antes da administração. Todas as terapêuticas baseadas em nuclease têm algum nível de efeitos fora do alvo. A realização da correção gênica *ex vivo* permite caracterizar completamente a população celular corrigida antes da implantação. Aspectos da invenção incluem sequenciar todo o genoma das células corrigidas para garantir que os cortes fora do alvo, se houver, estejam em localizações genômicas associadas a um risco mínimo para o paciente. Além disso, populações de células específicas, incluindo populações clonais, podem ser isoladas antes da implantação.

[00423] Outra modalidade desse método é uma terapia com base *in vivo*. Neste método, o DNA cromossômico das células no paciente é corrigido usando os materiais e métodos no presente documento descritos. Em algumas modalidades, as células são hepatócitos.

[00424] Uma vantagem da terapia gênica *in vivo* é a facilidade de produção e administração terapêutica. A mesma abordagem terapêutica e terapia pode ser usada para tratar mais de um paciente, por exemplo, um número de pacientes que partilham o mesmo ou similar genótipo ou alelo. Por outro lado, a terapia celular *ex vivo* normalmente usa as células de um paciente, que são isoladas, manipuladas e devolvidas ao mesmo paciente.

[00425] Em algumas modalidades, o indivíduo que está com necessidade do método de tratamento de acordo com as invenções é um paciente com sintomas de Hemofilia A. Em algumas modalidades, o indivíduo pode ser um ser humano suspeito de ter Hemofilia A. Em alternativa, o indivíduo pode ser um ser humano diagnosticado com um risco de Hemofilia A. Em algumas modalidades, o indivíduo que está

com necessidade do tratamento pode ter um ou mais defeitos genéticos (por exemplo, deleção, inserção e/ou mutação) no gene endógeno FVIII ou em suas sequências reguladoras, de modo que a atividade incluindo o nível de expressão ou funcionalidade da proteína FVIII é substancialmente reduzida em comparação com um indivíduo saudável, normal.

[00426] Em algumas modalidades, é no presente documento fornecido um método de tratamento da Hemofilia A em um indivíduo, o método compreendendo fornecer o seguinte a uma célula no indivíduo: (a) um RNA guia (gRNA) direcionado ao locus da albumina no genoma da célula (b) uma endonuclease de DNA ou um ácido nucleico que codifica a referida endonuclease de DNA; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional Em algumas modalidades, o gRNA tem como alvo o íntron 1 do gene da albumina. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104.

[00427] Em algumas modalidades, é no presente documento fornecido um método de tratamento da Hemofilia A em um indivíduo, o método compreendendo fornecer o seguinte a uma célula no indivíduo: (a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 18-44 e 104; (b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e (c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 21, 22, 28 e 30. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22. Em

algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28. Em algumas modalidades, o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30. Em algumas modalidades, a célula é uma célula humana, por exemplo, uma célula de hepatócito humana. Em algumas modalidades, o indivíduo é um paciente com ou é suspeito de ter Hemofilia A. Em algumas modalidades, o indivíduo é diagnosticado com um risco de Hemofilia A.

[00428] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1, ou um seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, a Cas9 é de *Streptococcus pyogenes* (spCas9). Em algumas modalidades, a Cas9 é de *Staphylococcus lugdunensis* (SluCas9).

[00429] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão em uma célula. Em algumas modalidades, a célula é uma célula humana.

[00430] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, o método usa um ácido nucleico que codifica a endonuclease de DNA. Em algumas modalidades, o ácido nucleico que codifica a

endonuclease de DNA é otimizado quanto a códon para expressão na célula. Em algumas modalidades, a célula é uma célula humana, por exemplo, uma célula de hepatócito humana. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é DNA, tal como um plasmídeo de DNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um RNA, tal como mRNA.

[00431] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV). Em algumas modalidades, o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA. Em algumas modalidades, o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA. Em algumas modalidades, o local alvo de gRNA é um local alvo para o gRNA de (a). Em algumas modalidades, o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA do genoma da célula para o gRNA de (a). Em algumas modalidades, fornecer o molde doador à célula compreende administrar o molde doador ao indivíduo. Em algumas modalidades, a administração é por via intravenosa.

[00432] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, o lipossomo ou nanopartícula lipídica também compreende o gRNA. Em algumas modalidades, fornecer o gRNA e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA à célula compreende administrar o

lipossomo ou nanopartícula lipídica ao indivíduo. Em algumas modalidades, a administração é por via intravenosa. Em algumas modalidades, o lipossomo ou nanopartícula lipídica é uma nanopartícula lipídica. Em algumas modalidades, o método usa uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA.

[00433] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a endonuclease de DNA é pré-complexada com o gRNA, formando um complexo de ribonucleoproteína (RNP).

[00434] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 17 dias após o molde doador de (c) ser fornecido à célula. Em algumas modalidades, o fornecimento de (a) e (b) à célula compreende a administração (tal como por via intravenosa) ao indivíduo de uma nanopartícula lipídica compreendendo ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas

modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA. Em algumas modalidades, o fornecimento de (c) à célula compreende a administração (tal como por via intravenosa) ao indivíduo do molde doador codificado em um vetor AAV.

[00435] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b). Em algumas modalidades, uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado. Em algumas modalidades, o fornecimento de (a) e (b) à célula compreende a administração (tal como por via intravenosa) ao indivíduo de uma nanopartícula lipídica compreendendo ácido nucleico que codifica a endonuclease de DNA e o gRNA. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é um mRNA que codifica a endonuclease de DNA.

[00436] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é expressa sob o controle do

promotor endógeno da albumina.

[00437] Em algumas modalidades, de acordo com qualquer um dos métodos de tratamento da Hemofilia A no presente documento descritos, a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é expressa no fígado do indivíduo.

IMPLANTANDO CÉLULAS EM UM indivíduo

[00438] Em algumas modalidades, os métodos *ex vivo* da invenção envolvem a implantação das células editadas por genoma em um indivíduo que está com necessidade de tal método. Esta etapa de implantação pode ser realizada usando qualquer método de implantação conhecido na técnica. Por exemplo, as células geneticamente modificadas podem ser injetadas diretamente no sangue do indivíduo ou de outra forma administradas ao indivíduo.

[00439] Em algumas modalidades, os métodos no presente documento descritos incluem a administração, que pode ser usada indistintamente com "introduzindo" e "transplantando" células terapêuticas geneticamente modificadas em um indivíduo, por um método ou via que resulta em pelo menos localização parcial das células introduzidas em um local desejado, de modo que um efeito(s) desejado(s) seja(m) produzido(s). As células terapêuticas ou sua progênie diferenciada podem ser administradas por qualquer via apropriada que resulte na entrega a um local desejado no indivíduo, onde pelo menos uma porção das células implantadas ou das células componentes permanecem viáveis. O período de viabilidade das células após a administração a um indivíduo pode ser tão curto quanto algumas horas, por exemplo, vinte e quatro horas, a alguns dias, até vários anos, ou até o tempo de vida do paciente, ou seja, enxerto de longo prazo.

[00440] Quando fornecidas profilaticamente, as células terapêuticas descritas no presente documento podem ser administradas a um

indivíduo antes de qualquer sintoma de Hemofilia A. Em conformidade, em algumas modalidades, a administração profilática de uma população de células de hepatócitos geneticamente modificadas serve para impedir a ocorrência de sintomas de Hemofilia A.

[00441] Quando fornecidas terapeuticamente em algumas modalidades, as células de hepatócitos geneticamente modificadas são fornecidas no (ou após) início de um sintoma ou indicação de Hemofilia A, por exemplo, no início da doença.

[00442] Em algumas modalidades, uma população de células de hepatócitos terapêuticas sendo administrada de acordo com os métodos descritos no presente documento tem células de hepatócitos alogênicas obtidas de um ou mais doadores. "Alogênico" refere-se a uma célula de hepatócito ou amostras biológicas com células de hepatócitos obtidas de um ou mais doadores diferentes da mesma espécie, onde os genes em um ou mais loci não são idênticos. Por exemplo, uma população de células de hepatócitos sendo administrada a um indivíduo pode ser derivada de um ou mais indivíduos doadores não relacionados ou de um ou mais irmãos não idênticos. Em algumas modalidades, populações de células de hepatócitos singênicas podem ser usadas, tal como aquelas obtidas de animais geneticamente idênticos ou de gêmeos idênticos. Em outras modalidades, as células de hepatócitos são células autólogas; isto é, as células de hepatócitos são obtidas ou isoladas de um indivíduo e administradas ao mesmo indivíduo, isto é, o doador e o receptor são os mesmos.

[00443] Em uma modalidade, uma quantidade eficaz refere-se à quantidade de uma população de células terapêuticas necessárias para prevenir ou aliviar pelo menos um ou mais sinais ou sintomas da Hemofilia A, e refere-se a uma quantidade suficiente de uma composição para fornecer o efeito desejado, por exemplo, para tratar um indivíduo com Hemofilia A. Em modalidades, uma quantidade

terapeuticamente eficaz refere-se, portanto, a uma quantidade de células terapêuticas ou a uma composição tendo células terapêuticas que é suficiente para promover um efeito específico quando administrada a um indivíduo típico, como aquele que tem ou corre risco de Hemofilia A. Uma quantidade eficaz também incluiria uma quantidade suficiente para impedir ou atrasar o desenvolvimento de um sintoma da doença, alterar o curso de um sintoma da doença (por exemplo, mas não limitado a retardar a progressão de um sintoma da doença) ou reverter um sintoma da doença. É entendido que para qualquer caso, uma quantidade eficaz apropriada pode ser determinada por um vulgar perito na técnica usando experimentação de rotina.

[00444] Para uso nas várias modalidades descritas no presente documento, uma quantidade eficaz de células terapêuticas, por exemplo, células de hepatócitos editadas por genoma podem ser pelo menos 10^2 células, pelo menos 5 X 10^2 células, pelo menos 10^3 células, pelo menos 5 X 10^3 células, pelo menos 10^4 células, pelo menos 5 X 10^4 células, pelo menos 10^5 células, pelo menos 2 X 10^5 células, pelo menos 3 X 10^5 células, pelo menos 4 X 10^5 células, pelo menos 5 X 10^5 células, pelo menos 6 X 10^5 células, pelo menos 7 X 10^5 células, pelo menos 8 X 10^5 células, pelo menos 9 X 10^5 células, pelo menos 1 X 10^6 células, pelo menos 2 X 10^6 células, pelo menos 3 X 10^6 células, pelo menos 4 X 10^6 células, pelo menos 5 X 10^6 células, pelo menos 6 X 10^6 células, pelo menos 7 X 10^6 células, pelo menos 8 X 10^6 células, pelo menos 9 X 10^6 células ou seus múltiplos. As células terapêuticas podem ser derivadas de um ou mais doadores, ou são obtidas de uma fonte autóloga. Em algumas modalidades descritas no presente documento, as células terapêuticas são expandidas em cultura antes da administração a um indivíduo em necessidade das mesmas.

[00445] Em algumas modalidades, aumentos modestos e incrementais nos níveis de FVIII funcional expressos em células de

pacientes com Hemofilia A podem ser benéficos para melhorar um ou mais sintomas da doença, aumentar a sobrevivência a longo prazo e/ou reduzir os efeitos secundários associados com outros tratamentos. Após a administração dessas células a pacientes humanos, é benéfica a presença de células terapêuticas que produzem níveis aumentados de FVIII funcional. Em algumas modalidades, o tratamento eficaz de um indivíduo dá origem a pelo menos cerca de 1%, 3%, 5% ou 7% de FVIII funcional em relação ao total de FVIII no indivíduo tratado. Em algumas modalidades, a FVIII funcional é pelo menos cerca de 10% do total de FVIII. Em algumas modalidades, a FVIII funcional é pelo menos, aproximadamente ou no máximo 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% ou 100% do total de FVIII. Da mesma forma, a introdução de subpopulações de células relativamente limitadas com níveis significativamente elevados de FVIII funcional pode ser benéfica em vários pacientes porque em algumas situações as células normalizadas terão uma vantagem seletiva em relação às células doentes. No entanto, mesmo níveis modestos de células terapêuticas com níveis elevados de FVIII funcional podem ser benéficos para melhorar um ou mais aspectos da Hemofilia A em pacientes. Em algumas modalidades, cerca de 10%, cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 80%, cerca de 90% ou mais da terapêutica em pacientes para os quais essas células administradas estão produzindo níveis aumentados de FVIII funcional.

[00446] Em modalidades, a entrega de uma composição celular terapêutica a um indivíduo por um método ou via resulta em pelo menos localização parcial da composição celular em um local desejado. Uma composição celular pode ser administrada por qualquer via apropriada que resulte em tratamento eficaz no indivíduo, isto é, a administração resulta na entrega em um local desejado no indivíduo onde pelo menos uma porção da composição é entregue, isto é, pelo menos 1 x 10^4

células são entregues para o local desejado por um período de tempo. Os modos de administração incluem injeção, infusão, instilação ou ingestão. "Injeção" inclui, sem limitação, injeção e infusão intravenosa, intramuscular, intra-arterial, intratecal, intraventricular, intracapsular, intraorbital, intracardíaca, intradérmica, intraperitoneal, transtraqueal, subcutânea, subcuticular, intra-articular, subcapsular, subaracnoidea, intraespinhal, intracérebro-espinhal e intraesternal. Em algumas modalidades, a via é intravenosa. Para a entrega das células, pode ser feita a administração por injeção ou infusão.

[00447] Em uma modalidade, as células são administradas sistemicamente, em outras palavras, uma população de células terapêuticas é administrada de modo diferente que diretamente em um local, tecido ou órgão alvo, de modo que entre no sistema circulatório do indivíduo e, portanto, esteja sujeita ao metabolismo e outros processos similares.

[00448] A eficácia de um tratamento tendo uma composição para o tratamento da Hemofilia A pode ser determinada pelo médico especialista. No entanto, um tratamento é considerado tratamento efetivo se qualquer um ou todos os sinais ou sintomas de, como apenas um exemplo, os níveis de FVIII funcional forem alterados de uma maneira benéfica (por exemplo, aumentados em pelo menos 10%) ou outros sintomas ou marcadores de doença clinicamente aceites são otimizados ou melhorados. A eficácia também pode ser medida pela falha de um indivíduo em piorar, como avaliado por hospitalização ou necessidade de intervenções médicas (por exemplo, a progressão da doença é interrompida ou pelo menos desacelerada). Os métodos para medir estes indicadores são conhecidos dos peritos na técnica e/ou no presente documento descritos. O tratamento inclui qualquer tratamento de uma doença em um indivíduo ou um animal (alguns exemplos não limitativos incluem um ser humano ou um mamífero) e inclui: (1) inibir a

doença, por exemplo, interromper ou retardar a progressão dos sintomas; ou (2) aliviar a doença, por exemplo, causando regressão dos sintomas; e (3) prevenir ou reduzir a probabilidade de desenvolvimento de sintomas.

COMPOSIÇÃO

[00449] Em um aspecto, a presente invenção fornece composições para levar a cabo os métodos no presente documento descritos. Uma composição pode incluir um ou mais dos seguintes itens: um ácido nucleico direcionado ao genoma (por exemplo, gRNA); um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA) ou uma sequência nucleotídica que codifica o polipeptídeo direcionado ao local; e um polinucleotídeo a ser inserido (por exemplo, um molde doador) para efetuar a modificação gênica desejada dos métodos no presente documento descritos.

[00450] Em algumas modalidades, uma composição tem uma sequência nucleotídica que codifica um ácido nucleico direcionado ao genoma (por exemplo, gRNA).

[00451] Em algumas modalidades, uma composição tem um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA). Em algumas modalidades, uma composição tem uma sequência de nucleotídeos que codifica o polipeptídeo direcionado ao local.

[00452] Em algumas modalidades, uma composição tem um polinucleotídeo (por exemplo, um molde doador) para ser inserido em um genoma.

[00453] Em algumas modalidades, uma composição tem (i) uma sequência de nucleotídeos que codifica um ácido nucleico direcionado ao genoma (por exemplo, gRNA) e (ii) um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA) ou uma sequência nucleotídica que codifica o polipeptídeo direcionado ao local.

[00454] Em algumas modalidades, uma composição tem (i) uma

sequência de nucleotídeos que codifica um ácido nucleico direcionado ao genoma (por exemplo, gRNA) e (ii) um polinucleotídeo (por exemplo, um molde doador) a ser inserido em um genoma.

[00455] Em algumas modalidades, uma composição tem (i) um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA) ou uma sequência de nucleotídeos que codifica o polipeptídeo direcionado ao local e (ii) um polinucleotídeo (por exemplo, um molde doador) a ser inserido em um genoma.

[00456] Em algumas modalidades, uma composição tem (i) uma sequência de nucleotídeos que codifica um ácido nucleico direcionado ao genoma (por exemplo, gRNA), (ii) um polipeptídeo direcionado ao local (por exemplo, endonuclease de DNA) ou uma sequência nucleotídica que codifica o polipeptídeo direcionado ao local e (iii) um polinucleotídeo (por exemplo, um molde doador) a ser inserido em um genoma.

[00457] Em algumas modalidades de qualquer uma das composições acima, a composição tem um ácido nucleico guia de molécula única direcionado ao genoma. Em algumas modalidades de qualquer uma das composições acima, a composição tem um ácido nucleico de molécula dupla direcionado ao genoma. Em algumas modalidades de qualquer uma das composições acima, a composição tem dois ou mais guias de molécula dupla ou guias de molécula única. Em algumas modalidades, a composição tem um vetor que codifica o ácido nucleico direcionado ao ácido nucleico. Em algumas modalidades, o ácido nucleico direcionado ao genoma é uma endonuclease de DNA, em particular, Cas9.

[00458] Em algumas modalidades, uma composição pode conter composição que inclui um ou mais gRNA que podem ser usados para edição de genoma, em particular, inserção de um gene FVIII ou seu derivado em um genoma de uma célula. O gRNA para a composição

pode ter como alvo um local no genoma no, dentro ou próximo do gene da albumina endógena. Portanto, em algumas modalidades, o gRNA pode ter uma sequência espaçadora complementar a uma sequência genômica no, dentro ou perto do gene da albumina.

[00459] Em algumas modalidades, um gRNA para uma composição é uma sequência selecionada entre aquelas listadas na Tabela 3 e suas variantes tendo pelo menos cerca de 50%, cerca de 55%, cerca de 60%, cerca de 65%, cerca de 70%, cerca de 75%, cerca de 80%, cerca de 85%, cerca de 90% ou cerca de 95% de identidade ou homologia com as listadas na Tabela 3. Em algumas modalidades, as variantes do gRNA para o kit têm pelo menos cerca de 85% de homologia com qualquer uma das listadas na Tabela 3.

[00460] Em algumas modalidades, um gRNA para uma composição tem uma sequência espaçadora que é complementar a um local alvo no genoma. Em algumas modalidades, a sequência espaçadora tem 15 bases a 20 bases de comprimento. Em algumas modalidades, uma complementaridade entre a sequência espaçadora e a sequência genômica é de pelo menos 80%, pelo menos 85%, pelo menos 90%, pelo menos 95%, pelo menos 96%, pelo menos 97%, pelo menos 98%, pelo menos 99% ou pelo menos 100%.

[00461] Em algumas modalidades, uma composição pode ter uma endonuclease de DNA ou um ácido nucleico que codifica a endonuclease de DNA e/ou um molde doador que tem uma sequência de ácido nucleico de um gene FVIII ou seu derivado funcional. Em algumas modalidades, a endonuclease de DNA é Cas9. Em algumas modalidades, o ácido nucleico que codifica a endonuclease de DNA é DNA ou RNA.

[00462] Em algumas modalidades, um ou mais de qualquer uma das oligonucleotídeos ou sequências de ácidos nucleicos para o kit podem ser codificados em um vetor de Vírus Adeno-Associado (AAV). Portanto,

em algumas modalidades, um gRNA pode ser codificado em um vetor AAV. Em algumas modalidades, um ácido nucleico que codifica uma endonuclease de DNA pode ser codificado em um vetor AAV. Em algumas modalidades, um molde doador pode ser codificado em um vetor AAV. Em algumas modalidades, dois ou mais oligonucleotídeos ou sequências de ácidos nucleicos podem ser codificados em um único vetor AAV. Assim, em algumas modalidades, uma sequência de gRNA e um ácido nucleico que codifica a endonuclease de DNA podem ser codificados em um único vetor AAV.

[00463] Em algumas modalidades, uma composição pode ter um lipossomo ou uma nanopartícula lipídica. Portanto, em algumas modalidades, qualquer uma das compostos (por exemplo, uma endonuclease de DNA ou um ácido nucleico que a codifica, gRNA e molde doador) da composição podem ser formulados em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, um ou mais desses compostos estão associados a um lipossomo ou nanopartícula lipídica através de uma ligação covalente ou ligação não covalente. Em algumas modalidades, qualquer um dos compostos pode estar separadamente ou conjuntamente contido em um lipossomo ou nanopartícula lipídica. Portanto, em algumas modalidades, cada um de uma endonuclease de DNA ou um ácido nucleico que a codifica, gRNA e molde doador é formulado separadamente em um lipossomo ou nanopartícula lipídica. Em algumas modalidades, uma endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica com gRNA. Em algumas modalidades, uma endonuclease de DNA ou um ácido nucleico que codifica a mesma, gRNA e molde doador são formulados em um lipossomo ou nanopartícula lipídica em conjunto.

[00464] Em algumas modalidades, uma composição descrita acima tem ainda um ou mais reagentes adicionais, onde esses reagentes adicionais são selecionados a partir de um tampão, um tampão para a

introdução de um polipeptídeo ou polinucleotídeo em uma célula, um tampão de lavagem, um reagente de controle, um vetor de controle, um polinucleotídeo de RNA de controle, um reagente para produção *in vitro* do polipeptídeo a partir de DNA, adaptadores para sequenciamento e similares. Um tampão pode ser um tampão de estabilização, um tampão de reconstituição, um tampão de diluição ou similares. Em algumas modalidades, uma composição também pode incluir um ou mais componentes que podem ser usados para facilitar ou potenciar a ligação no alvo ou a clivagem do DNA pela endonuclease, ou melhorar a especificidade do direcionamento.

[00465] Em algumas modalidades, qualquer uma das componentes de uma composição são formulados com excipientes farmaceuticamente aceitáveis, como transportadores, solventes, estabilizadores, adjuvantes, diluentes, etc., dependendo do modo particular de administração e forma de dosagem. Em modalidades, as composições de RNA guia são geralmente formuladas para atingir um pH fisiologicamente compatível e variam de um pH de cerca de 3 a um pH de cerca de 11, cerca de pH 3 a cerca de pH 7, dependendo da formulação e via de administração. Em algumas modalidades, o pH é ajustado para uma gama de cerca de pH 5,0 a cerca de pH 8. Em algumas modalidades, a composição tem uma quantidade terapeuticamente eficaz de pelo menos um composto como no presente documento descrito, juntamente com um ou mais excipientes farmaceuticamente aceitáveis. Opcionalmente, a composição pode ter uma combinação dos compostos descritos no presente documento, ou pode incluir um segundo ingrediente ativo útil no tratamento ou prevenção do crescimento bacteriano (por exemplo, e sem limitação, agentes antibacterianos ou antimicrobianos), ou pode incluir uma combinação de reagentes da invenção. Em algumas modalidades, os gRNAs são formulados com outros, um ou mais, oligonucleotídeos, por

exemplo, um ácido nucleico que codifica a endonuclease de DNA e/ou um molde doador. Em alternativa, um ácido nucleico que codifica a endonuclease de DNA e um molde doador, separadamente ou em combinação com outros oligonucleotídeos, são formulados com o método descrito acima para a formulação de gRNA.

[00466] Os excipientes adequados podem incluir, por exemplo, moléculas transportadoras que incluem macromoléculas grandes e lentamente metabolizadas, tais como proteínas, polissacarídeos, ácidos poliláticos, ácidos poliglicólicos, aminoácidos poliméricos, copolímeros de aminoácidos e partículas de vírus inativas. Outros excipientes exemplificativos incluem antioxidantes (por exemplo, e sem limitação, ácido ascórbico), agentes quelantes (por exemplo e sem limitação, EDTA), carboidratos (por exemplo e sem limitação, dextrina, hidroxialquilcelulose e hidroxialquilmetilcelulose), ácido esteárico, líquidos (por exemplo e sem limitação, óleos, água, solução salina, glicerol e etanol), agentes umectantes ou emulsificantes, substâncias tamponantes de pH e similares.

[00467] Em algumas modalidades, qualquer uma dos compostos (por exemplo, uma endonuclease de DNA ou um ácido nucleico que a codifica, gRNA e molde doador) de uma composição podem ser entregues através de transfecção, como eletroporação. Em algumas modalidades exemplificativas, uma endonuclease de DNA pode ser pré-complexada com um gRNA, formando um complexo de Ribonucleoproteína (RNP), antes do fornecimento para a célula e o complexo de RNP pode ser eletroporado. Em tais modalidades, o molde doador pode ser entregue através eletroporação.

[00468] Em algumas modalidades, uma composição refere-se a uma composição terapêutica tendo células terapêuticas que são usadas em um método de tratamento *ex vivo*.

[00469] Nas modalidades, as composições terapêuticas contêm um

transportador fisiologicamente tolerável juntamente com a composição de células e, opcionalmente, pelo menos um agente bioativo adicional como descrito no presente documento, dissolvido ou disperso na mesma como um ingrediente ativo. Em algumas modalidades, a composição terapêutica não é substancialmente imunogênica quando administrada a um mamífero ou paciente humano para fins terapêuticos, a menos que desejado.

[00470] Em geral, as células terapêuticas geneticamente modificadas no presente documento descritas são administradas como uma suspensão com um transportador farmaceuticamente aceitável. Um perito na técnica reconhecerá que um transportador farmaceuticamente aceitável para ser usado em uma composição de células não incluirá tampões, compostos, agentes de criopreservação, conservantes ou outros agentes em quantidades que interfiram substancialmente na viabilidade das células a serem administradas ao indivíduo. Uma formulação com células pode incluir, por exemplo, tampões osmóticos que permitem a manutenção da integridade da membrana celular e, opcionalmente, nutrientes para manter a viabilidade celular ou potenciar o enxerto após a administração. Tais formulações e suspensões são conhecidas dos peritos na técnica e/ou podem ser adaptadas para uso com as células progenitoras, como descrito no presente documento, usando experimentação de rotina.

[00471] Em algumas modalidades, uma composição de células também pode ser emulsionada ou apresentada como uma composição lipossômica, desde que o procedimento de emulsificação não afete adversamente a viabilidade celular. As células e qualquer outro ingrediente ativo podem ser misturados com excipientes que são farmaceuticamente aceitáveis e compatíveis com o ingrediente ativo, e em quantidades adequadas para uso nos métodos terapêuticos no presente documento descritos.

[00472] Agentes adicionais incluídos em uma composição de células podem incluir sais farmaceuticamente aceitáveis dos seus componentes. Os sais farmaceuticamente aceitáveis incluem os sais de adição de ácidos (formados com os grupos amino livres do polipeptídeo) que são formados com ácidos inorgânicos, tais como, por exemplo, ácidos clorídrico ou fosfórico ou ácidos orgânicos como acético, tartárico, mandélico e similares. Os sais formados com os grupos carboxila livres também podem ser derivados de bases inorgânicas, tal como, por exemplo, hidróxido de sódio, de potássio, de amônio, de cálcio ou férrico e bases orgânicas como isopropilamina, trimetilamina, 2-etilamino etanol, histidina, procaína e similares.

[00473] Os transportadores fisiologicamente toleráveis são bem conhecidos na técnica. Transportadores líquidos exemplificativos são soluções aquosas estéreis que não contêm materiais além dos ingredientes ativos e água, ou contêm um tampão tal como fosfato de sódio a valor fisiológico de pH, solução salina fisiológica ou ambos, tal como solução salina tamponada com fosfato. Além disso, os transportadores aquosos podem conter mais de um sal tampão, bem como sais como cloretos de sódio e potássio, dextrose, polietilenoglicol e outros solutos. As composições líquidas também podem conter fases líquidas além e com a exclusão de água. Exemplos de tais fases líquidas adicionais são glicerina, óleos vegetais, como óleo de semente de algodão e emulsões óleo em água. A quantidade de um composto ativo usado nas composições de células que é eficaz no tratamento de um distúrbio ou afecção particular irá depender da natureza do distúrbio ou afecção e pode ser determinada por técnicas clínicas padrão.

KIT

[00474] Algumas modalidades fornecem um kit que contém qualquer uma das composições descritas acima, por exemplo, uma composição para edição do genoma ou uma composição de células terapêutica e

um ou mais componentes adicionais.

[00475] Em algumas modalidades, um kit pode ter um ou mais agentes terapêuticos adicionais que podem ser administrados simultaneamente ou em sequência com a composição para uma finalidade desejada, por exemplo, edição do genoma ou terapia celular.

[00476] Em algumas modalidades, um kit pode incluir adicionalmente instruções para o uso dos componentes do kit para a prática dos métodos. As instruções para a prática dos métodos são geralmente gravadas em um meio de gravação adequado. Por exemplo, as instruções podem ser impressas em um substrato, tal como papel ou plástico, etc. As instruções podem estar presentes nos kits como um folheto informativo, na rotulagem do recipiente do kit ou seus componentes (isto é, associados a embalagem ou subembalagem), etc. As instruções podem estar presentes como um ficheiro de dados de armazenamento eletrônico presente em um meio de armazenamento legível por computador adequado, por exemplo, CD-ROM, disquete, unidade flash, etc. Em alguns casos, as instruções reais não estão presentes no kit, mas podem ser fornecidos meios para obter as instruções de uma fonte remota (por exemplo, através da Internet). Um exemplo desta modalidade é um kit que inclui um endereço da Web em que as instruções podem ser visualizadas e/ou do qual as instruções podem ser transferidas. Como nas instruções, este meio para obter as instruções pode ser gravado em um substrato adequado.

OUTRAS ABORDAGENS TERAPÊUTICAS POSSÍVEIS

[00477] A edição de genes pode ser realizada usando nucleases desenhadas para direcionar sequências específicas. Até o momento, existem quatro tipos principais de nucleases: meganucleases e seus derivados, nucleases de dedo de zinco (ZFNs), nucleases efetoras similares a ativador de transcrição (TALENs) e sistemas de nuclease CRISPR-Cas9. As plataformas de nuclease variam em dificuldade de

concepção, densidade alvo e o modo de ação, particularmente porque a especificidade de ZFNs e TALENs é através de interações proteína-DNA, enquanto as interações RNA-DNA orientam principalmente Cas9. A clivagem Cas9 também requer um motivo adjacente, o PAM, que difere entre os diferentes sistemas CRISPR. Cas9 de *Streptococcus pyogenes* cliva usando um NRG PAM, CRISPR de *Neisseria meningitidis* pode clivar em locais com PAMs incluindo NNNNGATT (SEQ ID NO: 101), NNNNNGTTT (SEQ ID NO: 102) e NNNNGCTT (SEQ ID NO: 103). Vários outros ortólogos de Cas9 têm como alvo o protoespaçador adjacente aos PAMs alternativos.

[00478] As endonucleases CRISPR, tal como Cas9, podem ser usadas em várias modalidades dos métodos da invenção. No entanto, os ensinamentos descritos neste documento, tal como locais alvo terapêuticos, podem ser aplicados a outras formas de endonucleases, como ZFNs, TALENs, HEs ou MegaTALs, ou usando combinações de nucleases. No entanto, para aplicar os ensinamentos da presente invenção a essas endonucleases, seria necessário, entre outras coisas, desenhar proteínas direcionadas para os locais alvo específicos.

[00479] Domínios de ligação adicionais podem ser fundidos à proteína Cas9 para aumentar a especificidade. Os locais alvo destes construtos mapeariam para o local especificado de gRNA identificado, mas exigiriam motivos de ligação adicionais, como para um domínio de dedo de zinco. No caso da Mega-TAL, uma meganuclease pode ser fundida com um domínio de ligação ao DNA TALE. O domínio da meganuclease pode aumentar a especificidade e fornecer a clivagem. Da mesma forma, a Cas9 inativada ou morta (dCas9) pode ser fundida com um domínio de clivagem e requer o local alvo sgRNA/Cas9 e o local de ligação adjacente para o domínio de ligação ao DNA fundido. Isto provavelmente exigiria alguma manipulação de proteína de dCas9, além da inativação catalítica, para diminuir a ligação sem o local de ligação

adicional.

[00480] Em algumas modalidades, as composições e métodos de edição do genoma, de acordo com as presentes invenções (por exemplo, inserção de uma sequência de codificação de FVIII no locus da albumina), podem usar ou ser realizados usando qualquer uma das seguintes abordagens.

Nucleases de Dedo de Zinco

[00481] As nucleases de dedo de zinco (ZFNs) são proteínas modulares com um domínio de ligação de DNA de dedo de zinco manipulado ligado ao domínio catalítico da endonuclease tipo II FokI. Como a FokI funciona apenas como um dímero, um par de ZFNs deve ser manipulado para se ligar às sequências cognatas de "metade-local" alvo em fitas de DNA opostas e com espaçamento preciso entre elas para permitir a formação do dímero FokI cataliticamente ativo. Após a dimerização do domínio FokI, que por si só não possui especificidade de sequência, é gerada uma quebra de fita dupla de DNA entre os locais-metade ZFN como o passo inicial na edição do genoma.

[00482] O domínio de ligação ao DNA de cada ZFN geralmente possui 3-6 dedos de zinco da abundante arquitetura Cys2-His2, com cada dedo reconhecendo principalmente um tripleto de nucleotídeos em uma fita da sequência de DNA alvo, embora a interação cruzada da fita com um quarto nucleotídeo também pode ser importante. A alteração dos aminoácidos de um dedo em posições que fazem contatos chave com o DNA altera a especificidade da sequência de um dado dedo. Assim, uma proteína de dedo de zinco com quatro dedos reconhecerá seletivamente uma sequência alvo de 12 pb, onde a sequência alvo é um composto das preferências de tripletos contribuídas por cada dedo, embora a preferência de tripletos possa ser influenciada em graus variados pelos dedos vizinhos. Um aspecto importante das ZFNs é que elas podem ser facilmente redirecionadas para quase qualquer

endereço genômico, simplesmente modificando dedos individuais, embora seja necessário um conhecimento considerável para fazer isso bem. Na maioria das aplicações de ZFNs, proteínas de 4-6 dedos são usadas, reconhecendo 12-18 pb, respectivamente. Portanto, um par de ZFNs reconhecerá tipicamente uma sequência alvo combinada de 24-36 pb, não incluindo o espaçador de 5-7 pb entre os locais-metade. Os locais de ligação podem ser separados ainda mais com espaçadores maiores, incluindo 15-17 pb. É provável que uma sequência alvo deste comprimento seja única no genoma humano, assumindo que sequências repetitivas ou homólogos de genes sejam excluídos durante o processo de projeção. No entanto, as interações proteína ZFN-DNA não são absolutas em sua especificidade, portanto ocorrem eventos de ligação e clivagem fora do alvo, como um heterodímero entre os dois ZFNs ou como um homodímero de um ou outro dos ZFNs. A última possibilidade foi efetivamente eliminada pela manipulação da interface de dimerização do domínio FokI para criar variantes "mais" e "menos", também conhecidas como variantes obrigatórias de heterodímero, que só podem dimerizar entre si e não consigo mesmas. Forçar o heterodímero obrigatório impede a formação do homodímero. Isto aumentou muito a especificidade de ZFNs, bem como qualquer outra nuclease que adota essas variantes FokI.

[00483] Uma variedade de sistemas baseados em ZFN foi descrita na técnica, suas modificações são relatadas regularmente e inúmeras referências descrevem regras e parâmetros que são usados para orientar o desenho de ZFNs; ver, por exemplo, Segal *et al.*, *Proc Natl Acad Sci EUA 96*(6):2758-63 (1999); Dreier B *et al.*, *J Mol Biol. 303*(4):489-502 (2000); Liu Q *et al.*, *J Biol Chem. 277*(6):3850-6 (2002); Dreier *et al.*, *J Biol Chem 280*(42):35588-97 (2005); e Dreier *et al.*, *J Biol Chem. 276*(31):29466-78 (2001).

Nucleases Efetoras do Tipo Ativador de Transcrição (TALENs)

[00484] As TALENs representam outro formato de nucleases modulares em que, assim como os ZFNs, um domínio de ligação de DNA manipulado é ligado ao domínio de nuclease FokI e um par de TALENs opera em conjunto para alcançar a clivagem de DNA direcionada. A principal diferença em relação aos ZFNs é a natureza do domínio de ligação ao DNA e as propriedades de reconhecimento de sequência de DNA alvo associadas. O domínio TALEN de ligação ao DNA deriva das proteínas TALE, que foram originalmente descritas no patógeno bacteriano de plantas *Xanthomonas sp*. As TALEs têm séries em tandem de 33-35 repetições de aminoácidos, com cada repetição reconhecendo um único par de bases na sequência de DNA alvo que normalmente tem até 20 pb de comprimento, fornecendo um comprimento total de sequência alvo de até 40 pb. A especificidade de nucleotídeos de cada repetição é determinada pela variável de repetição diresíduo (RVD), que inclui apenas dois aminoácidos nas posições 12 e 13. As bases guanina, adenina, citosina e timina são predominantemente reconhecidas pelos quatro RVDs: Asn-Asn, Asn-Ile, His-Asp e Asn-Gly, respectivamente. Isto constitui um código de reconhecimento muito mais simples do que para os dedos de zinco e, portanto, representa uma vantagem sobre o último para o desenho da nuclease. No entanto, assim como os ZFNs, as interações proteína-DNA das TALENs não são absolutas na sua especificidade, e as TALENs também beneficiam do uso de variantes heterodiméricas obrigatórias do domínio FokI para reduzir a atividade fora do alvo.

[00485] Foram criadas variantes adicionais do domínio FokI que são desativadas em sua função catalítica. Se metade de um par TALEN ou ZFN conter um domínio FokI inativo, apenas a clivagem de DNA de fita simples (corte, "nicking") irá ocorrer no local alvo, em vez de uma DSB. O resultado é comparável ao uso de mutantes "nickase" CRISPR/Cas9/Cpf1 nos quais um dos domínios de clivagem Cas9 foi

desativado. Os cortes de DNA podem ser usados para conduzir a edição do genoma por HDR, mas com menor eficiência do que com uma DSB. O principal benefício é que os cortes fora do alvo são reparados com rapidez e precisão, ao contrário da DSB, que é propensa a erros de reparação mediados por NHEJ.

[00486] Uma variedade de sistemas baseados em TALEN foi descrita na técnica e suas modificações são relatadas regularmente; ver, por exemplo, Boch, *Science 326*(5959):1509-12 (2009); Mak *et al.*, *Science 335*(6069):716-9 (2012); e Moscou *et al.*, *Science 326*(5959):1501 (2009). O uso de TALENs com base na plataforma "Golden Gate", ou esquema de clonagem, foi descrito por vários grupos; ver, por exemplo, Cermak *et al.*, *Nucleic Acids Res. 39*(12):e82 (2011); Li *et al.*, *Nucleic Acids Res. 39*(14):6315-25(2011); Weber *et al.*, *PLoS One. 6*(2):e16765 (2011); Wang *et al.*, *J Genet Genomics 41*(6):339-47, Epub 2014 Can 17 (2014); e Cermak T *et al.*, *Methods Mol Biol. 1239*:133-59 (2015).

Endonucleases de Endereçamento

[00487] As endonucleases de endereçamento (HEs) são endonucleases específicas de sequência que possuem longas sequências de reconhecimento (14-44 pares de bases) e clivam o DNA com alta especificidade – geralmente em locais únicos no genoma. Existem pelo menos seis famílias conhecidas de HEs classificadas pela sua estrutura, incluindo LAGLIDADG (SEQ ID NO:6), GIY-YIG, caixa His-Cis, H-N-H, PD-(D/E)xK, e do tipo Vsr que são derivadas de uma ampla gama de hospedeiros, incluindo eucariontes, protistas, bactérias, arqueobactérias, cianobactérias e fagos. Como nos ZFNs e TALENs, as HEs podem ser usadas para criar uma DSB em um local alvo como a etapa inicial na edição do genoma. Além disso, algumas HEs naturais e manipuladas cortam apenas uma única fita de DNA, funcionando assim como nickases específicas do local. A grande sequência alvo das HEs

e a especificidade que elas oferecem as tornam candidatos atraentes para criar DSBs específicas do local.

[00488] Uma variedade de sistemas baseados em HE foi descrita na técnica e suas modificações são relatadas regularmente; ver, por exemplo, as revisões por Steentoft *et al.*, *Glycobiology 24*(8):663-80 (2014); Belfort e Bonocora, *Methods Mol Biol. 1123*:1-26 (2014); Hafez e Hausner, *Genome 55*(8):553-69 (2012); e referências neles citadas.

MegaTAL / Tev-mTALEN / MegaTev

[00489] Como exemplos adicionais de nucleases híbridas, a plataforma MegaTAL e a plataforma Tev-mTALEN usam uma fusão dos domínios de ligação ao DNA TALE e HEs cataliticamente ativos, aproveitando a ligação e a especificidade ajustáveis do DNA das TALE, bem como a especificidade da sequência de clivagem das HE; ver, por exemplo, Boissel *et al., NAR 42*: 2591-2601 (2014); Kleinstiver *et al.*, *G3 4*:1155-65 (2014); e Boissel e Scharenberg, *Methods Mol. Biol. 1239*: 171-96 (2015).

[00490] Em uma variação adicional, a arquitetura MegaTev é a fusão de uma meganuclease (Mega) com o domínio nuclease derivado da endonuclease I-TevI (Tev) de endereçamento GIY-YIG. Os dois locais ativos são posicionados a ~30 pb de distância em um substrato de DNA e geram duas DSBs com extremidades coesivas não compatíveis; ver, por exemplo, Wolfs *et al.*, *NAR 42*, 8816-29 (2014). É previsto que outras combinações de abordagens existentes baseadas em nuclease evoluam e sejam úteis para alcançar as modificações de genoma direcionadas no presente documento descritas.

dCas9-FokI ou dCpf1-FokI e Outras Nucleases

[00491] A combinação das propriedades estruturais e funcionais das plataformas de nuclease descritas acima oferece uma abordagem adicional à edição do genoma que pode potencialmente superar algumas das deficiências inerentes. Como exemplo, o sistema de

edição do genoma CRISPR normalmente usa uma única endonuclease Cas9 para criar uma DSB. A especificidade do direcionamento é dirigida por uma sequência de 20 ou 22 nucleotídeos no RNA guia que sofre o emparelhamento de bases de Watson-Crick com o DNA alvo (mais 2 bases adicionais na sequência adjacente de NAG ou NGG PAM no caso da Cas9 de *S. pyogenes*). Esta sequência é longa o suficiente para ser única no genoma humano, no entanto, a especificidade da interação RNA/DNA não é absoluta, com promiscuidade significativa às vezes tolerada, particularmente na metade 5' da sequência alvo, reduzindo efetivamente o número de bases que dirigem a especificidade. Uma solução para isto foi desativar completamente a função catalítica Cas9 ou Cpf1 – mantendo apenas a função de ligação ao DNA guiada por RNA – e, em vez disso, fundindo um domínio FokI à Cas9 desativada; ver, por exemplo, Tsai *et al., Nature Biotech 32:* 569-76 (2014); e Guilinger *et al.*, *Nature Biotech. 32:* 577-82 (2014). Como a FokI deve dimerizar para se tornar cataliticamente ativa, são necessários dois RNAs guia para fixar duas fusões FokI nas proximidades para formar o dímero e clivar o DNA. Isto essencialmente duplica o número de bases nos locais alvo combinados, aumentando assim o estringência do direcionamento por sistemas baseados em CRISPR.

[00492] Como exemplo adicional, a fusão do domínio de ligação ao DNA TALE com uma HE cataliticamente ativa, tal como I-TevI, aproveita a ligação e a especificidade ajustáveis do DNA das TALE, bem como a especificidade da sequência de clivagem de I-TevI, com a expectativa de que a clivagem fora do alvo possa ser reduzida ainda mais.

[00493] Os detalhes de uma ou mais modalidades da invenção são estabelecidos na descrição anexa abaixo. Embora qualquer uma das materiais e métodos similares ou equivalentes àqueles descritos no presente documento possam ser usados na prática ou teste da presente invenção, os materiais e métodos preferenciais são agora descritos.

Outras características, objetos e vantagens da invenção serão evidentes a partir da descrição. Na descrição, as formas singulares também incluem o plural, a menos que o contexto dite claramente de outro modo. A não ser que definidos de outro modo, todos os termos técnicos e científicos usados no presente documento têm o mesmo significado como comumente entendido por um perito na técnica à qual esta invenção pertence. Em caso de conflito, a presente descrição prevalecerá.

[00494] É entendido que os exemplos e modalidades descritos no presente documento são somente para propósitos ilustrativos e que várias modificações ou mudanças no mesmo contexto serão sugeridas a pessoas peritas na técnica e são para serem incluídas dentro do espírito e alcance deste pedido e âmbito das reivindicações anexas. Todas as publicações, patentes e pedidos de patentes citados no presente documento são deste modo incorporados por referência na sua totalidade para todos os propósitos.

[00495] Algumas modalidades das invenções fornecidas com este documento são adicionalmente ilustradas pelos seguintes exemplos não limitativos.

Modalidades Exemplificativas

[00496] Modalidade 1. Um sistema compreendendo:

uma endonuclease de ácido desoxirribonucleico (DNA) ou ácido nucleico que codifica a referida endonuclease de DNA;

RNA guia (gRNA) compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104; e

um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional.

[00497] Modalidade 2. O sistema da modalidade 1, em que o gRNA

compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

[00498] Modalidade 3. O sistema da modalidade 2, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

[00499] Modalidade 4. O sistema da modalidade 2, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

[00500] Modalidade 5. O sistema da modalidade 2, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

[00501] Modalidade 6. O sistema da modalidade 2, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[00502] Modalidade 7. O sistema de qualquer uma das modalidades 1-6, em que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1 ou um seu derivado funcional.

[00503] Modalidade 8. O sistema de qualquer uma das modalidades 1-7, em que a referida endonuclease de DNA é Cas9.

[00504] Modalidade 9. O sistema de qualquer uma das modalidades 1-8, em que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão em uma célula hospedeira.

[00505] Modalidade 10. O sistema de qualquer uma das modalidades 1-9, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão em uma célula hospedeira.

[00506] Modalidade 11. O sistema de qualquer uma das modalidades

1-10, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[00507] Modalidade 12. O sistema de qualquer uma das modalidades 1-10, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

[00508] Modalidade 13. O sistema da modalidade 12, em que o RNA que codifica a referida endonuclease de DNA é um mRNA.

[00509] Modalidade 14. O sistema de qualquer uma das modalidades 1-13, em que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[00510] Modalidade 15. O sistema da modalidade 14, em que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA.

[00511] Modalidade 16. O sistema da modalidade 15, em que o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

[00512] Modalidade 17. O sistema da modalidade 15 ou 16, em que o local alvo de gRNA é um local alvo para um gRNA no sistema.

[00513] Modalidade 18. O sistema da modalidade 17, em que o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA genômico para um gRNA no sistema.

[00514] Modalidade 19. O sistema de qualquer uma das modalidades 1-18, em que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[00515] Modalidade 20. O sistema da modalidade 19, em que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

[00516] Modalidade 21. O sistema de qualquer uma das modalidades 1-20, compreendendo a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[00517] Modalidade 22. Um método de edição de um genoma em uma célula, o método compreendendo fornecer o seguinte para a célula:

(a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104;

(b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e

(c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

[00518] Modalidade 23. O método da modalidade 22, em que o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

[00519] Modalidade 24. O método da modalidade 23, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

[00520] Modalidade 25. O método da modalidade 23, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

[00521] Modalidade 26. O método da modalidade 23, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

[00522] Modalidade 27. O método da modalidade 23, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[00523] Modalidade 28. O método de qualquer uma das modalidades 22-27, em que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX,

Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1 ou um seu derivado funcional.

[00524] Modalidade 29. O método de qualquer uma das modalidades 22-28, em que a referida endonuclease de DNA é Cas9.

[00525] Modalidade 30. O método de qualquer uma das modalidades 22-29, em que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão na célula.

[00526]

[00527] Modalidade 31. O método de qualquer uma das modalidades 22-30, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão na célula.

[00528] Modalidade 32. O método de qualquer uma das modalidades 22-31, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[00529] Modalidade 33. O método de qualquer uma das modalidades 22-31, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

[00530] Modalidade 34. O método da modalidade 33, em que o RNA que codifica a referida endonuclease de DNA é um mRNA.

[00531] Modalidade 35. O método de qualquer uma das modalidades 22-34, em que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[00532] Modalidade 36. O método de qualquer uma das modalidades 22-35, em que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA.

[00533] Modalidade 37. O método da modalidade 36, em que o

cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

[00534] Modalidade 38. O método da modalidade 36 ou 37, em que o local alvo de gRNA é um local alvo para o gRNA de (a).

[00535] Modalidade 39. O método da modalidade 38, em que o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA no genoma da célula para o gRNA de (a).

[00536] Modalidade 40. O método de qualquer uma das modalidades 22-39, em que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[00537] Modalidade 41. O método da modalidade 40, em que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

[00538] Modalidade 42. O método de qualquer uma das modalidades 22-41, compreendendo fornecer à célula a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[00539] Modalidade 43. O método de qualquer uma das modalidades 22-42, em que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula.

[00540] Modalidade 44. O método de qualquer uma das modalidades 22-43, em que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após (c) ser fornecido à célula.

[00541] Modalidade 45. O método da modalidade 43 ou 44, em que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a

endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b).

[00542] Modalidade 46. O método da modalidade 45, em que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

[00543] Modalidade 47. O método de qualquer uma das modalidades 22-46, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00544] Modalidade 48. O método de qualquer uma das modalidades 22-47, em que a referida célula é um hepatócito.

[00545] Modalidade 49. Uma célula geneticamente modificada na qual o genoma da célula é editado pelo método de qualquer uma das modalidades 22-48.

[00546] Modalidade 50. A célula geneticamente modificada da modalidade 49, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00547] Modalidade 51. A célula geneticamente modificada da modalidade 49 ou 50, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão na célula.

[00548] Modalidade 52. A célula geneticamente modificada de

qualquer uma das modalidades 49-51, em que a referida célula é um hepatócito.

[00549] Modalidade 53. Um método de tratamento da Hemofilia A em um indivíduo, o método compreendendo fornecer o seguinte a uma célula no indivíduo:

(a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104;

(b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e

(c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

[00550] Modalidade 54. O método da modalidade 53, em que o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

[00551] Modalidade 55. O método da modalidade 54, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

[00552] Modalidade 56. O método da modalidade 54, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

[00553] Modalidade 57. O método da modalidade 54, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

[00554] Modalidade 58. O método da modalidade 54, em que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

[00555] Modalidade 59. O método de qualquer uma das modalidades 53-58, em que o referido indivíduo é um paciente com ou é suspeito de ter Hemofilia A.

[00556] Modalidade 60. O método de qualquer uma das modalidades 53-58, em que o referido indivíduo é diagnosticado com um risco de Hemofilia A.

[00557] Modalidade 61. O método de qualquer uma das modalidades

53-60, em que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1 ou um seu derivado funcional.

[00558] Modalidade 62. O método de qualquer uma das modalidades 53-61, em que a referida endonuclease de DNA é Cas9.

[00559] Modalidade 63. O método de qualquer uma das modalidades 53-62, em que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão na célula.

[00560] Modalidade 64. O método de qualquer uma das modalidades 53-63, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão na célula.

[00561] Modalidade 65. O método de qualquer uma das modalidades 53-64, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

[00562] Modalidade 66. O método de qualquer uma das modalidades 53-64, em que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

[00563] Modalidade 67. O método da modalidade 66, em que o RNA que codifica a referida endonuclease de DNA é um mRNA.

[00564] Modalidade 68. O método de qualquer uma das modalidades 53-67, em que um ou mais do gRNA de (a), a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) e o molde doador de (c) são formulados em um lipossomo ou nanopartícula lipídica.

[00565] Modalidade 69. O método de qualquer uma das modalidades 53-68, em que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

[00566] Modalidade 70. O método de qualquer uma das modalidades 53-69, em que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA.

[00567] Modalidade 71. O método da modalidade 70, em que o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

[00568] Modalidade 72. O método da modalidade 70 ou 71, em que o local alvo de gRNA é um local alvo para o gRNA de (a).

[00569] Modalidade 73. O método da modalidade 72, em que o local alvo de gRNA do molde doador é o complemento reverso do local alvo de gRNA no genoma da célula para o gRNA de (a).

[00570] Modalidade 74. O método de qualquer uma das modalidades 53-73, em que fornecer o molde doador à célula compreende administrar o molde doador ao indivíduo.

[00571] Modalidade 75. O método da modalidade 74, em que a administração é por via intravenosa.

[00572] Modalidade 76. O método de qualquer uma das modalidades 53-75, em que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

[00573] Modalidade 77. O método da modalidade 76, em que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

[00574] Modalidade 78. O método da modalidade 77, em que

fornecer o gRNA e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA à célula compreende administrar o lipossomo ou nanopartícula lipídica ao indivíduo.

[00575] Modalidade 79. O método da modalidade 78, em que a administração é por via intravenosa.

[00576] Modalidade 80. O método de qualquer uma das modalidades 53-79, compreendendo proporcionar à célula a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

[00577] Modalidade 81. O método de qualquer uma das modalidades 53-80, em que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula.

[00578] Modalidade 82. O método de qualquer uma das modalidades 53-81, em que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após o molde doador de (c) ser fornecido à célula.

[00579] Modalidade 83. O método da modalidade 81 ou 82, em que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b).

[00580] Modalidade 84. O método da modalidade 83, em que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência

de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

[00581] Modalidade 85. O método de qualquer uma das modalidades 81-84, em que fornecer o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) à célula compreende administrar ao indivíduo uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA.

[00582] Modalidade 86. O método de qualquer uma das modalidades 81-85, em que fornecer o molde doador de (c) à célula compreende administrar ao indivíduo o molde doador codificado em um vetor AAV.

[00583] Modalidade 87. O método de qualquer uma das modalidades 53-86, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

[00584] Modalidade 88. O método de qualquer uma das modalidades 53-87, em que a referida célula é um hepatócito.

[00585] Modalidade 89. O método de qualquer uma das modalidades 53-88, em que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa no fígado do indivíduo.

[00586] Modalidade 90. Um método de tratamento da Hemofilia A em um indivíduo compreendendo:

[00587] administrar a célula geneticamente modificada de qualquer uma das modalidades 49-52 ao indivíduo.

[00588] Modalidade 91. O método da modalidade 90, em que a referida célula geneticamente modificada é autóloga ao indivíduo.

[00589] Modalidade 92. O método da modalidade 90 ou 91,

compreendendo adicionalmente:

[00590] obter uma amostra biológica do indivíduo em que a amostra biológica compreende uma célula de hepatócito, em que a célula geneticamente modificada é preparada a partir do hepatócito.

[00591] Modalidade 93. Um kit compreendendo um ou mais elementos do sistema de qualquer uma das modalidades 1-21 e compreendendo adicionalmente instruções de uso.

EXEMPLOS

EXEMPLO 1: Identificação de gRNAs que direcionam a clivagem pela nuclease Cas9 no íntron 1 do gene da albumina de camundongo em células Hepa1-6 *in vitro*

[00592] Para fins de avaliação em modelos de animais pré-clínicos relevantes, foram testadas moléculas de gRNA que direcionam a clivagem eficiente pela nuclease Cas9 no íntron 1 da albumina de espécies animais pré-clínicas relevantes. Os modelos de camundongos da Hemofilia A estão bem estabelecidos (Bi L, Lawler AM, Antonarakis SE, High KA, Gearhart JD, Kazazian HH., Jr "Targeted disruption of the mouse factor VIII gene produces a model of hemophilia A." (Nat Genet. 1995;10:119–21. doi: 10.1038/ng0595-119) e representam um sistema modelo valioso para testar novas abordagens terapêuticas para esta doença. Para identificar o gRNA com potencial para cortar o íntron 1 da albumina do camundongo, a sequência do íntron foi analisada usando algoritmos (por exemplo, CCTOP; https://crispr.cos.uni-heidelberg.de/) que identificam todas as possíveis sequências alvo de gRNA usando uma sequência NGG de PAM que seriam alvos potenciais para a clivagem pela Cas9 de Streptococcus pyogenes (spCas9) na sequência de interesse e todas as sequências relacionadas no genoma do camundongo. Cada gRNA foi então classificado com base na frequência de sequências exatas ou relacionadas no genoma do camundongo para identificar o gRNA com o menor risco teórico do corte fora do alvo. Com

base em uma análise deste tipo, um gRNA chamado mALbgRNA_T1 foi selecionado para teste.

[00593] O mAlbgRNA_T1 exibiu homologia para apenas 4 outros locais no genoma do camundongo, cada um dos quais exibindo 4 não correspondências de nucleotídeos, como mostrado na Tabela 2 abaixo.

Tabela 2. Potenciais locais fora do alvo para o gRNA de mAlb_T1 no genoma do camundongo (MM = número de não correspondências)

Cromos-somo	fita	MM	seq_alvo	Alinha-mento	posição	nome do gene
chr5	+	4	TGCCTTTACCCCATCGT TAC (SEQ ID NO: 7)	‖‖‖--‖-[‖‖-‖‖‖‖‖‖]PAM	Exônica	Tmprss 11g
chr11	-	4	TGCCTCCTCCCGATAG TTAC (SEQ ID NO: 8)	‖‖‖---‖[‖‖‖‖‖-‖‖‖‖]PAM	Intrôni-ca	Dhrs7c
chr1	+	4	GGACAGTTCCTGATTG TTAC (SEQ ID NO: 9)	-‖-‖‖‖‖[‖‖-‖‖‖-‖‖‖‖]PAM	Exônica	Gm376 00
chrX	+	4	TGCCTTTTCCCGATTGT TAA (SEQ ID NO: 10)	‖‖‖--‖‖[‖‖‖‖‖-‖‖‖‖-]PAM	Intergêni ca	Zfp280c

[00594] Para avaliar a eficiência do mALbgRNA_T1 para promover a clivagem por Cas9 em células de camundongo, foi usada a linha celular Hepa1-6 derivada de células de fígado de camundongo. As células Hepa1-6 foram cultivadas em DMEM + 10% de FBS em uma incubadora a 5% de CO_2. Um complexo ribonuclear-proteico (RNP) composto pelo gRNA ligado à proteína Cas9 de Streptococcus pyogenes (spCas9) foi pré-formado pela mistura de 2,4 µL de spCas9 (0,8 µg/µL) e 3 µL de gRNA sintético (20 µMolar) e 7 µL de PBS (razão 1:5 spCas9:gRNA) e incubação à temperatura ambiente durante 10 minutos. Para a nucleofecção, o frasco inteiro do reagente suplemento SF (Lonza) foi adicionado ao reagente Nucleofector SF (Lonza) para preparar o reagente de nucleofecção completo. Para cada nucleofecção $1x10^5$ de células Hepa1-6 foram ressuspensas em 20 µL do reagente de nucleofecção completo, adicionadas ao RNP e depois transferidas para

uma cubeta de nucleofecção (tira de 16 poços) que foi colocada no dispositivo de nucleofecção 4D (Lonza) e nucleofectadas usando o programa EH-100. Depois de deixar as células descansarem durante 10 minutos, elas foram transferidas para uma placa de tamanho apropriado com meio completo fresco. Após 48 horas da nucleofecção, as células foram coletadas e o DNA genômico foi extraído e purificado usando o kit Qiagen DNeasy (cat 69506).

[00595] Para avaliar a frequência do corte mediado por Cas9/gRNA no local alvo no íntron 1 da albumina, um par de iniciadores (MALBF3; 5' TTATTACGGTCTCATAGGGC 3' (SEQ ID NO: 11) e MALBR5: AGTCTTTCTGTCAATGCACAC 3' (SEQ ID NO: 12)) flanqueando o local alvo foram usados em uma reação em cadeia da polimerase (PCR) usando uma temperatura de hibridação de 52 °C para amplificar uma região de 609 pb do DNA genômico. O produto de PCR foi purificado usando o Kit de Purificação Qiagen PCR (Cat no. 28106) e sequenciado diretamente usando o sequenciamento de Sanger com os mesmos iniciadores usados para a reação de PCR. Os dados da sequência foram analisados por um algoritmo chamado "Tracking of Indels by Decomposition" (TIDES), que determinou a frequência de inserções e deleções (INDELS) presentes no local de corte previsto para o complexo gRNA/Cas9 (Brinkman et al (2104); Nucleic Acids Research, 2014, 1). A frequência geral da geração de INDEL para o mAlbgRNA_T1 estava entre 85 e 95% quando testada em 3 experiências independentes, indicando um corte eficiente pelo gRNA/Cas9 no genoma dessas células. Um exemplo da análise TIDES em células Hepa1-6 nucleofectadas com o gRNA-T1 de mAlb é mostrado na FIG. 3. A maioria das inserções e deleções consiste em inserções de 1 pb e deleções de 1 pb com números menores de deleções de até 6 pb.

EXEMPLO 2: Avaliação da eficiência de clivagem do mAlbgRNA_T1 *in vivo* em camundongos

[00596] Para entregar Cas9 e mAlbgRNA-T1 aos hepatócitos de camundongos, foi usado um veículo de entrega de nanopartículas lipídicas (LNP). O sgRNA foi quimicamente sintetizado incorporando nucleotídeos quimicamente modificados para melhorar a resistência às nucleases. O gRNA em um exemplo é composto da seguinte estrutura: 5' usgscsCAGUUCCCGAUCGUUACGUUUUAGAgcuaGAAAuagcAAGU UAAAAUAAGGCUAGUCCGUUAUCaacuuGAAAaaguggcaccgagucgg ugcusususU-3' (SEQ ID NO: 13), onde "A, G, U, C" são nucleotídeos de RNA nativos, "a, g, u, c" são nucleotídeos 2'-O-metila e "s" representa uma estrutura principal do fosforotioato. A sequência de direcionamento da albumina do camundongo do gRNA está sublinhada, o restante da sequência de gRNA é a sequência de suporte comum. O mRNA de spCas9 foi projetado para codificar a proteína spCas9 fundida com um domínio de localização nuclear (NLS) que é necessário para transportar a proteína spCas9 para o compartimento nuclear, onde pode ocorrer a clivagem do DNA genômico. Componentes adicionais do mRNA de Cas9 são uma sequência KOZAK na extremidade 5' antes do primeiro códon para promover a ligação ao ribossomo e uma cauda de poliA na extremidade 3' composta por uma série de resíduos A. Um exemplo da sequência de um mRNA de spCas9 com sequências NLS é mostrado na SEQ ID NO: 81. O mRNA pode ser produzido por diferentes métodos bem conhecidos na técnica. Um desses métodos no presente documento usados é a transcrição *in vitro* usando a polimerase T7, na qual a sequência do mRNA é codificada em um plasmídeo que contém um promotor da polimerase T7. Resumidamente, após incubação do plasmídeo em um tampão apropriado contendo polimerase T7 e ribonucleotídeos, foi produzida uma molécula de RNA que codifica a sequência de aminoácidos da proteína desejada. Os ribonucleotídeos naturais ou ribonucleotídeos quimicamente modificados na mistura de

reação foram usados para gerar moléculas de mRNA com estrutura química natural ou com estruturas químicas modificadas que podem ter vantagens em termos de expressão, estabilidade ou imunogenicidade. Além disso, a sequência da sequência de codificação de spCas9 foi otimizada para o uso de códons por usação do códon mais frequentemente usado para cada aminoácido. Além disso, a sequência de codificação foi otimizada para remover os locais de ligação ao ribossomo críptico e os quadros de leitura abertos a montante, a fim de promover a tradução mais eficiente do mRNA na proteína spCas9.

[00597] Um componente primário da LNP usado nesses estudos é o lipídio C12-200 (Love et al (2010), PNAS vol. 107, 1864–1869). O lipídio C12-200 forma um complexo com as moléculas de RNA altamente carregadas. O C12-200 foi combinado com 1,2-Dioleoil-sn-glicero-3-fosfoetanolamina (DOPE), DMPE-mPEG2000 e colesterol. Quando misturados sob condições controladas, por exemplo, em um dispositivo NanoAssemblr (Precision NanoSystems) com ácidos nucléicos como gRNA e mRNA, ocorreu uma automontagem da LNP na qual o ácido nucleico foi encapsulado dentro da LNP. Para montar o gRNA e mRNA de Cas9 na LNP, foram pipetados estoques de etanol e lipídios em frascos de vidro, como apropriado. A proporção de C12-200 para DOPE, DMPE-mPEG2000 e colesterol foi ajustada para otimizar a formulação. Uma razão típica foi composta por C12-200, DOPE, colesterol e mPEG2000-DMG a uma razão molar de 50:10:38,5:1,5. O gRNA e o mRNA foram diluídos em citrato de Na a 100 mM, pH 3,0 e NaCl a 300 mM, em tubos livres de RNase. O cartucho NanoAssemblr (Precision NanoSystems) foi lavado com etanol no lado do lipídio e com água no lado do RNA. O estoque de lipídios foi puxado para uma seringa, o ar removido da seringa e inserido no cartucho. O mesmo procedimento foi usado para carregar uma seringa com a mistura de gRNA e mRNA de Cas9. A operação Nanoassemblr foi então realizada sob condições

padrão. A suspensão de LNP foi então dialisada usando cartuchos de diálise de limite de exclusão de 20 Kd em 4 litros de PBS durante 4 h e depois concentrada usando centrifugação através de cartuchos de centrifugação de limite de exclusão de 20 Kd (Amicon) incluindo lavagem três vezes em PBS durante a centrifugação. Finalmente, a suspensão de LNP foi filtrada estéril através do filtro de seringa de 0,2 μM. Os níveis de endotoxina foram verificados usando o kit comercial de endotoxina (ensaio LAL) e a distribuição do tamanho de partícula foi determinada por dispersão de luz dinâmica. A concentração de RNA encapsulado foi determinada usando um ensaio de Ribogreen (Thermo Fisher). Em alternativa, o gRNA e o mRNA de Cas9 foram formulados separadamente em LNP e depois misturados antes do tratamento de células em cultura ou injeção em animais. O uso de gRNA e mRNA de Cas9 formulados separadamente permitiu a análise de razões precisas de gRNA e mRNA de Cas9.

[00598] Formulações alternativas de LNP que usavam moléculas lipídicas catiônicas alternativas também foram usadas para entrega *in vivo* do gRNA e mRNA de Cas9. A LNP preparada de fresco encapsulando o gRNA de mALB T1 e mRNA de Cas9, foi misturada a uma razão de massa de 1:1 do RNA e injetada na veia da cauda (injeção TV) de camundongos com Hemofilia A. Em alternativa, a LNP foi doseada por injeção retro-orbital (RO). A dose de LNP administrada aos camundongos variou de 0,5 a 2 mg de RNA por kg de peso corporal. Três dias após a injeção da LNP os camundongos foram sacrificados e uma parte dos lobos esquerdo e direito do fígado e uma parte do baço foram coletados e o DNA genômico foi purificado de cada um. O DNA genômico foi então submetido à análise TIDES para medir a frequência de corte e o perfil de clivagem no local alvo no íntron 1 da albumina. Um exemplo dos resultados é mostrado na FIG. 4, em que, em média, 25% dos alelos foram clivados na dose de 2 mg/kg. Foi observada uma

resposta à dose com dose de 0,5 mg/kg, resultando em cerca de 5% de corte e 1 mg/kg, resultando em cerca de 10% de corte. Os camundongos injetados apenas com tampão PBS mostraram um sinal baixo de cerca de 1 a 2% no ensaio TIDES, que é uma medida de fundo do próprio ensaio TIDES.

Exemplo 3: AVALIANDO FREQUÊNCIAS INDEL DE SGRNAS DIRECIONADOS AO ÍNTRON 1 DA ALBUMINA HUMANA

[00599] Todas as possíveis sequências de gRNA usando uma sequência NGG de PAM que seriam alvos de clivagem pela Cas9 de *Streptococcus pyogenes* (spCas9) dentro do íntron 1 do gene da albumina humana foram identificadas usando um algoritmo proprietário chamado "Guido", baseado no algoritmo publicado chamado "CCTop" (ver, por exemplo, https://crispr.cos.uni-heidelberg.de/). Esse algoritmo identifica possíveis locais fora do alvo no genoma humano e classifica cada gRNA com base no potencial previsto de corte fora do alvo. As sequências de gRNA identificadas são fornecidas na tabela abaixo.

Tabela 3. Sequências de gRNA do íntron 1 da albumina humana

Nome gRNA	Sequência gRNA (com PAM)
Albumina Humana Íntron-1_T1	TAATTTTCTTTTGCGCACTAAGG (SEQ ID NO: 18)
Albumina Humana Íntron-1_T2	TAGTGCAATGGATAGGTCTTTGG (SEQ ID NO: 19)
Albumina Humana Íntron-1_T3	AGTGCAATGGATAGGTCTTTGGG (SEQ ID NO: 20)
Albumina Humana Íntron-1_T4	TAAAGCATAGTGCAATGGATAGG (SEQ ID NO: 21)
Albumina Humana Íntron-1_T5	ATTTATGAGATCAACAGCACAGG (SEQ ID NO: 22)
Albumina Humana Íntron-1_T6	TGATTCCTACAGAAAAACTCAGG (SEQ ID NO: 23)
Albumina Humana Íntron-1_T7	TGTATTTGTGAAGTCTTACAAGG (SEQ ID NO: 24)
Albumina Humana Íntron-1_T8	GACTGAAACTTCACAGAATAGGG (SEQ ID NO: 25)
Albumina Humana Íntron-1_T9	AATGCATAATCTAAGTCAAATGG (SEQ ID NO: 26)

Albumina Humana Íntron-1_T10	TGACTGAAACTTCACAGAATAGG (SEQ ID NO: 27)
Albumina Humana Íntron-1_T11	TTAAATAAAGCATAGTGCAATGG (SEQ ID NO: 28)
Albumina Humana Íntron-1_T12	GATCAACAGCACAGGTTTTGTGG (SEQ ID NO: 29)
Albumina Humana Íntron-1_T13	TAATAAAATTCAAACATCCTAGG (SEQ ID NO: 30)
Albumina Humana Íntron-1_T14	TTCATTTTAGTCTGTCTTCTTGG (SEQ ID NO: 31)
Albumina Humana Íntron-1_T15	ATTATCTAAGTTTGAATATAAGG (SEQ ID NO: 32)
Albumina Humana Íntron-1_T16	ATCATCCTGAGTTTTTCTGTAGG (SEQ ID NO: 33)
Albumina Humana Íntron-1_T17	GCATCTTTAAAGAATTATTTTGG (SEQ ID NO: 34)
Albumina Humana Íntron-1_T18	TACTAAAACTTTATTTTACTGGG (SEQ ID NO: 35)
Albumina Humana Íntron-1_T19	TGAATTATTCTTCTGTTTAAAGG (SEQ ID NO: 36)
Albumina Humana Íntron-1_T20	AATTTTTAAAATAGTATTCTTGG (SEQ ID NO: 37)
Albumina Humana Íntron-1_T21	ATGCATTTGTTTCAAAATATTGG (SEQ ID NO: 38)
Albumina Humana Íntron-1_T22	TTTGGCATTTATTTCTAAAATGG (SEQ ID NO: 39)
Albumina Humana Íntron-1_T23	AAAGTTGAACAATAGAAAAATGG (SEQ ID NO: 40)
Albumina Humana Íntron-1_T24	TTACTAAAACTTTATTTTACTGG (SEQ ID NO: 41)
Albumina Humana Íntron-1_T25	ACCTTTTTTTTTTTTTACCTAGG (SEQ ID NO: 104)
Albumina Humana Íntron-1_T26	TGCATTTGTTTCAAAATATTGGG (SEQ ID NO: 42)
Albumina Humana Íntron-1_T27	TGGGCAAGGGAAGAAAAAAAAGG (SEQ ID NO: 43)
Albumina Humana Íntron-1_T28	TCCTAGGTAAAAAAAAAAAAGG (SEQ ID NO: 44)

[00600] A proteína Cas9 nuclease (Platinum™, GeneArt™) a 5 µg/µL foi adquirida da Thermo Fisher Scientific (número de catálogo A27865, Carlsbad, CA), depois diluída 1:6 até uma concentração de trabalho de 0,83 µg/µL ou 5,2 µM. O RNA guia único sintético (sgRNA) quimicamente modificado (Synthego Corp, Menlo Park, CA) foi

ressuspenso a 100 µM com tampão TE como uma solução estoque. Em alternativa, o gRNA usado pode ser produzido por transcrição *in vitro* (IVT). Esta solução foi diluída com água livre de nuclease até uma concentração de trabalho de 20 µM.

[00601] Para produzir complexos de ribonucleoproteínas, a proteína Cas9 (12,5 pmol) e o sgRNA (60 pmol) foram incubados durante 10-20 minutos à temperatura ambiente. Durante essa incubação, as células HepG2 (American Type Culture Collection, Manassas, Virgínia) ou células HuH7 (American Type Tissue Culture Collection, Manassas, Virgínia) foram dissociadas usando Tripsina-EDTA a 0,25% (Thermo Fisher Scientific) durante 5 minutos a 37 °C. Cada reação de transfecção continha 1 x 10^5 células, e o número apropriado de células por experiência foi centrifugado a 350xG durante 3 minutos e depois ressuspenso em 20 µL solução de nucleofecção adicionalmente suplementada da Lonza SF mais (número de catálogo V4XC-2032, Basel, Suíça) por reação de transfecção. As células ressuspensas em 20 µL de solução de nucleofecção foram adicionadas a cada tubo de RNP e todo o volume foi transferido para um poço de uma tira de nucleofecção de 16 poços. As células HepG2 ou HuH7 foram transfectadas usando o programa EH-100 no Sistema Amaxa 4D-Nucleofector (Lonza). As HepG2 e HuH7 são linhas celulares de hepatócitos humanos que são assim relevantes para a avaliação do gRNA que é usado para clivar um gene no fígado. Após a transfecção, as células foram incubadas na tira de nucleofecção durante 10 minutos, transferidas para uma placa de 48 poços contendo meio quente, consistindo no Meio Essencial Mínimo de Eagle (número de catálogo 10-009-CV, Corning, Corning, NI) suplementado com 10% de soro bovino fetal (número de catálogo 10438026, Thermo Fisher Scientific). As células foram novamente alimentadas com meio fresco no dia seguinte.

[00602] Após 48 horas da transfecção, as células HepG2 ou HuH7 foram dissociadas e o DNA genômico foi extraído usando o kit Qiagen DNeasy (número de catálogo 69506, Hilden, Alemanha). A PCR foi realizada usando DNA genômico extraído, com a Platinum SuperFi Green PCR Master Mix (Thermo Fisher Scientific) e os seguintes iniciadores a 0,2 μM: Albumina iniciador direto: 5'-CCCTCCGTTTGTCCTAGCTT-3' (SEQ ID NO: 14); Albumina iniciador reverso: 5'-TCTACGAGGCAGCACTGTT-3' (SEQ ID NO: 15); AAVS1 iniciador direto:: 5'-AACTGCTTCTCCTCTTGGGAAGT-3' (SEQ ID NO: 16); AAVS1 iniciador reverso: 5'-CCTCTCCATCCTCTTGCTTTCTTTG-3'(SEQ ID NO: 17). As condições de PCR foram 2 minutos a 98 °C (1X), seguido de 30 segundos a 98 °C, 30 s a 62,5 °C e 1 min a 72 °C (35x). O produto de PCR correto foi confirmado usando um E-Gel a 1,2% (Thermo Fisher Scientific) e purificado usando o kit de purificação por Qiagen PCR (número de catálogo 28106). Os produtos de PCR purificados foram submetidos a sequenciamento de Sanger usando o iniciador direto ou reverso para o produto de PCR correspondente. As frequências de inserções ou deleções no local de clivagem previsto para o gRNA/Cas9 foram determinadas usando o algoritmo de análise TIDE como descrito por Brinkman, et al. (Brinkman, E.K., Chen, T., Amendola, M, e van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research, 2014, Vol. 42, No. 22* ***e168****)*. Resumidamente, os ficheiros de sequenciamento do cromatograma foram comparados com um cromatograma de controle derivado de células não tratadas para determinar a abundância relativa de nucleotídeos anômalos. Os resultados estão resumidos na Tabela 4. Também é de interesse identificar sequências de gRNA no ser humano que são homólogas em espécies pré-clínicas relevantes, como primatas não humanos. O alinhamento das sequências de gRNA potenciais identificadas no íntron 1 da albumina humana com as sequências do

íntron 1 da albumina dos primatas *Macaca fascicularis* e *Macaca mulatta* identificou várias moléculas de gRNA com correspondências perfeitas ou não correspondência de 1 a 2 nucleotídeos, como mostrado na Tabela 4. As frequências INDEL geradas usando guias IVT foram medidas em células HuH7, e as frequências INDEL geradas com guias sintéticos foram medidas em células HepG2. As frequências INDEL geradas pelos diferentes guias nas células HuH7 variaram de 0,3% a 64%, demonstrando que um gRNA que cliva com eficiência no íntron 1 da albumina não pode ser selecionado puramente com base em uma sequência baseada no algoritmo *in silico*. Com base nas frequências INDEL do IVT gRNA nas células HuH7 e do gRNA sintético nas células HepG2, foram identificados vários gRNAs com frequências de clivagem superiores a 40%. De particular interesse são os gRNA T5 e T12 que exibiram 46% e 43% de corte como guias sintéticos e são 100% idênticos em humanos e primatas.

Tabela 4. Eficiências de clivagem de candidatos sgRNA no íntron 1 da albumina humana e sua homologia com primatas. sgRNA = gRNA sintético, IVT gRNA = gRNA produzido por transcrição in vitro. *Alinhamento da sequência para *Macaca fascicularis* e *Macaca mulatta* com até 2 não correspondências em negrito e sublinhadas. Dados INDEL para o IVR gRNA N=1-2; sgRNA sintético, N=2-3

sgRNA	Sequência	Fita	Frequência Indel - IVT gRNA (%±SEM)	Frequência Indel - gRNA sintético (%±SEM)	*Macaca fascicularis**	*Macaca mulatta**
T1	TAATTTTCTTTTGC GCACTAAGG (SEQ ID NO: 18)	+	6,8 ± 1,7		TAATTTTCTTTTG C**C**CACTAAGG (SEQ ID NO: 45)	TAATTTTCTTTTGC **C**CACTAAGG (SEQ ID NO: 62)
T2	TAGTGCAATGGAT AGGTCTTTGG (SEQ ID NO: 19)	-	9,9 ± 2,2		TAGTGCAATGGA TAGGTCTT**A**GG (SEQ ID NO: 46)	TAGTGCAATGGAT AGGTCTT**A**GG (SEQ ID NO: 63)
T3	AGTGCAATGGATA GGTCTTTGGG (SEQ ID NO: 20)	-	28,6 ± 20,5	52,3 ± 5,4	AGTGCAATGGAT AGGTCTT**A**GGG (SEQ ID NO: 47)	AGTGCAATGGATA GGTCTT**A**GGG (SEQ ID NO: 64)
T4	TAAAGCATAGTGC AATGGATAGG (SEQ ID NO: 21)	-	26,3 ± 10,8	38,1 ± 11,9	TAAAGCATAGTG CAATGGATAGG (SEQ ID NO: 48)	TAAAGCATAGTGC AATGGATAGG (SEQ ID NO: 65)
T5	ATTTATGAGATCA ACAGCACAGG (SEQ ID NO: 22)	-	64,4 ± 6,4	46,1 ± 3,4	ATTTATGAGATC AACAGCACAGG (SEQ ID NO: 49)	ATTTATGAGATCA ACAGCACAGG (SEQ ID NO: 66)
T6	TGATTCCTACAGA	-	43,1 ± 13,3	37,7 ± 6,8	TGATTCCTACAG	TGATTCCTACAGA

	AAAACTCAGG (SEQ ID NO: 23)				AAAAA**G**TCAGG (SEQ ID NO: 50)	AAAA**G**TCAGG (SEQ ID NO: 67)
T7	TGTATTTGTGAAG TCTTACAAGG (SEQ ID NO: 24)	+	28,5 ± 17,2	38,2 ± 0,55		
T8	GACTGAAACTTCA CAGAATAGGG (SEQ ID NO: 25)	+	45,7 ± 24,6	38,5 ± 3,7		
T9	AATGCATAATCTA AGTCAAATGG (SEQ ID NO: 26)	+	5,6 ± 0,1		AATGCATAATCT AAGTCAAATGG (SEQ ID NO: 51)	AATGCATAATCTA AGTCAAATGG (SEQ ID NO: 68)
T10	TGACTGAAACTTC ACAGAATAGG (SEQ ID NO: 27)	+	24,4 ± 14,0			
T11	TTAAATAAAGCATA GTGCAATGG (SEQ ID NO: 28)	-	20,3 ± 9,4	61,3	TTAAATAAAGCA TAGTGCAATGG (SEQ ID NO: 52)	TTAAATAAAGCATA GTGCAATGG (SEQ ID NO: 69)
T12	GATCAACAGCACA GGTTTTGTGG (SEQ ID NO: 29)	-	60,0 ± 0,6	43,5 ± 5,9	ATTTATGAGATC AACAGCACAGG (SEQ ID NO: 53)	ATTTATGAGATCA ACAGCACAGG (SEQ ID NO: 70)
T13	TAATAAAATTCAAA CATCCTAGG (SEQ ID NO: 30)	+	13,9 ± 9,3	38,5 ± 2,7	TAATAAAATTCA AACATCCTAGG (SEQ ID NO: 54)	TAATAAAATTCAAA CATCCTAGG (SEQ ID NO: 71)
T14	TTCATTTTAGTCTG TCTTCTTGG (SEQ ID NO: 31)	+	1,1 ± 0,3			
T15	ATTATCTAAGTTTG AATATAAGG (SEQ ID NO: 32)	+	14,6 ± 3,9			
T16	ATCATCCTGAGTT TTTCTGTAGG (SEQ ID NO: 33)	+	14,7 ± 6,4		AT**T**ATCCTGA**C**T TTTTCTGTAGG (SEQ ID NO: 55)	AT**T**ATCCTGA**C**TTT TTCTGTAGG (SEQ ID NO: 72)
T17	GCATCTTTAAAGA ATTATTTTGG (SEQ ID NO: 34)	+	39,2 ± 25,9	22,5 ± 2,7		
T18	TACTAAAACTTTAT TTTACTGGG (SEQ ID NO: 35)	-	2,1 ± 0,2		TACTAAAACTTT ATTTTACT**T**GG (SEQ ID NO: 56)	TACTAAAACTTTAT TTTACT**T**GG (SEQ ID NO: 73)
T19	TGAATTATTCTTCT GTTTAAAGG (SEQ ID NO: 36)	+	2,4 ± 1,1		TGAATTATTC**C**T CTGTTTAAAGG (SEQ ID NO: 57)	TGAATTATTC**C**TCT GTTTAAAGG (SEQ ID NO: 74)
T20	AATTTTTAAAATAG TATTCTTGG (SEQ ID NO: 37)	+	0,3			
T21	ATGCATTTGTTTCA AAATATTGG (SEQ ID NO: 38)	-	2,2 ± 0,1		ATGCATTTGTTT CAAAATATTGG (SEQ ID NO: 58)	ATGCATTTGTTTCA AAATATTGG (SEQ ID NO: 75)
T22	TTTGGCATTTATTT CTAAAATGG (SEQ ID NO: 39)	+	1,7 ± 0,6		TTTGGCATTTAT TTCTAAAATGG (SEQ ID NO: 59)	TTTGGCATTTATTT CTAAAATGG (SEQ ID NO: 76)
T23	AAAGTTGAACAAT AGAAAAATGG (SEQ ID NO: 40)	-	4,7 ± 0,3		AAAGTTGAACAA TAGAAAAATGG (SEQ ID NO: 60)	AAAGTTGAACAAT AGAAAAATGG (SEQ ID NO: 77)
T24	TTACTAAAACTTTA TTTTACTGG (SEQ ID NO: 41)	-	1,4 ± 0,7			
T26	TGCATTTGTTTCAA AATATTGGG (SEQ ID NO: 42)	-	3,2 ± 0,0		TGCATTTGTTTC AAAATATTGGG (SEQ ID NO: 61)	TGCATTTGTTTCAA AATATTGGG (SEQ ID NO: 78)
T27	TGGGCAAGGGAA GAAAAAAAGG (SEQ ID NO: 43)	-	32,0 ± 4,9	48,4 ± 13,3		TGGG**G**AAGGG**G**A GAAAAAAAGG (SEQ ID NO: 79)

Exemplo 4: INTEGRAÇÃO DIRECIONADA DE UM GENE

TERAPÊUTICO DE INTERESSE NO ÍNTRON 1 DA ALBUMINA DE CAMUNDONGO

[00603] Uma abordagem para expressar uma proteína terapêutica necessária para tratar uma doença é a integração direcionada do cDNA ou a sequência de codificação do gene que codifica essa proteína no locus da albumina no fígado *in vivo*. A integração direcionada é um processo pelo qual um molde de DNA doador é integrado no genoma de um organismo no local de uma quebra de fita dupla, ocorrendo essa integração por HDR ou NHEJ. Esta abordagem usa a introdução nas células do organismo de uma nuclease de DNA específica de sequência e de um molde de DNA doador que codifica o gene terapêutico. Avaliamos se uma nuclease CRISPR-Cas9 direcionada para o íntron 1 de albumina era capaz de promover a integração direcionada de um molde de DNA doador. O molde de DNA doador é entregue em um vírus AAV, preferencialmente um vírus AAV8 no caso de camundongos, que preferencialmente transduz os hepatócitos do fígado após injeção intravenosa. O gRNA de mAlb_T1 específico da sequência e o mRNA de Cas9 são entregues aos hepatócitos do fígado dos mesmos camundongos por injeção intravenosa ou RO de uma formulação da LNP que encapsula o gRNA e mRNA de Cas9. Em um caso, o AAV8 com o molde doador é injetado nos camundongos antes da LNP, pois é sabido que a transdução dos hepatócitos pelo AAV leva várias horas a dias e o DNA doador entregue é mantido de forma estável nos núcleos dos hepatócitos durante semanas a meses. Por outro lado, o gRNA e o mRNA entregues por uma LNP persistirão nos hepatócitos por apenas 1 a 4 dias devido à instabilidade inerente às moléculas de RNA. Em outro caso, a LNP é injetada nos camundongos entre 1 dia e 7 dias após o AAV com o molde doador. O molde de DNA doador incorpora várias características projetadas com o objetivo de (i) maximizar a integração e (ii) maximizar a expressão da proteína terapêutica codificada.

[00604] Para que a integração ocorra através de HDR, os braços de homologia precisam ser incluídos em ambos os lados do cassete do gene terapêutico. Estes braços de homologia são compostos pelas sequências de ambos os lados do local de corte do gRNA no íntron 1 da albumina de camundongo. Enquanto braços de homologia mais longos geralmente promovem HDR mais eficiente, o comprimento dos braços de homologia pode ser limitado pelo limite de empacotamento para o vírus AAV de cerca de 4,7 a 5,0 Kb. Assim, a identificação do comprimento ideal do braço de homologia requer testes. A integração também pode ocorrer através de mecanismos NHEJ, nos quais as extremidades livres de um DNA doador de fita dupla são unidas às extremidades de uma quebra de fita dupla. Neste caso, os braços de homologia não são necessários. No entanto, a incorporação de locais de corte de gRNA em ambos os lados do cassete de genes pode melhorar a eficiência da integração ao gerar fragmentos lineares de fita dupla. Usando locais de clivagem de gRNA na orientação reversa, integração na orientação direta desejada pode ser favorecida. A introdução de uma mutação no local de clivagem da furina do FVIII pode gerar uma proteína FVIII que não pode ser clivada pela furina durante a expressão da proteína, resultando em um polipeptídeo de FVIII de uma cadeia que demonstrou ter estabilidade melhorada no plasma, enquanto mantendo a funcionalidade total.

[00605] DNA doadores exemplificativos projetados para integrar um gene FVIII no íntron 1 de albumina são mostrados na FIG. 5. Sequências projectadas de doadores específicos estão em sequência das SEQ ID NOs: 87-92.

[00606] A produção de AAV8 ou outro vírus do sorotipo AAV empacotado com o DNA doador FVIII é realizada usando métodos de empacotamento viral bem estabelecidos. Em um desses métodos, as células HEK293 são transfectadas com 3 plasmídeos, um que codifica

as proteínas de empacotamento de AAV, o segundo que codifica proteínas auxiliares de Adenovírus e o 3° que contém a sequência de DNA doador de FVIII flanqueada por sequências ITR de AAV. As células transfectadas dão origem a partículas de AAV do sorotipo especificado pela composição das proteínas do capsídeo do AAV codificadas no primeiro plasmídeo. Essas partículas de AAV são coletadas do sobrenadante celular ou do sobrenadante e das células lisadas e purificadas sobre um gradiente de CsCl ou um gradiente de Iodixanol ou por outros métodos, como desejado. As partículas virais purificadas são quantificadas medindo o número de cópias do genoma do DNA doador por PCR quantitativo (Q-PCR).

[00607] A entrega *in vivo* do gRNA e do mRNA de Cas9 é realizada por vários métodos. No primeiro caso, o gRNA e a proteína Cas9 são expressos a partir de um vetor viral AAV. Neste caso, a transcrição do gRNA é dirigida por um promotor U6 e a transcrição de mRNA de Cas9 é dirigida a partir de um promotor ubíquo como EF1-alfa ou preferencialmente de um promotor e potenciador específico do fígado, como o promotor/potenciador de transtirretina. O tamanho do gene spCas9 (4,4 Kb) impede a inclusão dos cassetes spCas9 e gRNA em um único AAV, exigindo assim um AAV separado para entregar o gRNA e spCas9. Em um segundo caso, é usado um vetor AAV que tem elementos de sequência que promovem a auto-inativação do genoma viral. Neste caso, a inclusão de locais de clivagem para o gRNA no DNA do vetor resulta na clivagem do DNA do vetor *in vivo*. Ao incluir locais de clivagem em locais que bloqueiam a expressão da Cas9 quando clivados, a expressão da Cas9 é limitada a um período de tempo mais curto. Na terceira abordagem alternativa para entregar o gRNA e Cas9 às células *in vivo*, é usado um método de entrega não viral. Em um exemplo, nanopartículas lipídicas (LNP) são usadas como um método de entrega não viral. Vários lipídios catiônicos ionizáveis diferentes

estão disponíveis para uso nas LNP. Estes incluem C12-200 (Love et al (2010), PNAS vol. 107, 1864–1869), MC3, LN16, MD1, entre outros. Em um tipo de LNP, uma fração GalNac é anexada à parte externa da LNP e atua como um ligante para absorção no fígado por meio do receptor da asialoglicoproteína. Qualquer um desses lipídios catiônicos é usado para formular LNP para a entrega de gRNA e mRNA de Cas9 ao fígado.

[00608] Para avaliar a integração direcionada e expressão de FVIII, os camundongos com Hemofilia A são injetados primeiro intravenosamente com um vírus AAV, preferencialmente um vírus AAV8 que encapsula o molde de DNA doador de FVIII. A dose de AAV varia de 10^{10} a 10^{12} genomas de vetor (VG) por equivalente de camundongo a $4x10^{11}$ a $4\ x10^{13}$ VG/kg. Entre 1 h e 7 dias após a injeção do doador em AAV, os mesmos camundongos recebem injeções iv de uma LNP que encapsula o gRNA e o mRNA de Cas9. O mRNA de Cas9 e o gRNA são encapsulados em LNP separadas e depois misturados antes da injeção a uma razão de massa de RNA de 1:1. A dose de LNP administrada varia de 0,25 a 2 mg de RNA por kg de peso corporal. A LNP é dosada por injeção nas veias da cauda ou por injeção retro-orbital. O impacto do tempo de injeção da LNP em relação à injeção de AAV na eficiência da integração direcionada e na expressão da proteína FVIII é avaliado por tempos de teste de 1 h, 24 h, 48 h, 72 h, 96 h, 120 h, 144 h e 168 h após a administração do AAV.

[00609] Em outro exemplo, o molde de DNA doador é administrado *in vivo* usando um sistema de administração não viral que é uma LNP. As moléculas de DNA são encapsuladas em partículas de LNP similares às descritas acima e entregues aos hepatócitos no fígado após a injeção iv. Embora a fuga do DNA do endossoma para o citoplasma ocorra de maneira relativamente eficiente, a translocação de grandes moléculas de DNA carregadas para o núcleo não é eficiente. Em um caso, a maneira de melhorar a entrega de DNA ao núcleo é imitando o genoma

do AAV pela incorporação do ITR do AAV no molde de DNA doador. Neste caso, as sequências ITR estabilizam o DNA ou de outra forma melhoram a translocação nuclear. A remoção dos dinucleotídeos CG (sequências CpG) da sequência do molde de DNA doador também melhora a entrega nuclear. O DNA contendo dinucleotídeos CG é reconhecido pelo sistema imune inato e eliminado. A remoção de sequências CpG que estão presentes em sequências de DNA artificial melhora a persistência da entrega de DNA por vetores não virais e virais. O processo de otimização de códons normalmente aumenta o conteúdo dos dinucleotídeos CG porque, em muitos casos, os códons mais frequentes têm um resíduo C na 3ª posição, o que aumenta a possibilidade de criar um CG quando o próximo códon começa com um G. Uma combinação da entrega de LNP do molde de DNA doador seguida 1 hora a 5 dias depois com uma LNP contendo o gRNA e o mRNA de Cas9 é avaliada em camundongos com Hemofilia A.

[00610] Para avaliar a eficácia da administração *in vivo* de gRNA/Cas9 e molde de DNA doador, os camundongos com Hemofilia injetados são avaliados quanto aos níveis de FVIII no sangue em diferentes tempos, começando cerca de 7 dias após a administração do segundo componente. As amostras de sangue são coletadas por sangramento RO e o plasma é separado e analisado quanto à atividade do FVIII usando um ensaio cromogênico (Diapharma). Os padrões de proteína FVIII são usados para calibrar o ensaio e calcular as unidades por mL da atividade do FVIII no sangue.

[00611] A expressão do mRNA do FVIII também é medida nos fígados dos camundongos no final do estudo. O RNA total extraído dos fígados dos camundongos é analisado quanto aos níveis de mRNA da albumina e mRNA do FVIII usando Q-PCR. A razão de mRNA do FVIII para mRNA da albumina quando comparada a camundongos não tratados é uma indicação da % de transcritos de albumina que foram

coescolhidos para produzir um mRNA híbrido de albumina-FVIII.

[00612] O DNA genômico dos fígados dos camundongos tratados é avaliado quanto a eventos de integração direcionados no local alvo do gRNA, especificamente no íntron 1 da albumina. Os pares de iniciadores de PCR são projetados para amplificar os fragmentos de junção em cada extremidade da integração direcionada prevista. Esses iniciadores foram desenhados para detectar a integração nas orientações direta e reversa. A sequência dos produtos de PCR confirma se o evento de integração esperado ocorreu. Para quantificar a porcentagem de alelos de albumina que sofreram integração direcionada é sintetizado um padrão que corresponde aos fragmentos de junção esperados. Quando adicionado ao DNA genômico de camundongos não tratados em diferentes concentrações e depois submetido à mesma reação de PCR, uma curva padrão é gerada e usada para calcular o número de cópias de alelos com eventos de integração nas amostras de camundongos tratados.

Exemplo 5: INTEGRAÇÃO DIRECIONADA AO ÍNTRON 1 DA ALBUMINA DE PRIMATAS

[00613] As mesmas metodologias descritas no Exemplo 4 para o camundongo são aplicadas a espécies de primatas usando um gRNA que tem como alvo o íntron 1 da albumina do primata. O AAV8 ou uma LNP é usado para administrar primeiro o molde de DNA doador por injeção iv. As doses usadas são baseadas nas que foram bem-sucedidas nos camundongos. Subsequentemente, os mesmos primatas recebem injeções iv da LNP encapsulando o gRNA e o mRNA de Cas9. São usadas a mesma formulação e doses da LNP eficazes nos camundongos. Como não existe um modelo de primatas para a Hemofilia, a proteína FVIII precisa ser medida usando um teste ELISA humano específico para o FVIII. As mesmas análises moleculares da integração direcionada e dos níveis de mRNA do FVIII descritos no

Exemplo 4 são realizadas no primata. O primata é um bom modelo pré-clínico para permitir a avaliação translacional para a clínica.

Exemplo 6: AVALIAÇÃO DA CLIVAGEM NO E FORA DO ALVO POR GRNA/CAS9 E INTEGRAÇÃO DIRECIONADA EM HEPATÓCITOS PRIMÁRIOS HUMANOS

[00614] Os hepatócitos humanos primários são o tipo de célula mais relevante para a avaliação da potência e clivagem fora do alvo de um gRNA/Cas9 que será administrado ao fígado dos pacientes. Estas células são cultivadas em cultura como monocamadas aderentes durante um período limitado. Foram estabelecidos métodos para a transfecção de células aderentes com mRNA, por exemplo, Message Max (Thermo Fisher). Após a transfecção com uma mistura de mRNA de Cas9 e gRNA, a eficiência de clivagem no alvo é medida usando a análise TIDES. As mesmas amostras de DNA genômico são submetidas a análises fora do alvo para identificar locais adicionais no genoma que foram clivados pelo complexo gRNA/Cas9. Um desses métodos é "GuideSeq" (Tsai et al Nat Biotechnol. 2015 Fev; 33(2):187-197). Outros métodos incluem sequenciamento profundo, Sequenciamento de todo o genoma, ChIP-seq (Nature Biotechnology 32,677–683 2014), BLESS (2013 Crosetto et al. Doi: 10.1038/nmeth.2408), sequenciamento de translocação, de elevado rendimento, à escala do genoma (HTGTS) como descrito em 2015 Frock et al. doi: 10.1038/nbt.3101, Digenome-seq (2015 Kim et al. doi: 10.1038/nmeth.3284) e IDLV (2014 Wang et al. doi: 10.1038/nbt.3127).

[00615] Os hepatócitos humanos primários também são transduzidos por vírus AAV contendo o molde de DNA doador. Em particular, os sorotipos AAV6 ou AAVDJ são particularmente eficientes na transdução de células em cultura. Entre 1 e 48 h após a transdução pelo AAV-DNA doador, as células são então transfectadas com o gRNA e o mRNA de Cas9 para induzir a integração direcionada. Os eventos

de integração direcionados são medidos usando as mesmas abordagens baseadas em PCR descritas no Exemplo 4.

Exemplo 7: IDENTIFICAÇÃO E SELEÇÃO DE RNA GUIA QUE CLIVA EFICAZMENTE NO ÍNTRON 1 DA ALBUMINA HUMANA EM HEPATÓCITOS HUMANOS PRIMÁRIOS EM CULTURA

[00616] Quatro gRNA (T4, T5, T11, T13) foram selecionados com base na homologia perfeita com o primata não humano e na análise da eficiência de corte em células HuH7 e HepG2 (Tabela 4), para avaliação da eficiência de corte em hepatócitos humanos primários. Os hepatócitos humanos primários (obtidos da BioIVT) foram descongelados, transferidos para o Meio de Recuperação de Hepatócitos Criopreservados (CHRM) (Gibco), peletizados a baixa velocidade e depois plaqueados em Meio *InVitro*GRO™ CP (BioIVT) mais Mistura de Antibióticos *Torpedo*™ (BioIVT) a uma densidade de 0,7 x 10^6 células/mL em placas de 24 poços pré-revestidas com Colágeno IV (Corning). As placas foram incubadas em 5% de CO2 a 37 °C. Depois de as células terem aderido (3-4 horas após o plaqueamento), as células mortas que não aderiram à placa foram lavadas com meio completo quente e fresco, e depois as células foram incubadas em 5% de CO2 a 37 °C. Para transfectar as células, o mRNA de Cas9 (Trilink) e o RNA guia (Synthego Corp, Menlo Park, CA) foram descongelados em gelo e depois adicionados a 30 µL de meio OptiMem (Gibco) a 0,6 µg de mRNA e 0,2 µg de guia por poço. MessengerMax (ThermoFisher) diluído em 30 µL em OptiMem a um volume de 2:1 para o peso total de ácido nucleico foi incubado com a solução OptMem com mRNA de Cas9/gRNA à temperatura ambiente durante 20 minutos. Esta mistura foi adicionada gota a gota aos 500 µL de meio de plaqueamento de hepatócitos por poço de hepatócitos cultivados em uma placa de 24 poços e as células incubadas em 5% de CO2 a 37 °C. As células foram lavadas e novamente alimentadas na manhã seguinte e 48 h após a

transfecção as células foram coletadas para extração de DNA genômico por adição de 200 µL de Tripsina-EDTA (Gibco) a 0,25% a cada poço e incubação de 5 a 10 minutos a 37 °C. Uma vez as células tendo sido deslocadas, 200 µL de FBS (Gibco) foram adicionados para inativar a tripsina. Após adição a 1 mL de PBS (Gibco), as células foram peletizadas a 1200 rpm durante 3 minutos e depois ressuspensas em 50 µL de PBS. O DNA genômico foi extraído usando o kit MagMAX DNA Multi-Sample Ultra 2.0 (Applied Biosytems) seguindo as instruções no kit. A qualidade e concentração do DNA genômico foram analisadas usando um espectrofotômetro. Para a análise TIDE, o DNA genômico foi amplificado por PCR usando iniciadores que flanqueiam o local de clivagem previsto no alvo (AlbF: CCCTCCGTTTGTCCTAGCTTTTC, SEQ ID NO: 178, e AlbR: CCAGATACAGAATATCTTCCTCAACGCAGA, SEQ ID NO: 179) e Platinum PCR SuperMix High Fidelity (Invitrogen) usando 35 ciclos de PCR e uma temperatura de hibridação de 55 ^{0}C. Os produtos de PCR foram analisados primeiro por eletroforese em gel de agarose para confirmar que o tamanho certo do produto (1053 pb) havia sido gerado, depois purificados e sequenciados usando iniciadores (For: CCTTTGGCACAATGAAGTGG, SEQ ID NO: 180, rev: GAATCTGAACCCTGATGACAAG, SEQ ID NO: 181). Os dados do sequenciamento foram então analisados usando uma versão modificada do algoritmo TIDES (Brinkman et al (2104); Nucleic Acids Research, 2014, 1) chamado Tsunami. Este determina a frequência de inserções e deleções (INDELS) presentes no local de corte previsto para o complexo gRNA/Cas9.

[00617] Foram testados os RNA guia contendo a sequência alvo padrão de 20 nucleotídeos ou uma sequência alvo de 19 nucleotídeos (1 pb mais curto na extremidade 5') dos guias T4, T5, T11 e T13 (sintetizados quimicamente nos AxoLabs, Kulmbach, Alemanha ou

Synthego Corp, Menlo Park, CA). Um gRNA de 19 nucleotídeos pode ser mais específico para a sequência, mas um guia mais curto pode ter menor potência. Guias de controle direcionados ao locus AAVS1 humano e fator de complemento humano foram incluídos para comparação entre doadores. A frequência de INDEL no local alvo no íntron 1 da albumina foi medida 48 h após a transfecção usando o método TIDES. A FIG. 6 resume os resultados das transfecções de hepatócitos primários de 4 doadores humanos diferentes. Os resultados demonstram eficiências de corte que variam de 20% a 80% para os diferentes guias. A versão de 20 nucleotídeos de cada gRNA da albumina era consistentemente mais potente que a variante de 19 nucleotídeos. A potência superior do gRNA de 20 nucleotídeos pode compensar qualquer benefício potencial que um gRNA de 19 nucleotídeos possa ter em termos de corte fora do alvo. O RNA guia T4 exibiu o corte mais consistente nos 4 doadores de células com frequências INDEL de cerca de 60%. Os gRNA T4, T5, T11 e T13 foram selecionados para análise fora do alvo.

EXEMPLO 8: Identificação de locais fora do alvo para o RNA guia da albumina humana

[00618] Duas abordagens para identificação de locais fora do alvo para CRISPR/Cas9 são previsão *ab initio* e detecção empírica. A especificação do local de clivagem Cas9 pelo RNA guia é um processo imperfeito, pois a clivagem Cas9 tolera não correspondências entre a sequência do RNA guia e o genoma. É importante conhecer o espectro dos locais de clivagem Cas9 para entender o risco de segurança de diferentes guias e selecionar guias com o perfil fora do alvo mais favorável. O método preditivo é baseado no Guido, uma ferramenta de software adaptada do algoritmo CCTop para previsão fora do alvo (Stemmer *et al.*, 2015). O Guido usa o algoritmo Bowtie 1 para identificar potenciais locais de clivagem fora do alvo, procurando homologia entre

o RNA guia e toda a construção GRCh38/hg38 do genoma humano (Langmead *et al.*, 2009). Guido detecta sequências com até 5 não correspondências com o RNA guia, priorizando a homologia proximal do PAM e um NGG de PAM corretamente posicionado. Os locais foram classificados pelo número e posição das suas não correspondências. Para cada execução, a sequência guia e o PAM genômico são concatenados e executados com parâmetros padrão. Os principais acertos com três ou menos não correspondências são mostrados nas Tabelas 5-8 abaixo para os guias da albumina T4, T5, T11 e T13. A primeira linha de cada tabela mostra o local no alvo no genoma humano, as linhas abaixo que mostram os locais fora do alvo previstos.

Tabela 5

Locais fora do alvo previstos por Guido para hALB T4						
Cro-mos-somo	**Posição**	**Gene**	**Tipo**	**Não Corres-pon-dên-cias**	**Sequência**	**PAM**
4	73404720	ALB	Intrônica	0	TAAAGCATAGTGCAA TGGAT (SEQ ID NO: 106)	AGG
1	105184629		Intergê-nica	2	GAAAGCATGGTGCA ATGGAT (SEQ ID NO: 107)	TGG
20	51270388		Intergê-nica	3	TATTGCACAGTGCAA TGGAT (SEQ ID NO: 108)	GGG
4	30923943	PCDH7	Intrônica	3	TGATGCATATTGCAA TGGAT (SEQ ID NO: 109)	TGG
1	58844572	RP11-63G10,2	Intrônica	3	TAATGAATAGGGCAA TGGAT (SEQ ID NO: 110)	TGG
1	107412556	NTNG1	Intrônica	3	TAAGGCACAGTGTA ATGGAT (SEQ ID NO: 111)	TGG
8	10123839	MSRA	Intrônica	3	AAAAGCATAGACCAA TGGAT (SEQ ID NO: 112)	TGG
Y	10935087		Intergê-nica	3	TAGAGTATAGTGCA GTGGAT (SEQ ID NO: 113)	TGG

X	21813781		Intergênica	3	CAAAGCAAAGTGCAATTGAT (SEQ ID NO: 114)	GGG
3	31414024		Intergênica	3	GGAAGCATAGTGCAATGGTT (SEQ ID NO: 115)	GGG
2	177957869	AC011998,1	Intrônica	3	TAAAGGATAGAGCAATGTAT (SEQ ID NO: 116)	AGG
Y	10775325		Intergênica	3	TAGAGTATAGTGCAATGGAG (SEQ ID NO: 117)	TGG
8	116113757	LINC00536	Intrônica	3	TAAAGAATAGTGAAATGGTT (SEQ ID NO: 118)	TGG

Tabela 6

Locais fora do alvo previstos por Guido para hALB T5						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Não Correspondências**	**Sequência**	**PAM**
4	73404759	ALB	Intrônica	0	ATTTATGAGATCAACAGCAC (SEQ ID NO: 119)	AGG
19	31798902		Intergênica	2	ATTTATGATATCATCAGCAC (SEQ ID NO: 120)	CGG
11	98512684		Intergênica	3	AAATATGACATCAACAGCAC (SEQ ID NO: 121)	AGG
17	12093264	MAP2K4	Intrônica	3	ATCTTTGAGATCATCAGCAC (SEQ ID NO: 122)	TGG
21	35820764	RUNX1	Intrônica	3	ATGTATCAGATCATCAGCAC (SEQ ID NO: 123)	GGG
19	29334372	CTC-525D6,1	Intrônica	3	AATTATGAGATTCACAGCAC (SEQ ID NO: 124)	AGG
2	116633233		Intergênica	3	ATTTATGTGTTCAACCGCAC (SEQ ID NO: 125)	AGG
9	90654432		Intergênica	3	ATATATGACATCAACAGAAC (SEQ ID NO: 126)	AGG
6	17047800		Intergênica	3	ACTTATGATATCAACAGCAT (SEQ ID NO: 127)	TGG

Tabela 7

Locais fora do alvo previstos por Guido para hALB T11						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Não Correspondências**	**Sequência**	**PAM**
4	73404725	ALB	Intrônica	0	TTAAATAAAGCATAGTGCAA (SEQ ID NO: 128)	TGG
2	229867834	TRIP12	Intrônica	1	TAAAATAAAGCATAGTGCAA (SEQ ID NO: 129)	AGG
14	91174270	C14orf159	Intrônica	2	TTAAATAAAGGATATTGCAA (SEQ ID NO: 130)	AGG
16	73177850		Intergênica	2	TTAAATAAAGCATTGAGCAA (SEQ ID NO: 131)	GGG
4	1839915	LETM1	Intrônica	3	TACTATAAAGCATAGTGCAA (SEQ ID NO: 132)	AGG
4	82950298	LIN54	Intrônica	3	TACTATAAAGCATAGTGCAA (SEQ ID NO: 133)	GGG
3	133084865	TMEM108	Intrônica	3	TTAAGGAAACCATAGTGCAA (SEQ ID NO: 134)	AGG
8	5026909		Intergênica	3	ATAAATATATCATAGTGCAA (SEQ ID NO: 135)	AGG
8	59960346		Intergênica	3	CTAAATAGAGAATAGTGCAA (SEQ ID NO: 136)	TGG
21	18677763	MIR548X	Intrônica	3	TTAAAGAAATTATAGTGCAA (SEQ ID NO: 137)	GGG
X	66550751		Intergênica	3	TTAAATATATAATAGTGCAA (SEQ ID NO: 138)	GGG
X	109390455	GUCY2F	Intrônica	3	TTAAAAACAGCACAGTGCAA (SEQ ID NO: 139)	AGG
10	20767685		Intergênica	3	TTAAAATAAGCATGGTGCAA (SEQ ID NO: 140)	GGG
15	54261380	UNC13C	Intrônica	3	TTTGATAAAGCATAGGGCAA (SEQ ID NO: 141)	TGG
1	230563372		Intergênica	3	TTTTATAAAGCATAGTCCAA (SEQ ID NO: 142)	AGG

15	56985313	TCF12	Intrônica	3	TTAAATGAAGAATATTGCAA (SEQ ID NO: 143)	AGG
3	153332862		Intergênica	3	ATAAATAAAGAATAGAGCAA (SEQ ID NO: 144)	GGG
14	31932077		Intergênica	3	TTGAATAAAGCAGAGTGGAA (SEQ ID NO: 145)	GGG
12	38399588		Intergênica	3	TTAATTAATGCATAGTGCCA (SEQ ID NO: 146)	GGG
7	141092721	TMEM178B	Intrônica	3	TTAGATAAAGCTTAGTGCTA (SEQ ID NO: 147)	AGG
4	60292980		Intergênica	3	TTAGATAAAGCATACTGGAA (SEQ ID NO: 148)	TGG
2	155632685		Intergênica	3	TTAAAGAAAGCATGGTGCAG (SEQ ID NO: 149)	TGG
8	19144500	RP11-1080G15,2	Intrônica	3	TTACATAAAGCATACTGCAT (SEQ ID NO: 150)	GGG
22	44584358		Intergênica	3	TTATATAAAGCATAGAGCAG (SEQ ID NO: 151)	GGG
20	47604347	NCOA3	Intrônica	3	TTAAATGAAGCATAGTGAAG (SEQ ID NO: 152)	AGG

Tabela 8

Locais fora do alvo previstos por Guido para hALB T13						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Não Correspondências**	**Sequência**	**PAM**
4	73404562	ALB	Intrônica	0	TAATAAAATTCAAACATCCT (SEQ ID NO: 153)	AGG
10	33567530		Intergênica	2	GAATAAAATTCTAACATCCT (SEQ ID NO: 154)	TGG
2	53855928	GPR75	Intrônica	2	TAATATAATTCCAACATCCT (SEQ ID NO: 155)	TGG
10	7439135		Intergênica	2	AAATAAAATTCAAACTTCCT (SEQ ID NO: 156)	TGG
11	106969296	GUCY1A2	Intrônica	3	GAGTTAAATTCAAACATCCT (SEQ ID NO: 157)	GGG

14	52353218		Inter-gênica	3	TTTTAAAAATCAA ACATCCT (SEQ ID NO: 158)	GGG
3	25222362	RARB	Intrônica	3	AAATGAAAGTCAA ACATCCT (SEQ ID NO: 159)	TGG
18	29352071	CTD-2515C13,2	Intrônica	3	GATTAAAATTTAA ACATCCT (SEQ ID NO: 160)	TGG
1	48069696	RP4-683M8,2	Intrônica	3	TCTTAAAATTCCA ACATCCT (SEQ ID NO: 161)	AGG
20	22206955		Inter-gênica	3	AAAAAAAATTCCA ACATCCT (SEQ ID NO: 162)	TGG
2	145716708		Inter-gênica	3	TACTGAAATTCTA ACATCCT (SEQ ID NO: 163)	AGG
4	135277467		Inter-gênica	3	TACAAAAATTCAC ACATCCT (SEQ ID NO: 164)	GGG
2	114502757	Clostridiales-1	Intrônica	3	TATTAGAATTCAG ACATCCT (SEQ ID NO: 165)	TGG
12	65459700	MSRB3	Intrônica	3	TAATAAAGCCCAA ACATCCT (SEQ ID NO: 166)	AGG
6	6201132	F13A1	Intrônica	3	TATTTAAATTCAA ATATCCT (SEQ ID NO: 167)	TGG
2	213458045	SPAG16	Intrônica	3	AAATAAAGTTCAA AGATCCT (SEQ ID NO: 168)	GGG
5	4307376		Inter-gênica	3	TACAAAAATTCAA ACTTCCT (SEQ ID NO: 169)	TGG
2	201942075		Inter-gênica	3	GAATAAAATTTAA ATATCCT (SEQ ID NO: 170)	AGG
8	141400355	CTD-3064M3,4	Intrônica	3	TATAAAAATTCAA ACAGCCT (SEQ ID NO: 171)	GGG
14	39294628	CTAGE5	Intrônica	3	TACTAAAATTTAA ACTTCCT (SEQ ID NO: 172)	GGG
14	72580607	RP3-514A23,2	Intrônica	3	TAATAACCTTCAA ACATTCT (SEQ ID NO: 173)	TGG
12	3628277	CRACR2A	Intrônica	3	TAGTAAAATTCAA ATGTCCT (SEQ ID NO: 174)	AGG

21	42611948	AP001626,1	Intrônica	3	CAATAAAATTCAA CCATCAT (SEQ ID NO: 175)	GGG
7	16480457	GS1-166A23,1	Intrônica	3	GAATAAAATTCAA ACTTCTT (SEQ ID NO: 176)	TGG
12	108648894	CORO1C	Intrônica	3	AAATAAAATTCAA AAATCCC (SEQ ID NO: 177)	AGG

[00619] Além disso, locais fora do alvo para o gRNA da albumina humana T4, T5, T11, T13 em células de hepatócitos humanos foram identificados usando um método chamado GUIDE-seq. O GUIDE-seq (Tsai *et al.* 2015) é um método empírico para encontrar locais de clivagem fora do alvo. O GUIDE-seq se baseia na captura espontânea de um oligonucleotídeo no local de uma quebra da fita dupla no DNA cromossômico. Em resumo, após a transfecção de células relevantes com o complexo gRNA/Cas9 e oligonucleotídeo de fita dupla, o DNA genômico é purificado a partir das células, sonicado e uma série de ligações adaptadoras é realizada para criar uma biblioteca. As bibliotecas contendo oligonucleotídeos são sujeitas a sequenciamento de DNA de elevado rendimento e o resultado é processado com o software GUIDE-seq padrão para identificar o local da captura de oligonucleotídeos.

[00620] Em detalhe, o oligo GUIDEseq de fita dupla foi gerado emparelhando dois oligonucleotídeos de fita simples complementares por aquecimento a 89°C e depois arrefecendo lentamente até à temperatura ambiente. Os complexos de proteínas ribonucleares (RNP) foram preparados misturando 240 pmol de RNA guia (Synthego Corp, Menlo Park, CA) e 48 pmol de Cas9 TruCut a 20 µMolar (ThermoFisher Scientific) em um volume final de 4,8 µL. Em um tubo separado, 4 µL do oligonucleotídeo de fita dupla GUIDeseq 10 µMolar foram misturados com 1,2 µL da mistura RNP e depois adicionados a um cassete Nucleofection (Lonza). A isto foram adicionados 16,4 µL de solução Nucleofector SF (Lonza) e 3,6 µL de suplemento (Lonza). As células

HepG2 cultivadas como culturas aderentes foram tratadas com tripsina para liberá-las da placa e após a desativação da tripsina foram peletizadas e ressuspensas a 12,5 e6 células/mL em solução Nucleofector e 20 µL (2,5 e5 células) adicionados a cada cubeta de nucleofecção. A nucleofecção foi realizada com o programa de células EH-100 na Unidade Nucleofector 4-D (Lonza). Após incubação à temperatura ambiente durante 10 minutos, foram adicionados 80 µL de meio HepG2 completo e a suspensão de células foi colocada em um poço de uma placa de 24 poços e incubada a 37 °C em 5% de CO_2 durante 48 horas. As células foram soltas com tripsina, peletizadas por centrifugação (300 g, 10 min) e depois o DNA genômico foi extraído usando o Kit DNAeasy Blood and Tissue (Qiagen). A região do íntron 1 da Albumina humana foi amplificada por PCR usando os iniciadores AlbF (CCCTCCGTTTGTCCTAGCTTTTC, SEQ ID NO: 178) e AlbR (CCAGATACAGAATATCTTCCTCAACGCAGA, SEQ ID NO: 179) e Platinum PCR SuperMix High Fidelity (Invitrogen) usando 35 ciclos de PCR e uma temperatura de hibridação de 55 ^{0}C. Os produtos de PCR foram analisados primeiro por eletroforese em gel de agarose para confirmar que o tamanho certo do produto (1053 pb) havia sido gerado, depois sequenciados diretamente usando iniciadores (For: CCTTTGGCACAATGAAGTGG, SEQ ID NO: 180, rev: GAATCTGAACCCTGATGACAAG, SEQ ID NO: 181). Os dados do sequenciamento foram então analisados usando uma versão modificada do algoritmo TIDES (Brinkman et al (2104); Nucleic Acids Research, 2014, 1) chamado Tsunami. Este determina a frequência de inserções e deleções (INDELS) presentes no local de corte previsto para o complexo gRNA/Cas9. Comparado ao protocolo descrito por Tsai *et al.* realizamos GUIDE-seq com o oligonucleotídeo de captura a 40 pmol (~1,67 µM) para aumentar a sensibilidade da identificação do local de clivagem fora do alvo. Para atingir uma sensibilidade de

aproximadamente 0,01%, definimos um mínimo de 10.000 leituras únicas de sequência no alvo por transfecção com um mínimo de 50% de clivagem no alvo. Amostras sem transfecção de RNPs foram processadas em paralelo. Os locais (+/- 1 kb) encontrados nas amostras que contêm RNP e que não contêm RNP são excluídos de análises posteriores.

[00621] GUIDE-seq foi realizado na linha celular de hepatoma humano HepG2. Em HepG2 a captura do oligonucleotídeo GUIDE-seq nos locais no alvo estava na gama de 70% a 200% da frequência NHEJ, demonstrando eficiente captura de oligo.

[00622] O adaptador Y foi preparado por emparelhamento do Adaptador Comum a cada um dos adaptadores de código de barras de amostra (A01 - A16) que contêm o índice molecular de 8 mer. O DNA genômico extraído das células HepG2 que foram nucleofectadas com RNP e o oligo GUIDEDseq foi quantificado usando Qubit e todas as amostras normalizadas para 400 ng em volume de 120 µL de Tampão TE. O DNA genômico foi cisalhado em um comprimento médio de 200 pb de acordo com o procedimento operacional padrão para o sonicador Covaris S220. Para confirmar o comprimento médio do fragmento, 1 µL da amostra foi analisada em uma TapeStation de acordo com o protocolo do fabricante. As amostras de DNA cisalhado foram limpas com esferas AMPure XP SPRI de acordo com o protocolo do fabricante e eluídas em 17 µL de Tampão TE. A reação de reparação de extremidades foi realizada no DNA genômico por mistura de 1,2 µL de mistura de dNTP (5 mM de cada dNTP), 3 µL de Tampão T4 DNA Ligase 10x, 2,4 µL de Mistura End-Repair, 2,4 µL de Tampão Taq Platinum 10x (livre de Mg2+) e 0,6 µL de Polimerase Taq (sem Hotstart) e 14 µL de amostra de DNA cisalhado (da etapa precedente) para um volume total de 22,5 µL por tubo e incubada em um termociclador (12ºC 15 min; 37ºC 15 min; 72ºC 15 min; manter a 4ºC). A isto foi adicionado 1 µL de

Adaptador Y emparelhado (10 µM), 2 µL de DNA Ligase T4 e a mistura incubada em um termociclador (16°C, 30 min; 22°C, 30 min; manter a 4°C). A amostra foi limpa usando umas esferas AMPure XP SPRI de acordo com o protocolo do fabricante e eluída em 23 µL de Tampão TE. Foi executado 1 µL de amostra em uma TapeStation de acordo com o protocolo do fabricante para confirmar a ligação dos adaptadores aos fragmentos. Para preparar a biblioteca GUIDEseq, foi preparada uma reação contendo 14 µL de H_2O livre de nuclease, 3,6 µL de Tampão Taq Platinum 10 x, 0,7 µL de mistura dNTP (10 mM cada), 1,4 µL de $MgCl_2$ a 50 mM, 0,36 µL de Polimerase Taq Platinum, 1,2 µL de iniciador sentido ou anti-sentido específico do gene (10 µM), 1,8 µM de TMAC (0,5 M), 0,6 µL de P5_1 (10 µM) e 10 µL da amostra da etapa precedente. Esta mistura foi incubada em um termociclador (95 °C 5 min, depois 15 ciclos de 95°C 30 s, 70°C (menos 1 °C por ciclo) durante 2 min, 72 °C 30 s, seguido por 10 ciclos de 95 °C 30 s, 55 °C 1min, 72°C 30 s, seguido por 72°C 5 min). A reação de PCR foi limpa usando esferas AMPure XP SPRI de acordo com o protocolo do fabricante e eluída em 15 µL de Tampão TE. Foi verificado 1 µL de amostra no TapeStation de acordo com o protocolo do fabricante para rastrear o progresso da amostra. Uma segunda PCR foi realizada misturando 6,5 µL de H_2O livre de Nuclease, 3,6 µL de tampão Taq Platinum 10x (livre de Mg2+), 0,7 µL de mistura de dNTP (10 mM cada), 1,4 µL de $MgCl_2$ (50 mM), 0,4 µL de Polimerase Taq Platinum, 1,2 µL de Iniciador Específico do Gene (GSP) 2 (sentido; + ou anti-sentido; -), 1,8 µL de TMAC (0,5 M), 0,6 µL P5_2 (10 µM) e 15 µL do produto de PCR da etapa precedente. Se GSP1+ foi usado na primeira PCR, então GSP2+ foi usado na PCR2. Se o iniciador GSP1- foi usado na primeira reação de PCR, então o iniciador GSP2- foi usado nesta segunda reação de PCR. Após a adição de 1,5 µL de P7 (10 µM), a reação foi incubada em um termociclador com o seguinte programa: 95ºC 5 min, depois 15

ciclos de 95°C 30 s, 70°C (menos 1°C por ciclo) durante 2 minutos, 72°C 30 s, seguidos por 10 ciclos de 95°C 30 s, 55°C 1 min, 72°C 30 s, seguido de 72°C 5 min. A reação de PCR foi limpa usando esferas AMPure XP SPRI de acordo com o protocolo do fabricante e eluída em 30 µL de Tampão TE e 1 µL analisados em um TapeStation de acordo com o protocolo do fabricante para confirmar a amplificação. A biblioteca de produtos de PCR foi quantificada usando o kit Kapa Biosystems para Illumina Library Quantification, de acordo com o protocolo fornecido pelo fabricante e submetida a sequenciamento de próxima geração no sistema Illumina para determinar os locais nos quais o oligonucleotídeo se integrou.

[00623] Os resultados do GUIDE-seq estão listados nas Tabelas 9 a 12. É importante levar em consideração a sequência alvo prevista identificada pelo GUIDE-seq. Se a sequência alvo prevista não possui um PAM ou carece de homologia significativa com o gRNA, por exemplo, mais de 5 não correspondências (mm), esses locais genômicos não são considerados locais fora do alvo verdadeiros, mas sinais de fundo do ensaio. A abordagem GUIDE-seq resultou em uma alta frequência de captura de oligo nas células HepG2, indicando que este método é apropriado neste tipo de célula. As contagens de leitura no alvo atenderam aos critérios predefinidos de um mínimo de 10.000 nas leituras no alvo para 3 dos 4 guias. Um pequeno número de locais fora do alvo para os 4 candidatos gRNA principais foi identificado. O número de locais fora do alvo verdadeiros (ou seja, contendo um PAM e tendo homologia significativa com o gRNA) variou de 0 a 6 para os 4 gRNA. O guia T4 exibiu 2 locais fora do alvo que parecem reais. A frequência destes eventos no GUIDE-seq, a julgar pela contagem da leitura de sequenciamento, foi de 2% e 0,6% da frequência de clivagem no alvo. Os guias T13 e T5 não exibiram locais fora do alvo pelo GUIDE-seq possuindo homologia com o gRNA e contendo um PAM e, portanto,

parecem ter o perfil fora do alvo mais desejável dos 4 guias testados. O gRNA T11 exibiu um local fora do alvo com uma contagem de leitura relativamente alta que era 23% da contagem de leitura no alvo, o que sugere que este guia é menos atraente para uso terapêutico.

Tabela 9

hAlb T4 (Os dez principais locais potenciais de clivagem fora do alvo nas células HepG2)						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Sequência Alvo Prevista**	**Leituras**	**Comentário**
4	74270442	ALB	Exônica	TAAAGCATAGTGCAATGGAT (SEQ ID NO: 106)	25960	Local alvo
14	55327909	GCH1	Intrônica	TAAAGCATAGTGCCAATGGAT (SEQ ID NO: 182)	515	protuberância 1 nt
6	132002620	ENPP3	Intrônica	GAAAGCATAATAGCAATGGAT (SEQ ID NO: 183)	158	2 mm, protuberância
9	135937394	CEL	Exônica	CTCACCATGGGGCGCCTGCAACTGGTT (SEQ ID NO: 184)	142	Nenhum PAM
1	107955183	NTNG1	Exônica	TAAGGCACAGTGTAATGGAT (SEQ ID NO: 111)	58	3mm, forma de splice alternativa
3	31455533		Intergênica	GGAAGCATAGTGCAATGGTT (SEQ ID NO: 115)	50	3 mm
2	148438664		Intergênica		37	Nenhum PAM, homologia
6	31941379	STK19	Intrônica		35	Nenhum PAM, homologia
1	121485148		Intergênica	TAAAGGATCGTTCAACTCTGTGAGT (SEQ ID NO: 185)	37	Nenhum PAM
21	28069029		Intergênica	CCAAGCATAGGTAATGGAT (SEQ ID NO: 186)	15	3 mm, anti-protuberância
1	245132152		Intergênica	AAAAGCATAGTGAATGAAT (SEQ ID NO: 187)	12	2 mm, anti-protuberância

Tabela 10

hAlb T5 (Os dez principais locais potenciais de clivagem fora do alvo nas células HepG2)						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Sequência Alvo Prevista**	**Leituras**	**Comentário**
4	74270481	ALB	Exônica	ATTTATGAGATCAACAGCAC (SEQ ID NO: 119)	15407	Local alvo
5	171126779		Intergênica		114	Nenhum PAM, nenhuma homologia
11	5271381		Intergênica		51	Nenhum PAM, nenhuma homologia
1	121485232		Intergênica	CTTCGTATAGAAACAAGACAG (SEQ ID NO: 188)	40	Nenhum PAM, repetição
	128137	LOC101928428	Intrônica	GAGAGAGAGAGAAAGAGACAG (SEQ ID NO: 189)	23	Repetição simples
12	25600072		Intergênica	ATTCTAGAGGCATAGAGAGTTCAACCT (SEQ ID NO: 190)	20	Nenhum PAM, homologia
1	121484872		Intergênica	TTTTCTGCCATTGACCTTAAAGCGC (SEQ ID NO: 191)	19	Nenhum PAM, repetição
	118041		Intergênica	ATCTGTGGGATTATGACTGAAC (SEQ ID NO: 192)	14	Nenhum PAM, homologia
6	58778779		Intergênica	CTTCTCATAAAACCTAGACAG (SEQ ID NO: 193)	13	Nenhum PAM, homologia
11	65429541	RELA	Exônica	ATGTGGAGATCATTGAGCA (SEQ ID NO: 194)	12	Nenhum PAM, homologia
12	19817224		Intergênica	ATTAATATGGTATCATGGGAGCAGGAC (SEQ ID NO: 195)	9	Nenhuma homologia

As duas entradas sem um mapa listado no cromossomo para GL000220.1, um contig não colocado de 161 kb.

Tabela 11

hAlb T11 (Os dez principais locais potenciais de clivagem fora do alvo nas células HepG2)						
Cromossomo	**Posição**	**Gene**	**Tipo**	**Sequência Alvo Prevista**	**Leituras**	**Comentário**
4	74270447	ALB	Exônica	TTAAATAAAGCATAGTGCAA (SEQ ID NO: 128)	20997	Local alvo
2	230732566		Intrônica	TTAAAATAAAGCATAGTGCAA (SEQ ID NO: 196)	4918	Protuberância
	131491		Intergênica		100	Nenhum PAM, nenhuma homologia
1	121485377		Intergênica	TTCCAACGAAGGCCTCAA (SEQ ID NO: 197)	86	Nenhum PAM, homologia
4	83871456	LIN54	Intrônica	TTACTATAAAGCATAGTGCAA (SEQ ID NO: 198)	86	1 mm, protuberância
4	158301206		Intergênica	AAAAAAAAAGAAAAGAAAAGAAA (SEQ ID NO: 199)	37	Nenhum PAM, homologia
MT	14042	ND5	Exônica	TTACCTAAAACAATTTCACA (SEQ ID NO: 200)	32	Nenhum PAM, homologia
9	13017306		Intergênica	TTAATACTGGGCCCTGAAGCCAAATACAGTT (SEQ ID NO: 201)	28	Nenhum PAM, homologia
	117905	RNA45SN4	Exônica	GGCAACAACACATCATCAGTAGGGTAA (SEQ ID NO: 202)	26	Nenhuma homologia
14	105307026		Intergênica	TTCAGAAATAGAAAAGCTGATCCTCAA (SEQ ID NO: 203)	21	Nenhum PAM, homologia
1	121484870		Intergênica	TTAAAGCGCTTGAAATCTACACTTGCAA (SEQ ID NO: 204)	19	Nenhuma homologia

As duas entradas sem um mapa listado no cromossomo para GL000220.1, um contig não colocado de 161 kb.

Tabela 12

hAlb T13 (Os dez principais locais potenciais de clivagem fora do alvo nas células HepG2)						
Cro-mos-somo	**Posição**	**Gene**	**Tipo**	**Sequência Alvo Prevista**	**Leitu-ras**	**Comen-tário**
4	74270295	ALB	Exôni-ca	TAATAAAATTCAAACA TCCT (SEQ ID NO: 153)	6620	Local alvo
1	37943291	Z3H12A	Intrô-nica	TGATAGGATGTGTGT GTAGAAGACTCC (SEQ ID NO: 205)	206	Nenhum a homo-logia
11	5276345		Inter-gênica	CCATAGAAGATACCA GGACTTCTT (SEQ ID NO: 206)	36	Nenhum PAM, homo-logia
	124450		Inter-gênica		29	Nenhum a homo-logia
14	102377003	PPP2R 5C	Intrô-nica		11	Nenhum a homo-logia
19	917722	KISS1R	Exônica	ATGTAGAAGTTGGTC ACGGTCCGCATCGGC T (SEQ ID NO: 207)	10	Nenhum PAM, homolo gia
3	80519749		Intergênic a	AAATAGAATACCTCAG CATTTCT (SEQ ID NO: 208)	5	Nenhum PAM, homo-logia
6	113169934		Intergênic a	AGATGAAAATCTATCA ATGGCACCAGCGCCT (SEQ ID NO: 209)	5	Nenhum PAM, homo-logia
7	98360198		Intergênic a	TAAAAAAGGGCTGAG CATAGTGGCTCACAC CT (SEQ ID NO: 210)	5	Nenhum PAM, homo-logia
1	121485138		Intergênic a	TATTCAACTCACAGAG TTGAACGATCCT (SEQ ID NO: 211)	4	Nenhum PAM, homol-ogia
1	121485228		Intergênic a		3	Nenhum a homo-logia

A entrada sem um mapa listado no cromossomo para GL000220.1, um contig não colocado de 161 kb.

[00624] Os candidatos a medicamentos terapêuticos são frequentemente avaliados em primatas não humanos, a fim de prever sua potência e segurança para uso humano. No caso da edição de

genes usando o sistema CRISPR-Cas9, a especificidade da sequência do RNA guia determina que a mesma sequência alvo deve estar presente em humanos e no primata não humano, a fim de testar um guia que será potencialmente usado em humanos. Os guias visando o íntron 1 da albumina humana foram analisados in silico para identificar aqueles que correspondiam à sequência genômica correspondente nos macacos Cynomolgus (ver Tabela 4). No entanto, a capacidade desses guias para cortar o genoma de primatas não humanos e a eficiência relativa com a qual eles cortam no local previsto no alvo precisa de ser determinada em um sistema celular relevante. Os hepatócitos primários de macacos Cynomolgus (obtidos de BioIVT, Westbury, NI) foram transfectados com os RNAs guia de albumina T4, T5, T11 ou T13 e mRNA de spCas9 usando o mesmo protocolo experimental descrito acima para hepatócitos humanos primários. A frequência de INDELS foi então determinada usando o mesmo protocolo TIDES descrito acima, mas usando iniciadores de PCR específicos para o íntron 1 da albumina de Cynomolgus. Os resultados estão resumidos na FIG. 7. Os dados correspondentes para o RNA T4 Guia em hepatócitos primários humanos são mostrados na mesma figura para comparação. Todos os 4 guias promoveram a clivagem no local esperado no íntron 1 da albumina nos hepatócitos de Cynomolgus de dois doadores animais diferentes em frequências que variavam de 10% a 25%. A ordem de classificação da eficiência de corte foi T5>T4>T11=T13. O RNA guia T5 foi o mais potente dos 4 guias e cortou 20% e 25% dos alelos alvo nos 2 doadores. A eficiência de corte foi inferior que os guias correspondentes nas células humanas, o que pode ser devido a diferenças na eficiência da transfecção. Em alternativa, estas guias e/ou a enzima spCas9 podem ser inerentemente menos potentes nas células de primatas. No entanto, a constatação de que o T5 foi o mais potente dos 4 guias, juntamente com o seu perfil favorável fora do alvo pelo

GUIDEseq, torna o T5 atraente para testes em NHP e também em humanos.

Exemplo 9: INTEGRAÇÃO DIRECIONADA DE UM GENE REPÓRTER SEAP DOADOR NO ÍNTRON 1 DA ALBUMINA DE CAMUNDONGO MEDIADA POR CRISPR/CAS9 RESULTA NA EXPRESSÃO DE SEAP E NA SECREÇÃO NO SANGUE

[00625] Para avaliar o potencial de usar a clivagem de sequência específica por CRISPR/Cas9 para mediar a integração de uma sequência molde doadora que codifica um gene de interesse na quebra de fita dupla criada pelo complexo Cas9/gRNA, projetamos e construímos um molde doador que codifica o gene repórter de fosfatase alcalina segregada de murino (mSEAP). O gene mSEAP não é imunogênico em camundongos, permitindo que a expressão da proteína mSEAP codificada seja monitorizada sem interferência de uma resposta imune à proteína. Além disso, mSEAP é prontamente secretada no sangue quando um peptídeo sinal apropriado é incluído na extremidade 5' da sequência de codificação e a proteína é facilmente detectável usando um ensaio que mede a atividade da proteína. Um construto mSEAP para empacotamento no Vírus Adeno-Associado (AAV) foi projetado como mostrado na FIG. 8 para integração direcionada no íntron 1 da albumina de camundongo através da clivagem com spCas9 e o RNA guia mALbT1 (tgccagttcccgatcgttacagg, SEQ ID NO: 80). A sequência de codificação de mSEAP da qual o peptídeo sinal foi removido foi otimizada quanto a códon para camundongo e precedida por dois pares de bases (TG) necessários para manter o quadro de leitura correto após o splicing do éxon 1 da albumina endógena do camundongo. Um aceitador de splice consistindo na sequência do aceitador de splice de consenso e um trato de polipirimidina (CTGACCTCTTCTCTTCCTCCCACAG, SEQ ID NO: 2) foi adicionado na extremidade 5' da sequência de codificação e um sinal de

poliadenilação (sPA) foi adicionado na extremidade 3' da sequência de codificação (AATAAAAGATCTTTATTTTCATTAGATCTGTGTGTTGGTTTTTTGTGTG, SEQ ID NO: 5). O complemento reverso do local alvo para o RNA guia mAlbT1 presente no genoma (TGCCAGTTCCCGATCGTTACAGG, SEQ ID NO: 80) foi incluído em ambos os lados deste cassete. A nossa hipótese foi de que, adicionando locais de corte para o RNA guia, o genoma do AAV deve ser clivado *in vivo* dentro do núcleo das células para as quais foi entregue, gerando assim fragmentos de DNA lineares que são moldes ideais para integração em uma quebra da fita dupla pela via de junção de extremidade não homóloga (NHEJ). Para permitir o empacotamento eficiente nos capsídeos do AAV, foi adicionado um fragmento de enchimento derivado da sequência de microssatélites humanos para atingir o tamanho total, incluindo a ITR de 4596 pb. Se esse cassete doador se integrar na orientação direta à quebra de fita dupla no íntron 1 da albumina criado pelo complexo de RNA guia Cas9/mALbT1, é previsto que a transcrição do promotor de albumina gere um transcrito primário que possa sofrer splicing do doador de splice do éxon 1 da albumina ao aceitador de splice de consenso e gerar um mRNA maduro no qual o éxon 1 da albumina é fundido enquadrado com a sequência de codificação de mSEAP. A tradução deste mRNA irá produzir uma proteína mSEAP precedida pelo peptídeo sinal da albumina de camundongo (que é codificada no éxon 1 da albumina). O peptídeo sinal irá direcionar a secreção de mSEAP para a circulação e será clivado no processo de secreção, deixando a proteína mSEAP madura. Uma vez que o éxon 1 da albumina de camundongo codifica o peptídeo sinal e o pró-peptídeo seguidos por 7 pb que codificam o terminal N da proteína madura da albumina (codificando Glu-Ala mais 1 pb (C)), após a clivagem do pró-peptídeo é previsto que a proteína SEAP contenha 3

aminoácidos adicionais no terminal N, ou seja, Glu-Ala-Leu (Leu é gerado pela última base C do éxon 1 da albumina que é unida a TG a partir do cassete do gene SEAP integrado). Optamos por codificar a Leucina (Leu) como o 3º dos 3 aminoácidos adicionais adicionados no terminal N porque a leucina é não carregada e não polar e, portanto, é improvável que interfira na função da proteína SEAP. Este cassete dador SEAP, designado pCB0047, foi empacotado no capsídeo do serotipo AAV8 usando um sistema de transfecção baseado em HEK293 e métodos padrão para purificação de vírus (Vector Biolabs Inc). O vírus foi titulado usando PCR quantitativo com iniciadores e sonda localizados na sequência de codificação mSEAP.

[00626] O vírus pCB0047 foi injetado na veia da cauda dos camundongos no dia 0, na dose de 2e12 vg/kg, seguida 4 dias depois por uma nanopartícula lipídica (LNP) que encapsula o RNA guia mALbT1 (sequência de RNA Guia 5' TGCCAGTTCCCGATCGTTACAGG 3', PAM sublinhado, SEQ ID NO: 80) e mRNA de spCas9. O RNA guia único foi quimicamente sintetizado e incorporou bases quimicamente modificadas essencialmente como descrito (Hendel et al, Nat Biotechnol. 2015 33 (9): 985–989) e usou uma sequência de RNA tracr padrão. O mRNA de spCas9 foi sintetizado usando técnicas padrão e incluiu sequências nucleotídicas que adicionam um sinal de localização nuclear no terminal N e no terminal C da proteína. O sinal de localização nuclear é necessário para direcionar a proteína spCas9 para o núcleo após o mRNA ter sido entregue ao citoplasma das células de interesse pela LNP e depois traduzido para a proteína spCas9. O uso de sequências NLS para direcionar proteínas Cas9 para o núcleo é bem conhecido na técnica, por exemplo, ver Jinek et al (eLife 2013; 2:e00471. DOI: 10.7554/eLife.00471). O mRNA de spCas9 também continha uma cauda de poliA e foi revestido na extremidade 5' para melhorar a

estabilidade e a eficiência da tradução. Para empacotar o gRNA e mRNA de Cas9 na LNP, usamos um protocolo essencialmente como descrito por Kaufmann et al (Nano Lett. 15(11):7300-6) para montar a LNP com base no lipídio ionizável C12-200 (adquirido da AxoLabs). Os outros componentes da LNP são o ácido cis-4,7,10,13,16,19-Docosa-hexaenoico (DHA, adquirido da Sigma), 1,2-dilinoleoil-sn-glicero-3-fosfocolina (DLPC, adquirido da Avanti), 1,2-dimiristoil-sn-glicero-3-fosfoetanolamina-N-[metóxi(polietilenoglicol)-2000] (DMPE-mPEG200, adquirido de Avanti) e Colestrol (adquirido da Avanti). A LNP foi produzida usando o instrumento Nanoassembler Benchtop (Precision Nanosystems), no qual a LNP se monta quando os componentes lipídio e ácido nucleico são misturados sob condições controladas em uma câmara microfluídica. O mRNA de spCas9 e o RNA guia foram encapsulados em LNP separadas. A LNP foi concentrada por diálise em solução salina tamponada com fosfato e armazenada a 4 °C durante até 1 semana antes do uso. A LNP foi caracterizada usando dispersão de luz dinâmica e tipicamente tinha um tamanho na gama de 50 a 60 nM. A concentração de RNA na LNP foi medida usando o kit de teste Ribogreen (Thermofisher Scientific) e usada para determinar a dose administrada aos camundongos. Para dosagem de camundongos, a LNP com spCas9 e RNA guia foram misturadas na razão de massa de 1:1 de RNA imediatamente antes da injeção. A capacidade desta LNP de entregar o mRNA de spCas9 e o RNA guia no fígado de camundongos foi demonstrada pela injeção IV de camundongos com uma gama de doses de LNP e pela medição da clivagem do genoma do camundongo no local alvo no íntron 1 da albumina no fígado usando o procedimento TIDES (Brinkman et al, Nucleic Acids Res. 2014 Dec 16; 42(22): e168). Ver o Exemplo 2 (FIG. 4) para um resultado típico em que até 25% dos alelos foram clivados no local alvo.

[00627] Duas coortes de 5 camundongos foram injetados na veia da

cauda com 2e12 vg/kg de vírus AAV8-CB0047. Três dias depois, uma das coortes foi injetada com LNP encapsulando mRNA de spCas9 e RNA guia mAlbT1 a uma dose total de 2 mg/kg de RNA (razão 1:1 de spCas9 e gRNA). As amostras de sangue foram coletadas semanalmente e o plasma foi analisado quanto à atividade de SEAP usando um kit comercial (InvivoGen). Os resultados (*ver* a Tabela 13) demonstram que nenhuma atividade SEAP foi detectável nos camundongos que receberam apenas o vírus AAV8-pCB0047. Os camundongos que receberam o vírus AAV8-pCB0047 seguido pela LNP tiveram atividade SEAP no plasma que permaneceu estável até o último ponto temporal, 4 semanas após a administração. A descoberta de que a SEAP foi expressa apenas quando os camundongos receberam o gene SEAP doador de AAV8 e os componentes de edição do gene CRISPR-Cas9 sugerem que a proteína SEAP estava sendo expressa a partir de cópias do gene SEAP integradas no local alvo no íntron 1 da albumina. Como o gene SEAP no pCB047 carece de um peptídeo sinal ou de um promotor, ele não pode ser expresso e segregado, a menos que esteja operacionalmente ligado a um promotor e a um peptídeo sinal que está grelha com a sequência de codificação de SEAP. É improvável que isto aconteça se o cassete do gene do pCB047 for integrado em um local aleatório no genoma.

[00628] Para confirmar que o cassete do gene SEAP do pCB0047 foi integrado no íntron 1 da albumina, usamos o Droplet Digital PCR (DD-PCR) para medir a frequência de integração no DNA genômico extraído dos fígados dos camundongos no final do estudo. DD-PCR é um método para quantificar com precisão o número de cópias de uma sequência de ácido nucleico em uma mistura complexa. Um par de iniciadores de PCR foi projetado com um localizado na sequência genômica da albumina do camundongo no lado 5' do local alvo para o guia mAlbT1 (o local previsto para a integração direcionada) e o outro iniciador

localizado na extremidade 5' do gene SEAP em pCB0047. Este PCR "in-out" amplificará a junção entre a sequência genômica da albumina do camundongo e o cassete SEAP integrado quando o cassete SEAP estiver integrado na orientação direta desejada. Foi projetada uma sonda fluorescente que hibrida com a sequência de DNA amplificada por esses 2 iniciadores. Como controle interno para o ensaio de DD-PCR foi usado um conjunto de sonda iniciador que detecta o gene da albumina de camundongo. Usando este ensaio de DD-PCR medimos uma frequência de integração direcionada de 0,24 +/- 0,07% (0,24 cópias por 100 cópias do gene da albumina), confirmando assim que o cassete SEAP foi integrado no íntron 1 da albumina.

Tabela 13: Atividade de SEAP no plasma de camundongos injetados apenas com o vírus AAV8 com pCB0047 ou seguido 3 dias depois com LNP que encapsula o mRNA de spCas9 e o RNA guia mAlbT1

Tempo após AAV	**Atividade de SEAP (micro U/mL de plasma)**	
	Apenas AAV-SEAP	**AAV-SEAP e LNP no dia 3**
Semana 1	-19,9 ± 8,1	2264 ± 766
Semana 2	-31,2 ± 14,6	3470 ± 1480
Semana 3	-423,8 ± 7,6	4575 ± 1737
Semana 4	-415,2 ± 8,1	2913 ± 1614

Exemplo 10: Integração direcionada de um gene doador FVIII humano no íntron 1 da albumina de camundongo mediada por CRISPR/Cas9 resulta na expressão de FVIII no sangue

[00629] A Hemofilia A é uma doença amplamente estudada (Coppola et al., J Blood Med. 2010; 1: 183–195), na qual os pacientes apresentam mutações no gene do Fator VIII que resultam em baixos níveis de proteína funcional do Fator VIII no sangue. O Fator VIII é um componente crítico da cascata de coagulação e, na ausência de quantidades suficientes de FVIII, o sangue não forma um coágulo

estável nos locais da lesão, resultando em sangramento excessivo. Os pacientes de Hemofilia A que não são tratados eficazmente experimentam sangramento nas articulações, resultando em destruição das articulações. Sangramento intracraniano também pode ocorrer e às vezes pode ser fatal.

[00630] Para avaliar se esta estratégia de edição de genes poderia ser usada para tratar a Hemofilia A, foi usado um modelo de camundongo no qual o gene FVIII do camundongo está inativado. Estes camundongos de Hemofilia A não têm FVIII detectável no sangue, o que torna possível medir o FVIII fornecido exogenamente usando um ensaio de atividade da FVIII (Diapharma, Chromogenix Coatest SP Factor FVIII, cat# K824086kit). Como padrões neste ensaio usamos o Kogenate (Bayer), um FVIII humano recombinante usado no tratamento de pacientes com hemofilia. Os resultados do ensaio são relatados como porcentagem da atividade normal de FVIII humano que é definida como 1 IU/mL. Um molde doador de FVIII humano foi construído com base em uma sequência de codificação de FVIII com deleção do domínio B que demonstrou funcionar quando administrada a camundongos com um vetor AAV sob o controle de um forte promotor específico do fígado (McIntosh et al, 2013; Blood; 121(17): 3335-3344). A sequência de DNA que codifica o peptídeo sinal nativo foi removida dessa sequência de codificação de FVIII e substituída por dois pares de bases (TG) necessários para manter o quadro de leitura correto após o splicing do éxon 1 da albumina de camundongo. Uma sequência aceitadora de splice derivada do íntron 1 da albumina de camundongo foi inserida imediatamente a 5' desta sequência de codificação de FVIII. Uma sequência não traduzida a 3' do gene da globina humana seguida por uma sequência sinal de poliadenilação sintética foi inserida no lado 3' da sequência de codificação de FVIII. O sinal de poliadenilação sintético é uma sequência curta de 49 pb mostrada direcionar

efetivamente a poliadenilação (Levitt et al, 1989; GENES & DEVELOPMENT 3:1019-1025). A sequência UTR 3' foi retirada do gene da globina B e pode funcionar para melhorar ainda mais a eficiência da poliadenilação. O complemento reverso dos locais alvo para o RNA guia mAlbT1 foi colocado em qualquer dos locais deste cassete do gene FVIII para criar um vetor chamado pCB056 contendo as sequências ITR de AAV2, como mostrado na FIG. 9. Este plasmídeo foi empacotado em capsídeos AAV8 para gerar o vírus AAV8-pCB056.

[00631] Uma coorte de 5 camundongos de hemofilia A (Grupo 2; G2) foi injetada na veia da cauda com o vírus AAV8-pCB056 na dose de 1 e13 vg/kg e 19 dias depois os mesmos camundongos foram injetados na veia da cauda com uma mistura de duas LNP à base de C12-200 encapsulando mRNA de spCas9 e RNA guia mAlbT1, cada um com uma dose de 1 mg de RNA/kg. A LNP foi formulada como descrito no Exemplo 2 acima. Uma coorte separada de 5 camundongos de hemofilia A (Grupo 6; G6) foi injetada na veia da cauda com o vírus AAV8-pCB056 na dose de 1 e13 vg/kg e a atividade do FVIII foi monitorizada nas 4 semanas seguintes. Quando apenas o AAV foi injetado, nenhuma atividade de FVIII era mensurável no sangue dos camundongos (G6 na FIG. 9). Os camundongos que receberam o vírus AAV8-pCB056 seguido pelos componentes de edição gênica CRISPR/Cas9 em uma LNP tinham atividade de FVIII no sangue que variava de 25% a 60% dos níveis humanos normais de atividade de FVIII. Pacientes com Hemofilia grave têm níveis de atividade do FVIII inferiores a 1% do normal, pacientes com Hemofilia A moderada têm níveis de atividade do FVIII entre 1 e 5% do normal e pacientes com Hemofilia A leve têm níveis entre 6% e 30% do normal. Uma análise de pacientes com Hemofilia A recebendo a terapia de proteína de reposição com FVIII relatou que nos níveis previstos de FVIII de 3%, 5%, 10%, 15% e 20%, a frequência na qual não ocorreram sangramentos foi de 71%, 79%,

91%, 97% e 100%, respectivamente (Spotts et al Blood 2014 124:689), sugerindo que quando os níveis de FVIII são mantidos acima de um nível mínimo de 15 a 20%, a taxa de eventos hemorrágicos foi reduzida para quase zero. Embora um nível preciso de FVIII necessário para curar a Hemofilia A não tenha sido definido e provavelmente varie entre os pacientes, é provável que níveis entre 5% e 30% proporcionem uma redução significativa nos eventos hemorrágicos. Assim, no modelo de camundongo de Hemofilia A descrito acima, os níveis de FVIII que foram alcançados (25 a 60%) estão em uma gama terapeuticamente relevante que se espera que seja curativa.

[00632] Quatro dos cinco camundongos na FIG. 10 exibiram níveis estáveis de FVIII (dentro da variabilidade normal do ensaio e da variação na fisiologia do rato) até o final do estudo no dia 36. A atividade do FVIII em um dos camundongos (2-3) caiu para níveis indetectáveis no dia 36 e isso provavelmente ocorreu devido a uma resposta imune contra a proteína humana FVIII que pode ser reconhecida como uma proteína estranha em camundongos (Meeks et al, 2012 Blood 120(12): 2512-2520). A observação de que nenhuma proteína FVIII foi expressa nos camundongos quando apenas o AAV-molde doador de FVIII foi injetado demonstra que a expressão de FVIII necessitava do fornecimento dos componentes de edição gênica CRISPR/Cas9. Como o cassete doador de FVIII não tem um promotor ou um peptídeo sinal é improvável que o FVIII seja produzido pela integração do cassete em locais aleatórios no genoma ou por algum outro mecanismo indefinido. Para confirmar que o cassete doador de FVIII foi integrado no íntron 1 da albumina, usamos PCR "in-out" em um formato DD-PCR. Os fígados inteiros dos camundongos do grupo 2 foram homogeneizados e o DNA genômico foi extraído e analisado por DD-PCR usando um iniciador localizado no gene da albumina do camundongo em uma posição 5' do local de corte para o gRNA mAlbT1 no qual a integração no alvo é

prevista ter ocorrido. O segundo iniciador de PCR foi localizado na extremidade 5' da sequência de codificação de FVIII dentro do cassete pCB056. Uma sonda fluorescente usada para detecção foi projetada para hibridar com uma sequência entre os dois iniciadores de PCR. A PCR usando estes 2 iniciadores amplificará a junção 5' dos eventos de integração nos quais o cassete FVIII foi integrado no local de corte do gRNA de mAlbT1 na orientação direta que seria capaz de expressar a proteína FVIII. Um teste de DD-PCR contra uma região dentro do gene da albumina de camundongo foi usado como controle para medir o número de cópias dos genomas de camundongo no ensaio. Este ensaio detectou entre 0,46 e 1,28 eventos de integração direcionados por 100 genomas haploides de camundongo (média de 1,0). Houve uma correlação entre a frequência de integração direcionada e o pico dos níveis de FVIII consistentes com o FVIII sendo produzido a partir do cassete do gene FVIII integrado. Assumindo que cerca de 70% das células no fígado de camundongo são hepatócitos e que tanto o AAV8 quanto a LNP são primariamente absorvidos por hepatócitos, pode ser estimado que 1,4% (1,0 * (1/0,7)) dos alelos da albumina de hepatócitos continham um cassete FVIII na orientação direta. Estes resultados demonstram que CRISPR/Cas9 pode ser usado para integrar um cassete do gene FVIII projetado de maneira apropriada no íntron 1 da albumina de camundongos resultando na expressão e secreção de níveis terapêuticos da proteína FVIII funcional no sangue. As modalidades de entrega usadas neste estudo, designadamente um vírus AAV que entrega o molde doador FVIII e uma LNP que entrega os componentes CRISPR/Cas9 são potencialmente passíveis de entrega *in vivo* aos pacientes. Como a Cas9 foi entregue como um mRNA que tem uma vida útil curta *in vivo* (na gama de 1 a 3 dias), o complexo de edição de genes CRISPR/Cas9 estará ativo apenas por um curto período de tempo, o que limita o tempo para a ocorrência dos eventos

de clivagem fora do alvo, proporcionando assim um benefício de segurança previsto. Estes dados demonstram que embora CRISPR/Cas9 estivesse ativo por apenas um curto período de tempo isso foi suficiente para induzir a integração direcionada a uma frequência suficiente para produzir níveis terapeuticamente relevantes da atividade de FVIII em camundongos.

Tabela 14: Frequências de integração direcionadas e níveis de FVIII em camundongos HemA do Grupo 2 que foram injetados com AAV8-pCB056 e LNP

ID do camundongo	Integração Direcionada (%)	Atividade de FVIII no dia 36 (% do normal)	Pico da atividade de FVIII (% do normal)
2-1	0,97	27	38
2-2	1,28	43	62
2-3	0,46	0	25
2-4	1,01	19	30
2-5	1,28	32	32
Camundongo HemA sem tratamento	0,00	0	0

Exemplo 11: O momento da dosagem do RNA guia e do mRNA de Cas9 em uma LNP em relação ao AAV doador afeta os níveis de expressão gênica

[00633] Para avaliar se o tempo entre a injeção do AAV com o molde doador e a dosagem da LNP que encapsula o mRNA de Cas9 e o RNA guia teve um impacto no nível de expressão do gene codificado no molde doador, injetamos duas coortes de 5 camundongos cada com AAV8-pCB0047 que codifica mSEAP. Quatro dias após a injeção do AAV, uma coorte de camundongos (grupo 3) foi injetada com LNP à base de C12-200 encapsulando mRNA de spCas9 e gRNA mAlbT1 (1 mg/kg de cada) e a atividade SEAP foi medida no plasma semanalmente durante as 4 semanas seguintes. A atividade de SEAP foi monitorizada na segunda coorte de camundongos durante 4

semanas, durante as quais nenhuma SEAP foi detectada. Aos 28 dias após a injeção do AAV, os camundongos do grupo 4 foram dosados com LNP à base de C12-200 encapsulando mRNA de spCas9 e gRNA mAlbT1 (1 mg/kg de cada) e a atividade de SEAP foi medida no plasma semanalmente durante as 3 semanas seguintes. Os dados da SEAP estão resumidos na Tabela 15. No grupo 3 que recebeu spCas9/gRNA encapsulado em LNP 4 dias após o AAV, a atividade de SEAP foi em média de 3306 microU/mL. No grupo 4 que recebeu spCas9/gRNA encapsulado em LNP 28 dias após o AAV, a atividade de SEAP foi em média 13389 microU/mL, o que é 4 vezes maior do que no grupo 3. Estes dados demonstraram que a dosagem de spCas9/gRNA encapsulado em LNP 28 dias após a LNP resulta em uma expressão 4 vezes maior do gene integrado no genoma do que se o spCas9/gRNA encapsulado em LNP for dosado apenas 4 dias após o AAV-molde doador. Esta expressão melhorada é provavelmente devido a uma maior frequência de integração de cassetes de genes codificados por doadores de comprimento total no íntron 1 da albumina.

Tabela 15: Atividade de SEAP no plasma de camundongos injetados com AAV8-pCB0047 e LNP, 4 dias ou 28 dias depois

Grupo: **Média ± SD**	**3**	**4**
	Atividade de SEAP média no plasma (microU/mL) em 5 camundongos	
	AAV8-SEAP **+ LNP 4 dias depois**	**AAV8-SEAP** **+ LNP 28 dias depois**
Semana 1	2264 ± 766	-19,9 ± 8,2
Semana 2	3470 ± 1480	-31,2 ± 14,6
Semana 3	4575 ± 1737	-423,8 ± 7,6
Semana 4	2913 ± 1614	-415,2 ± 8,1
Semana 5		19856 ± 12732
Semana 6		10657 ± 9642
Semana 7		9680 ± 3678
Média de SEAP após dose de LNP (Média de 3 ou 4 pontos no tempo)	**3306 ± 848**	**13389 ± 4584**

[00634] O impacto da calendarização da dosagem de AAV-doador e Cas9/gRNA encapsulado em LNP também foi avaliado usando o gene do Fator VIII como um exemplo de um gene de relevância terapêutica. Duas coortes de camundongos de hemofilia A foram injetadas com AAV8-pCB056, que codifica um cassete doador de FVIII humano a uma dose de 2e12 vg/kg no dia 0. Uma das coortes foi injetada 4 dias depois com a LNP à base de C12-200 encapsulando mRNA de spCas9 e gRNA mAlbT1 (1 mg/kg cada), enquanto a segunda coorte foi doseada 17 dias depois com a LNP à base de C12-200 encapsulando mRNA de spCas9 e gRNA mAlbT1 (1 mg/kg cada). A dosagem do AAV8-pCB056 foi escalonada de modo que o mesmo lote de LNP encapsulando mRNA de spCas9 e RNA guia fosse usado para ambos os grupos no mesmo dia. A atividade de FVIII no sangue dos camundongos foi medida no dia 10 e no dia 17 após a dose de LNP e os resultados são mostrados na FIG. 11. Os camundongos que receberam LNP 4 dias após o AAV não tinham FVIII detectável no sangue, enquanto todos os 4 camundongos do grupo que foram injetados com LNP 17 dias após o AAV tinham atividade detectável de FVIII que variava de 2% a 30% do normal no dia 17. Estes resultados demonstram que para um AAV doador que codifica FVIII, a dosagem dos componentes CRISPR/Cas9 pelo menos 17 dias após o AAV doador resulta em níveis terapeuticamente relevantes de FVIII, enquanto a dosagem 4 dias após o AAV não levou à expressão de FVIII.

[00635] O processo pelo qual o AAV infecta as células, incluindo as células do fígado, envolve a saída do endossoma, o revestimento do vírus e o transporte do genoma do AAV para o núcleo. No caso do AAV usado nesses estudos em que os genomas de fita simples são empacotados no vírus, os genomas de fita simples passam por um processo de síntese de DNA da segunda fita para formar genomas de DNA de fita dupla. O tempo necessário para a conversão completa de

genomas de fita simples em genomas de fita dupla não está bem estabelecido, mas é considerado um passo limitante da taxa (Ferrari et al 1996; J Virol. 70: 3227–3234). Os genomas lineares de fita dupla tornam-se concatemerizados em formas circulares multiméricas compostas por monômeros unidos cabeça a cauda e cauda a cabeça (Sun et al 2010; Human Gene Therapy 21: 750–762). Como os AAV com moldes doadores usados em nossos estudos não contêm braços de homologia, eles não serão modelos para HDR e, portanto, só podem ser integrados pela via NEHJ. Apenas fragmentos de DNA linear de fita dupla são moldes para integração mediada por NHEJ em uma quebra de fita dupla. Assim, levantamos a hipótese de que a entrega dos componentes CRISPR-Cas9 às células hepáticas logo após o AAV com o doador pode levar a uma baixa frequência de integração, porque a maioria dos genomas de AAV está na forma de uma fita simples e nessas circunstâncias a maioria das quebras de fita dupla no genoma serão reparadas com pequenas inserções e deleções sem integração de um molde doador. A entrega dos componentes de edição gênica CRISPR/Cas9 em um tempo posterior após o AAV-molde doador permite tempo para a formação de genomas de AAV de fita dupla que são moldes para a integração direcionada mediada por NHEJ. No entanto, esperar muito tempo após a entrega do AAV com o doador pode resultar na conversão de formas lineares de fita dupla em formas circulares (concateméricas) que não serão moldes para a integração direcionada mediada por NHEJ. A inclusão de locais de corte para o RNA guia/Cas9 no molde doador resultará na clivagem de formas circulares para gerar formas lineares. Qualquer uma das formas lineares restantes também serão clivadas para libertar fragmentos curtos contendo a sequência AAV ITR. A inclusão de 1 ou 2 locais de corte de RNA guia no AAV com o molde doador irá gerar uma variedade de fragmentos lineares a partir de formas concateméricas do genoma do

AAV. Os tipos de fragmentos lineares irão variar dependendo do número de locais de corte no genoma do AAV e do número de multímeros em cada concatâmero e da sua orientação relativa e, portanto, é difícil de prever. Um único local de gRNA colocado na extremidade 5' do cassete no AAV irá libertar moldes monoméricos de fita dupla dos círculos monoméricos e dos concatâmeros cabeça a cauda (cabeça a cauda significa a extremidade 5' de um genoma de AAV unido à extremidade 3' do próximo genoma do AAV). No entanto, um único local de gRNA na extremidade 5' não irá libertar um molde linear monomérico de fita dupla dos concatâmeros cabeça a cabeça (os concatâmeros cabeça a cabeça consistem na extremidade 5' de um genoma de AAV unido à extremidade 5' do próximo genoma do AAV). Uma possível vantagem do uso de um único local de gRNA na extremidade 5' é que ele irá libertar apenas ITR curta contendo fragmentos de fita dupla dos concatâmeros cabeça a cabeça, mas não dos concatâmeros cabeça a cauda. Com um único local de corte de gRNA na extremidade 5' do genoma de AAV, a ITR permanecerá na extremidade 3' dos cassetes de genes monoméricos lineares e, portanto, será integrada no genoma. Quando o cassete doador no AAV contém dois locais de gRNA (flanqueando o cassete), isto resultará na libertação de modelos monoméricos de fita dupla de todas as formas de DNA de fita dupla e, portanto, poderá libertar mais moldes para integração direcionada, especialmente se uma mistura dos concatâmeros cabeça a cauda e cauda a cabeça está presente. Uma desvantagem potencial de incluir 2 locais alvo de gRNA que flanqueiam o cassete é que isso irá libertar pequenos fragmentos lineares de fita dupla (cerca de 150 pares de bases) que contêm a sequência AAV ITR. Dois destes pequenos fragmentos (cerca de 150 pares de bases) serão gerados para cada cópia do cassete de gene que contém o gene terapêutico de interesse. Espera-se que os fragmentos que contenham ITR curtos também sejam

moldes para a integração direcionada mediada por NHEJ na quebra de fita dupla no genoma e irão, portanto, competir com o fragmento que contém o cassete do gene para integração na quebra de fita dupla no genoma e, assim, reduz a frequência na qual ocorre o evento desejado de integração do cassete de gene terapêutico no genoma da célula hospedeira. Dada a complexidade deste sistema biológico, no qual não são conhecidos muitos parâmetros, como a cinética da formação do concatâmero e a composição molecular dos concatâmeros (conteúdo de concatâmeros cabeça a cauda e cauda a cabeça e o número de unidades monoméricas nos concatâmeros), não é possível prever com certeza se 0, 1 ou 2 locais de corte guia no cassete doador irão alcançar a mais elevada integração direcionada do cassete doador desejado contendo o gene terapêutico ou como isto é afetado pelo momento da entrega dos componentes de edição gênica CRISPR/Cas9. Os nossos dados corroboram que a inclusão de 2 locais de corte de RNA guia leva a uma integração direcionada mensurável em um ambiente em que os componentes da edição gênica CRISPR/Cas9 são entregues por uma LNP encapsulando o mRNA de spCas9 e o RNA guia dosada pelo menos 17 dias após o AAV-cassete doador ser dosado, mas não quando a LNP foi doseada 4 dias após o AAV-cassete doador.

Exemplo 12: Impacto de diferentes sinais de poliadenilação na expressão de FVIII

[00636] Para avaliar o impacto de diferentes sequências de sinal de poliadenilação na expressão de um gene FVIII após integração direcionada no íntron 1 da albumina de camundongo, construímos uma série de plasmídeos mostrados na FIG. 12. Estes plasmídeos foram projetados com um único local alvo para o gRNA de mALbT1 na extremidade 5' que irá resultar na linearização do DNA plasmídico circular in vivo após a administração aos camundongos usando injeção hidrodinâmica (HDI). A HDI é uma técnica estabelecida para a

administração de DNA plasmídico no fígado de camundongos (Budker et al, 1996; Gene Ther., 3, 593–598), na qual o DNA plasmídico nu em solução salina é injetado rapidamente na veia da cauda de camundongos (2 a 3 mL de volume em 5 a 7 segundos).

[00637] Coortes de 6 camundongos de hemofilia A foram injetadas hidrodinamicamente com 25 µg por camundongo de pCB065, pCB076 ou pCB077. Vinte e quatro horas depois, os camundongos foram doseados por injeção retro-orbital com uma LNP à base de C12-200 encapsulando mRNA de spCas9 e gRNA mAlbT1 a uma dose de 1 mg/kg de cada RNA. A atividade de FVIII no sangue dos camundongos foi medida no dia 10 após a dosagem de LNP. No dia 10, os camundongos foram sacrificados, o fígado inteiro foi homogeneizado e o DNA genômico foi extraído do homogenato. A frequência da integração direcionada do cassete doador de FVIII na orientação direta para o íntron 1 da albumina foi quantificada usando PCR quantitativo em tempo real. Neste ensaio de PCR em tempo real, um iniciador foi localizado na sequência genômica do gene da albumina de camundongo 5' do local de integração esperado (o local de corte para o gRNA mAlbT1) e o segundo iniciador de PCR foi localizado na extremidade 5' da sequência de codificação de FVIII no plasmídeo doador. Uma sonda fluorescente foi localizada entre os dois iniciadores. Este ensaio irá detectar especificamente a junção entre o genoma do camundongo e o cassete doador quando a integração ocorreu na orientação direta (na qual o gene FVIII está na mesma orientação que o gene da albumina genômica do camundongo). Fragmentos de DNA sintético compostos pela sequência prevista do fragmento de junção inserida no DNA genômico do fígado de camundongo não tratado foram usados como padrões de número de cópias para calcular as cópias absolutas dos eventos de integração no DNA genômico do fígado. A atividade de FVIII em camundongos nos grupos 2 (injetados com

pCB065), 3 (injetados com pCB076) e 4 (injetados com pCB077) foi de 5,5%, 4,2% e 11,4%, respectivamente. O grupo 4 que foi injetado com pCB077 teve a maior atividade de FVIII. Como a entrega de DNA ao fígado por injeção hidrodinâmica é altamente variável entre os camundongos calculamos a atividade do FVIII dividida pela frequência de integração direcionada, como mostrado na FIG. 13 para cada camundongo individual. Esta razão representa a expressão de FVIII por cópia integrada do gene FVIII e demonstrou expressão superior de pCB077 (grupo 4) em comparação com pCB065 e pCB076. Quando excluímos os camundongos que não expressavam nenhum FVIII, as razões médias de FVIII/TI foram 42, 8 e 57 para pCB065, pCB076 e pCB077, respectivamente. Estes dados indicam que o sinal de poliadenilação aPA+ em pCB077 permite a expressão superior de FVIII em comparação com o sinal de poliadenilação sPA em pCB076. A expressão de FVIII usando o sinal de poliadenilação sPA+ foi similar à do sinal de poliadenilação do hormônio de crescimento bovino (bGH). Existe uma vantagem em usar uma sequência curta do sinal de poliadenilação, como o sPA (49 pb) ou sPA+ (54 pb), em comparação com o poliA da bGH (225 pb) ao entregar o doador usando o vírus AAV, especialmente no caso do gene FVIII cujo tamanho de 4,3 Kb está próximo do limite de empacotamento do AAV (4,4 Kb excluindo o ITR). O sinal de poliadenilação sPA+ difere do sinal de poliadenilação sPA somente pela presença de um espaçador de 5 pb (tcgcg, SEQ ID NO: 212) entre o códon de terminação do gene FVIII e a sequência sinal de poliadenilação sintética (aataaaagatctttattttcattagatctgtgtgttggttttttgtgtg, SEQ ID NO: 5). Embora esta sequência sinal de poliadenilação sintética tenha sido descrita precedentemente (Levitt et al., 1989; Genes Dev. (7):1019-25) e usado por outros em vetores de terapia gênica baseados em AAV (McIntosh et al, 2013; Blood 121:3335-3344), um benefício da inclusão de uma sequência espaçadora não foi explicitamente

demonstrado. Os nossos dados demonstram que a inclusão de um espaçador curto de 5 pb melhorou a expressão de um gene FVIII integrado no íntron 1 da albumina, no qual a transcrição foi dirigida pelo forte promotor de albumina no genoma. É possível que a vantagem do espaçador seja exclusiva para a configuração da integração direcionada em um locus altamente expresso no genoma.

Exemplo 13: Dosagem repetida dos componentes CRISPR/Cas9 usando uma LNP resulta em aumentos incrementais na expressão de um cassete doador entregue por AAV direcionado para o íntron 1 da albumina de camundongo

[00638] No cenário de administração a um paciente de uma terapia gênica baseada na edição de genes na qual um gene terapêutico é integrado no íntron 1 da albumina, seria vantajoso alcançar um nível de expressão gênica que forneça o benefício terapêutico ótimo para o paciente. Por exemplo, na Hemofilia A, o nível mais desejável da proteína FVIII no sangue estaria na gama de 20% a 100% ou 30% a 100% ou 40% a 100%, ou mais preferencialmente 50% a 100%. Os níveis de FVIII que excedem 100% aumentam o risco de eventos trombóticos (Jenkins et al, 2012; Br J Haematol. 157:653-63) e são, portanto, indesejáveis. As terapias gênicas padrão baseadas em AAV que usam um forte promotor para direcionar a expressão do gene terapêutico a partir de cópias episssomais do genoma de AAV não permitem nenhum controle do nível de expressão alcançado porque o vírus AAV pode ser dosado apenas uma vez e os níveis da expressão alcançada varia significativamente entre os pacientes (Rangarajan et al, 2017; N Engl J Med 377:2519-2530). Após a dosagem do paciente com um vírus AAV, eles desenvolvem anticorpos de alto título contra as proteínas do capsídeo do vírus que, com base em modelos pré-clínicos, são esperados impedir a readministração eficaz do vírus (Petry et al, 2008; Gene Ther. 15:54-60). Uma abordagem em que o gene

terapêutico entregue por um vírus AAV é integrado ao genoma em um locus ‘safe harbor', tal como o íntron 1 da albumina, e essa integração direcionada ocorre através da criação de uma quebra de fita dupla no genoma fornece uma oportunidade para controlar o nível de integração direcionada e, portanto, os níveis do produto gênico terapêutico. Após a transdução do fígado por um AAV que encapsula um genoma de AAV contendo um cassete de DNA doador que codifica o gene terapêutico de interesse, o genoma de AAV será mantido episssomalmente dentro do núcleo das células transduzidas. Esses genomas episssômicos de AAV são relativamente estáveis ao longo do tempo e, portanto, fornecem um conjunto de moldes doadores para integração direcionada em quebras de fita dupla criadas por CRISPR/Cas9. O potencial para usar doses repetidas dos componentes CRISPR/Cas9 entregues em uma LNP não imunogênica para induzir aumentos graduais na expressão de uma proteína codificada em um molde doador entregue por AAV foi avaliado usando o AAV8-pCB0047 e o mRNA de spCas9 e o gRNA mALbT1 encapsulados em LNP C12 -200. Uma coorte de 5 camundongos foi injetada na veia da cauda com AAV8-pCB0047 a 2e12 vg/kg e 4 dias depois foram injetados iv com a LNP à base de C12-200 encapsulando mRNA de spCas9 a 1 mg/kg e gRNA mAlbT1 a 1mg/kg. Os níveis de SEAP no sangue foram medidos semanalmente durante as 4 semanas seguintes e em média 3306 microU/mL (Tabela 16). Após a última medição de SEAP na semana 4, os mesmos camundongos foram redosados com LNP C12-200 encapsulando mRNA de spCas9 e gRNA mALbT1 a 1 mg/kg cada. Os níveis de SEAP no sangue foram medidos semanalmente durante as 3 semanas seguintes e em média 6900 microU/mL, 2 vezes maior que os níveis semanais médios após a primeira dose de LNP. Os mesmos 5 camundongos receberam então uma terceira injeção de LNP C12-200 encapsulando mRNA de spCas9 e gRNA mALbT1 a 1 mg/kg cada. Os níveis de SEAP no sangue foram

medidos semanalmente durante as 4 semanas seguintes e em média 13117 microU/mL, 2 vezes maior que os níveis semanais médios após a segunda dose de LNP. Estes dados demonstram que a dosagem repetida de componentes de edição gênica CRISPR/Cas9 compreendendo mRNA de spCas9 e gRNA encapsulado em uma LNP pode resultar em aumentos graduais na expressão gênica de um molde doador entregue por AAV. O fato de o gene SEAP codificado no molde doador é dependente após ligação covalente a um promotor e uma sequência de peptídeo sinal para expressão sugere fortemente que o aumento da expressão é devido ao aumento da integração direcionada no íntron 1 da albumina. Na semana 12, os camundongos foram sacrificados, o fígado inteiro foi homogeneizado e o DNA genômico foi extraído e analisado quanto à integração direcionada no íntron 1 da albumina, usando DD-PCR com iniciadores flanqueando a junção 5' prevista na orientação direta (a orientação necessária para produzir proteína SEAP funcional). A frequência de integração foi em média de 0,3% (0,3 cópias por 100 alelos de albumina).

Tabela 16: Atividade de SEAP no sangue de camundongos injetados com AAV8-pCB0047, seguido por LNP C12-200 encapsulando mRNA de spCas9 e gRNA de mAlbT1 (1 mg/kg cada) 4 dias, 4 semanas e 7 semanas após o AAV

Tempo após AAV dosagem	**Média da atividade de SEAP no plasma (microU/mL) em 5 camundongos**
	AAV8-SEAP + LNP aos 4 dias, 4 semanas e 7 semanas
Semana 1	2264 ± 765,6
Semana 2	3470 ± 1480
Semana 3	4575 ± 1737
Semana 4	2913 ± 1614

Média da atividade de SEAP após 1 dose de LNP	**3306 ± 848**
Semana 5	9817 ± 4322
Semana 6	6042 ± 2858
Semana 7	4840 ± 2355
Média da atividade de SEAP após 2 doses de LNP	**6900 ± 2120**
Semana 8	12066 ± 3460
Semana 9	12886 ± 9014
Semana 10	15333 ± 4678
Semana 11	12181 ± 2986
Média da atividade de SEAP após 3 doses de LNP	**13117 ± 1318**

Exemplo 14: Integração direcionada de uma FVIII ou SEAP doadora no íntron 1 da albumina em hepatócitos humanos primários mediada por CRISPR/Cas9 resulta na expressão de FVIII ou SEAP

[00639] Para demonstrar que o conceito de integração direcionada de um cassete genético no íntron 1 da albumina mediada pela clivagem CRISPR/Cas9 também funciona em células humanas usando um RNA guia específico para o genoma humano, realizamos experiências em hepatócitos humanos primários. Hepatócitos humanos primários são hepatócitos humanos coletados dos fígados de doadores humanos que foram submetidos a manipulação *in vitro* mínima para manter seu fenótipo normal. Dois moldes doadores foram construídos como mostrado na FIG. 14 e foram empacotados no sorotipo AAV-DJ (Grimm et al., 2008; J Virol. 82: 5887–5911) que é particularmente eficaz na transdução de hepatócitos *in vitro*. Os vírus AAV-DJ foram titulados por PCR quantitativo usando iniciadores e sondas localizadas na sequência de codificação do gene relevante (FVIII ou mSEAP), resultando em um título expresso como cópias do genoma (GC) por mL.

[00640] Os hepatócitos humanos primários (obtidos da BioIVT,

Westbury, NI) foram descongelados, transferidos para o Meio de Recuperação de Hepatócitos (CHRM) (Gibco), peletizados a baixa velocidade e depois plaqueados em Meio *InVitro*GRO™ CP (BioIVT) mais Mistura de Antibióticos *Torpedo*™ (BioIVT) a uma densidade de 0,7 x 10^6 células/mL em placas de 24 poços pré-revestidas com Colágeno IV (Corning). As placas foram incubadas em 5% de CO2 a 37°C. Depois das células aderirem (3-4 horas após o plaqueamento), as células mortas que não aderiram à placa foram lavadas e meio completo quente e fresco foi adicionado às células. Foram preparadas misturas de transfecção à base de lipídios do mRNA de spCas9 (produzido na Trilink) e do RNA guia hAlb T4 (produzido na Synthego Corp, Menlo Park, CA) por adição do RNA ao meio OptiMem (Gibco) na concentração final de 0,02 µg/µL de mRNA e 0,2 µMolar de guia. A isto foi adicionado um volume igual de Lipofectamina diluída 30 vezes em Optimem e incubada à temperatura ambiente durante 20 minutos. O AAV-DJ-pCB0107 ou o AAV-DJ-pCB0156 foi adicionado aos poços relevantes em várias multiplicidades de infecção, variando de 1.000 GC por célula a 100.000 GC por célula, seguido imediatamente (no período de 5 minutos) com a mistura de transecção de lipídios mRNA de spCas9 / gRNA. As placas foram então incubadas em 5% de CO_2 a 37 °C durante 72 h, após as quais o meio foi coletado e analisado quanto à atividade de FVIII usando um ensaio cromogênico (Diapharma, Chromogenix Coatest SP Factor FVIII, cat# K824086kit) ou atividade de SEAP usando um kit comercial (InvivoGen). Os resultados estão resumidos nas FIGS. 15 e 16. Controles nos quais as células foram transfectadas apenas com o mRNA de spCas9 e o gRNA ou o vírus SEAP isoladamente ou o vírus FVIII isoladamente apresentaram um nível baixo de atividade SEAP representando a atividade de base nas células. Quando o vírus AAV-DJ-pCB0107 e o mRNA de Cas9/gRNA hAlbT4 foram transfectados, a atividade SEAP estava significativamente

acima dos níveis de base na MOI mais alta de 50.000 e 100.000. Estes dados indicam que a combinação de componentes de edição gênica CRISPR/Cas9 e um doador entregue por AAV contendo locais de corte para o mesmo gRNA pode resultar na expressão do transgene codificado pelo doador. Como o gene SEAP codificado no doador AAV não possui um promotor ou peptídeo sinal e porque a expressão de SEAP requer os componentes de edição gênica, é provável que o SEAP tenha sido expresso a partir de cópias do doador integradas ao íntron 1 da albumina humana. A PCR "in-out" é um método que pode ser usado para confirmar a integração do doador SEAP no íntron 1 da albumina humana.

[00641] Os controles nos quais as células foram transfectadas com MOI de 100.000 dos vírus AAV-DJ-pCB0107 ou AAV-DJ-pCB0156 isoladamente (sem mRNA de Cas9 ou gRNA) exibiram níveis baixos ou indetectáveis de atividade do FVIII nos meios às 72 h (FIG. 16). As células transfectadas com o vírus AAV-DJ-pCB0156 a várias MOI, juntamente com o mRNA de spCas9 e o gRNA hAlbT4, apresentaram níveis mensuráveis de atividade do FVIII nos meios às 72 h, que variaram de 0,2 a 0,6 mIU/mL. Estes dados indicam que a combinação de componentes de edição gênica CRISPR/Cas9 e um doador entregue por AAV contendo locais de corte para o mesmo gRNA pode resultar na expressão do transgene de FVIII codificado pelo doador. Como o gene FVIII codificado no doador AAV não possui um promotor ou peptídeo sinal e porque a expressão de FVIII requer os componentes de edição gênica, é provável que o FVIII tenha sido expresso a partir de cópias do doador integradas ao íntron 1 da albumina humana. A PCR "in-out" é um método que pode ser usado para confirmar a integração do doador FVIII no íntron 1 da albumina humana.

[00642] Embora a presente invenção tenha sido descrita detalhadamente e com alguma particularidade em relação às várias

modalidades descritas, não se pretende que ela deva ser limitada a essas particularidades ou modalidades ou qualquer modalidade específica, mas deve ser interpretada com referências às reivindicações anexas, a fim de fornecer a interpretação mais ampla possível de tais reivindicações na perspectiva da técnica precedente e, portanto, abranger efetivamente o escopo pretendido da invenção.

LISTAGEM DE SEQUÊNCIAS

[00643] Além das sequências descritas em outras partes da presente invenção, as seguintes sequências são fornecidas à medida que são mencionadas ou usadas em várias modalidades exemplificativas das invenções, que são fornecidas com a finalidade de ilustração.

SEQ ID NO	**Sequência**	**Descrição**
1	SFSQNPPVLKRHQR	
2	ctgacctcttctcttcctcccacag	aceitador de splice sintético
3	TTAACAATCCTTTTTTTTCTTCCCTTGCCCAG	aceitador de splice íntron 1/éxon 2 da albumina nativa, humano
4	ttaaatatgttgtgtggtttttctctccctgtttccacag	aceitador de splice íntron 1/éxon 2 da albumina nativa, camundongo
5	AATAAAAGATCTTTATTTTCATTAGATCTGTGTGTTGGTTTTTTGTGTG	sequência sinal poliA sintética consenso
6	LAGLIDADG	
7	TGCCTTTACCCCATCGTTAC	local alvo
8	TGCCTCCTCCCGATAGTTAC	local alvo
9	GGACAGTTCCTGATTGTTAC	local alvo
10	TGCCTTTTCCCGATTGTTAA	local alvo
11	TTATTACGGTCTCATAGGGC	Iniciador MALBF3
12	AGTCTTTCTGTCAATGCACAC	Iniciador MALBR5

13	usgscsCAGUUCCCGAUCGUUACGUUUUAGAgcuaGAAAuagcAAGUUAAAAUAAGGCUAGUCCGUUAUCaacuuGAAAaaguggcaccgagucggugcusususU "A,G,U,C" são nucleotídeos RNA nativos "a,g,u,c" são nucleotídeos 2'-O-metila, e "s" representa uma estrutura principal fosforotioato	gRNA
14	CCCTCCGTTTGTCCTAGCTT	Iniciador direto da albumina
15	TCTACGAGGCAGCACTGTT	Iniciador reverso da albumina
16	AACTGCTTCTCCTCTTGGGAAGT	Iniciador direto de AAVS1
17	CCTCTCCATCCTCTTGCTTTCTTTG	Iniciador reverso de AAVS1
18	TAATTTTCTTTTGCGCACTAAGG	Albumina Humana Íntron-1_T1
19	TAGTGCAATGGATAGGTCTTTGG	Albumina Humana Íntron-1_T2
20	AGTGCAATGGATAGGTCTTTGGG	Albumina Humana Íntron-1_T3
21	TAAAGCATAGTGCAATGGATAGG	Albumina Humana Íntron-1_T4
22	ATTTATGAGATCAACAGCACAGG	Albumina Humana Íntron-1_T5
23	TGATTCCTACAGAAAAACTCAGG	Albumina Humana Íntron-1_T6
24	TGTATTTGTGAAGTCTTACAAGG	Albumina Humana Íntron-1_T7
25	GACTGAAACTTCACAGAATAGGG	Albumina Humana Íntron-1_T8
26	AATGCATAATCTAAGTCAAATGG	Albumina Humana Íntron-1_T9
27	TGACTGAAACTTCACAGAATAGG	Albumina Humana Íntron-1_T10
28	TTAAATAAAGCATAGTGCAATGG	Albumina Humana Íntron-1_T11
29	GATCAACAGCACAGGTTTTGTGG	Albumina Humana Íntron-1_T12
30	TAATAAAATTCAAACATCCTAGG	Albumina Humana Íntron-1_T13
31	TTCATTTTAGTCTGTCTTCTTGG	Albumina Humana Íntron-1_T14

32	ATTATCTAAGTTTGAATATAAGG	Albumina Humana Íntron-1_T15
33	ATCATCCTGAGTTTTTCTGTAGG	Albumina Humana Íntron-1_T16
34	GCATCTTTAAAGAATTATTTTGG	Albumina Humana Íntron-1_T17
35	TACTAAAACTTTATTTTACTGGG	Albumina Humana Íntron-1_T18
36	TGAATTATTCTTCTGTTTAAAGG	Albumina Humana Íntron-1_T19
37	AATTTTTAAAATAGTATTCTTGG	Albumina Humana Íntron-1_T20
38	ATGCATTTGTTTCAAAATATTGG	Albumina Humana Íntron-1_T21
39	TTTGGCATTTATTTCTAAAATGG	Albumina Humana Íntron-1_T22
40	AAAGTTGAACAATAGAAAAATGG	Albumina Humana Íntron-1_T23
41	TTACTAAAACTTTATTTTACTGG	Albumina Humana Íntron-1_T24
42	TGCATTTGTTTCAAAATATTGGG	Albumina Humana Íntron-1_T26
43	TGGGCAAGGGAAGAAAAAAAAGG	Albumina Humana Íntron-1_T27
44	TCCTAGGTAAAAAAAAAAAAAGG	Albumina Humana Íntron-1_T28
45	TAATTTTCTTTTGCCCACTAAGG	
46	TAGTGCAATGGATAGGTCTTAGG	
47	AGTGCAATGGATAGGTCTTAGGG	
48	TAAAGCATAGTGCAATGGATAGG	
49	ATTTATGAGATCAACAGCACAGG	
50	TGATTCCTACAGAAAAAGTCAGG	
51	AATGCATAATCTAAGTCAAATGG	
52	TTAAATAAAGCATAGTGCAATGG	
53	ATTTATGAGATCAACAGCACAGG	
54	TAATAAAATTCAAACATCCTAGG	
55	ATTATCCTGACTTTTTCTGTAGG	
56	TACTAAAACTTTATTTTACTTGG	
57	TGAATTATTCCTCTGTTTAAAGG	
58	ATGCATTTGTTTCAAAATATTGG	
59	TTTGGCATTTATTTCTAAAATGG	
60	AAAGTTGAACAATAGAAAAATGG	
61	TGCATTTGTTTCAAAATATTGGG	

62	TAATTTTCTTTTGCCCACTAAGG	
63	TAGTGCAATGGATAGGTCTTAGG	
64	AGTGCAATGGATAGGTCTTAGGG	
65	TAAAGCATAGTGCAATGGATAGG	
66	ATTTATGAGATCAACAGCACAGG	
67	TGATTCCTACAGAAAAAGTCAGG	
68	AATGCATAATCTAAGTCAAATGG	
69	TTAAATAAAGCATAGTGCAATGG	
70	ATTTATGAGATCAACAGCACAGG	
71	TAATAAAATTCAAACATCCTAGG	
72	ATTATCCTGACTTTTTCTGTAGG	
73	TACTAAAACTTTATTTTACTTGG	
74	TGAATTATTCCTCTGTTTAAAGG	
75	ATGCATTTGTTTCAAAATATTGG	
76	TTTGGCATTTATTTCTAAAATGG	
77	AAAGTTGAACAATAGAAAAATGG	
78	TGCATTTGTTTCAAAATATTGGG	
79	TGGGGAAGGGGAGAAAAAAAAGG	
80	tgccagttcccgatcgttacagg	Sequências de gRNA do íntron 1 da albumina de camundongo, mALbgRNA_T1
81	GGAAATAAGAGAGAAAAGAAGAGTAAGAAGA AATATAAGAGCCACCATGGCCCCAAAGAAGAA GCGGAAGGTCGGTATCCACGGAGTCCCAGCA GCCGACAAGAAGTACAGCATCGGCCTGGACA TCGGCACCAACTCTGTGGGCTGGGCCGTGAT CACCGACGAGTACAAGGTGCCCAGCAAGAAA TTCAAGGTGCTGGGCAACACCGACCGGCACA GCATCAAGAAGAACCTGATCGGAGCCCTGCT GTTCGACAGCGGCGAAACAGCCGAGGCCACC CGGCTGAAGAGAACCGCCAGAAGAAGATACA CCAGACGGAAGAACCGGATCTGCTATCTGCA AGAGATCTTCAGCAACGAGATGGCCAAGGTG GACGACAGCTTCTTCCACAGACTGGAAGAGT CCTTCCTGGTGGAAGAGGACAAGAAGCACGA GAGACACCCCATCTTCGGCAACATCGTGGAC GAGGTGGCCTACCACGAGAAGTACCCCACCA TCTACCACCTGAGAAAGAAACTGGTGGACAG CACCGACAAGGCCGACCTGAGACTGATCTAC CTGGCCCTGGCCCACATGATCAAGTTCAGAG GCCACTTCCTGATCGAGGGCGACCTGAACCC CGACAACAGCGACGTGGACAAGCTGTTCATC CAGCTGGTGCAGACCTACAACCAGCTGTTCG AGGAAAACCCCATCAACGCCAGCGGCGTGGA CGCCAAGGCTATCCTGTCTGCCAGACTGAGC	mRNA spCas9

	AAGAGCAGAAGGCTGGAAAATCTGATCGCCC AGCTGCCCGGCGAGAAGAAGAACGGCCTGTT CGGCAACCTGATTGCCCTGAGCCTGGGCCTG ACCCCCAACTTCAAGAGCAACTTCGACCTGG CCGAGGATGCCAAACTGCAGCTGAGCAAGGA CACCTACGACGACGACCTGGACAACCTGCTG GCCCAGATCGGCGACCAGTACGCCGACCTGT TCCTGGCCGCCAAGAACCTGTCTGACGCCAT CCTGCTGAGCGACATCCTGAGAGTGAACACC GAGATCACCAAGGCCCCCCTGAGCGCCTCTA TGATCAAGAGATACGACGAGCACCACCAGGA CCTGACCCTGCTGAAAGCTCTCGTGCGGCAG CAGCTGCCTGAGAAGTACAAAGAAATCTTCTT CGACCAGAGCAAGAACGGCTACGCCGGCTAC ATCGATGGCGGCGCTAGCCAGGAAGAGTTCT ACAAGTTCATCAAGCCCATCCTGGAAAAGATG GACGGCACCGAGGAACTGCTCGTGAAGCTGA ACAGAGAGGACCTGCTGAGAAAGCAGAGAAC CTTCGACAACGGCAGCATCCCCCACCAGATC CACCTGGGAGAGCTGCACGCTATCCTGAGAA GGCAGGAAGATTTTTACCCATTCCTGAAGGAC AACCGGGAAAAGATCGAGAAGATCCTGACCT TCAGGATCCCCTACTACGTGGGCCCCCTGGC CAGAGGCAACAGCAGATTCGCCTGGATGACC AGAAAGAGCGAGGAAACCATCACCCCCTGGA ACTTCGAGGAAGTGGTGGACAAGGGCGCCAG CGCCCAGAGCTTCATCGAGAGAATGACAAAC TTCGATAAGAACCTGCCCAACGAGAAGGTGC TGCCCAAGCACAGCCTGCTGTACGAGTACTT CACCGTGTACAACGAGCTGACCAAAGTGAAAT ACGTGACCGAGGGAATGAGAAAGCCCGCCTT CCTGAGCGGCGAGCAGAAAAAGGCCATCGTG GACCTGCTGTTCAAGACCAACAGAAAAGTGAC CGTGAAGCAGCTGAAAGAGGACTACTTCAAG AAAATCGAGTGCTTCGACTCCGTGGAAATCTC CGGCGTGGAAGATAGATTCAACGCCTCCCTG GGCACATACCACGATCTGCTGAAAATTATCAA GGACAAGGACTTCCTGGATAACGAAGAGAAC GAGGACATTCTGGAAGATATCGTGCTGACCCT GACACTGTTTGAGGACCGCGAGATGATCGAG GAAAGGCTGAAAACCTACGCTCACCTGTTCGA CGACAAAGTGATGAAGCAGCTGAAGAGAAGG CGGTACACCGGCTGGGGCAGGCTGAGCAGA AAGCTGATCAACGGCATCAGAGACAAGCAGA GCGGCAAGACAATCCTGGATTTCCTGAAGTC CGACGGCTTCGCCAACCGGAACTTCATGCAG CTGATCCACGACGACAGCCTGACATTCAAAGA GGACATCCAGAAAGCCCAGGTGTCCGGCCAG GGCGACTCTCTGCACGAGCATATCGCTAACC TGGCCGGCAGCCCCGCTATCAAGAAGGGCAT CCTGCAGACAGTGAAGGTGGTGGACGAGCTC GTGAAAGTGATGGGCAGACACAAGCCCGAGA ACATCGTGATCGAGATGGCTAGAGAGAACCA GACCACCCAGAAGGGACAGAAGAACTCCCGC	

	GAGAGGATGAAGAGAATCGAAGAGGGCATCA AAGAGCTGGGCAGCCAGATCCTGAAAGAACA CCCCGTGGAAAACACCCAGCTGCAGAACGAG AAGCTGTACCTGTACTACCTGCAGAATGGCC GGGATATGTACGTGGACCAGGAACTGGACAT CAACAGACTGTCCGACTACGATGTGGACCATA TCGTGCCTCAGAGCTTTCTGAAGGACGACTC CATCGATAACAAAGTGCTGACTCGGAGCGAC AAGAACAGAGGCAAGAGCGACAACGTGCCCT CCGAAGAGGTCGTGAAGAAGATGAAGAACTA CTGGCGACAGCTGCTGAACGCCAAGCTGATT ACCCAGAGGAAGTTCGATAACCTGACCAAGG CCGAGAGAGGCGGCCTGAGCGAGCTGGATA AGGCCGGCTTCATCAAGAGGCAGCTGGTGGA AACCAGACAGATCACAAAGCACGTGGCACAG ATCCTGGACTCCCGGATGAACACTAAGTACGA CGAAAACGATAAGCTGATCCGGGAAGTGAAA GTGATCACCCTGAAGTCCAAGCTGGTGTCCG ATTTCCGGAAGGATTTCCAGTTTTACAAAGTG CGCGAGATCAACAACTACCACCACGCCCACG ACGCCTACCTGAACGCCGTCGTGGGAACCGC CCTGATCAAAAAGTACCCTAAGCTGGAAAGCG AGTTCGTGTACGGCGACTACAAGGTGTACGA CGTGCGGAAGATGATCGCCAAGAGCGAGCAG GAAATCGGCAAGGCTACCGCCAAGTACTTCTT CTACAGCAACATCATGAACTTTTTCAAGACCG AAATCACCCTGGCCAACGGCGAGATCAGAAA GCGCCCTCTGATCGAGACAAACGGCGAAACC GGGGAGATCGTGTGGGATAAGGGCAGAGACT TCGCCACAGTGCGAAAGGTGCTGAGCATGCC CCAAGTGAATATCGTGAAAAAGACCGAGGTG CAGACAGGCGGCTTCAGCAAAGAGTCTATCC TGCCCAAGAGGAACAGCGACAAGCTGATCGC CAGAAAGAAGGACTGGGACCCCAAGAAGTAC GGCGGCTTCGACAGCCCTACCGTGGCCTACT CTGTGCTGGTGGTGGCTAAGGTGGAAAAGGG CAAGTCCAAGAAACTGAAGAGTGTGAAAGAG CTGCTGGGGATCACCATCATGGAAAGAAGCA GCTTTGAGAAGAACCCTATCGACTTTCTGGAA GCCAAGGGCTACAAAGAAGTGAAAAAGGACC TGATCATCAAGCTGCCTAAGTACTCCCTGTTC GAGCTGGAAAACGGCAGAAAGAGAATGCTGG CCTCTGCCGGCGAACTGCAGAAGGGAAACGA GCTGGCCCTGCCTAGCAAATATGTGAACTTCC TGTACCTGGCCTCCCACTATGAGAAGCTGAA GGGCAGCCCTGAGGACAACGAACAGAAACAG CTGTTTGTGGAACAGCATAAGCACTACCTGGA CGAGATCATCGAGCAGATCAGCGAGTTCTCC AAGAGAGTGATCCTGGCCGACGCCAATCTGG ACAAGGTGCTGTCTGCCTACAACAAGCACAG GGACAAGCCTATCAGAGAGCAGGCCGAGAAT ATCATCCACCTGTTCACCCTGACAAACCTGGG CGCTCCTGCCGCCTTCAAGTACTTTGACACCA CCATCGACCGGAAGAGGTACACCAGCACCAA	

	AGAGGTGCTGGACGCCACCCTGATCCACCAG AGCATCACCGGCCTGTACGAGACAAGAATCG ACCTGTCTCAGCTGGGAGGCGACAAGAGACC TGCCGCCACTAAGAAGGCCGGACAGGCCAAA AAGAAGAAGTGAGCGGCCGCTTAATTAAGCT GCCTTCTGCGGGGCTTGCCTTCTGGCCATGC CCTTCTTCTCTCCCTTGCACCTGTACCTCTTG GTCTTTGAATAAAGCCTGAGTAGGAAGAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAA	
82	ctgacctcttctcttcctcccacag	Aceitador de splice sintético
83	ctttaaatatgttgtgtggtttttctctccctgtttccacag	Aceitador de splice Íntron 1 da albumina de camundongo
84	cctcatactgaggtttttgtgtctgcttttcag	Aceitador de splice Íntron 2 da albumina de camundongo
85	ttaacaatccttttttttcttcccttgcccag	Aceitador de splice Íntron 1 da albumina humana
86	attatactacatttttctacatcctttgtttcag	Aceitador de splice Íntron 2 da albumina humana
87	AATTGCTGACCTCTTCTCTTCCTCCCACAGTG GCCACCAGAAGATACTACCTCGGAGCCGTCG AATTGAGCTGGGATTACATGCAATCCGACCTG GGAGAACTGCCCGTGGATGCCAGGTTTCCTC CTCGGGTCCCCAAGTCCTTCCCGTTCAACAC CTCAGTCGTCTACAAGAAAACCCTCTTCGTGG AGTTCACCGACCATCTGTTCAACATCGCCAAG CCAAGACCCCCGTGGATGGGACTCCTCGGTC CGACCATCCAAGCCGAAGTGTACGACACTGT GGTCATTACCCTGAAGAACATGGCCTCCCATC CTGTGTCCCTGCATGCAGTGGGCGTGTCCTA CTGGAAGGCTTCCGAAGGGGCCGAGTACGAC GATCAAACCAGCCAGCGGGAAAAGGAGGATG ACAAAGTGTTCCCGGGTGGTTCGCACACCTA CGTGTGGCAAGTGCTCAAGGAGAACGGTCCT ATGGCCTCTGATCCCCTGTGTCTGACCTACTC CTACCTGTCCCATGTCGACCTCGTGAAGGATC TGAACAGCGGGCTGATTGGCGCCCTGCTCGT GTGCCGGGAAGGCTCCCTGGCCAAGGAAAAG ACCCAGACACTGCACAAGTTCATCTTGCTGTT CGCCGTGTTTGATGAGGGAAAGTCCTGGCAT AGCGAGACTAAGAACTCCCTTATGCAAGACC GGGATGCTGCCTCCGCTAGGGCTTGGCCTAA GATGCATACTGTGAACGGATACGTGAACAGAT CCCTGCCTGGCCTTATCGGTTGCCACCGGAA GTCCGTGTATTGGCATGTGATCGGCATGGGA ACCACTCCAGAGGTGCACTCCATTTTCTTGGA	MAB8A

	GGGGCATACCTTCTTGGTGCGCAACCACAGA CAGGCCTCCCTGGAAATTTCTCCGATCACTTT CCTGACTGCCCAGACCCTCCTTATGGACCTG GGTCAGTTCCTGCTGTTCTGCCACATTTCGTC CCACCAACACGATGGCATGGAAGCCTACGTG AAAGTGGACTCGTGCCCGGAAGAACCACAGC TGCGGATGAAGAACAACGAAGAGGCAGAGGA CTACGATGATGATCTTACCGATTCGGAAATGG ATGTGGTCCGATTCGACGACGATAATAGCCCA TCCTTCATCCAAATTAGGAGCGTGGCCAAGAA GCACCCCAAAACTTGGGTGCATTACATTGCG GCCGAGGAAGAGGATTGGGACTACGCACCCC TCGTGCTTGCACCCGATGATCGGTCCTACAA GTCCCAATACCTGAACAACGGCCCGCAGAGG ATCGGTCGGAAGTATAAGAAAGTGCGCTTCAT GGCCTACACCGACGAGACTTTCAAGACCAGA GAGGCCATTCAGCACGAAAGCGGCATTCTGG GGCCGCTGTTGTACGGGGAGGTCGGAGATAC ACTGCTCATCATTTTCAAGAACCAGGCGTCCA GACCCTACAACATCTACCCGCACGGAATCACT GACGTCCGCCCCCTGTACTCCCGGAGACTCC CGAAGGGAGTCAAGCACTTGAAAGACTTCCC CATCCTGCCTGGGGAAATCTTCAAGTACAAGT GGACCGTGACCGTCGAGGATGGGCCGACCA AGTCCGATCCAAGATGCCTCACTAGATACTAC TCATCCTTCGTCAACATGGAACGGGACCTGG CCTCAGGACTGATTGGCCCCCTGCTCATCTG CTACAAGGAGTCCGTGGATCAGCGCGGAAAC CAGATCATGTCGGACAAACGCAACGTCATCCT CTTCTCCGTCTTTGACGAGAACCGCTCATGGT ACCTTACGGAGAACATCCAGCGGTTCCTCCC CAACCCTGCCGGAGTGCAGCTCGAGGACCCG GAATTCCAGGCATCAAACATTATGCACTCCAT CAACGGTTACGTGTTCGACAGCCTCCAGCTTA GCGTGTGCCTCCATGAAGTCGCATATTGGTAC ATCCTGTCCATTGGAGCACAAACCGACTTTCT CTCCGTGTTCTTCTCCGGATATACCTTCAAGC ACAAGATGGTGTACGAGGATACCCTGACCCT CTTCCCCTTCTCCGGAGAGACTGTGTTTATGT CGATGGAAAACCCAGGCCTGTGGATTTTGGG GTGCCACAACTCGGATTTCCGAAACCGGGGC ATGACTGCCTTGCTCAAGGTGTCCTCCTGTGA CAAGAACACGGGAGACTACTACGAGGACTCC TACGAGGATATTTCCGCCTACCTCCTGTCCAA GAACAACGCCATCGAACCCAGGTCCTTCAGC CAGAACCCTCCTGTCCTCAAGCGCCATCAGA GAGAAATCACCCGCACGACCCTGCAGTCCGA CCAGGAAGAGATCGATTACGACGACACTATCT CCGTCGAAATGAAGAAGGAGGACTTTGACAT CTACGACGAAGATGAAAATCAGTCCCCTCGCT CGTTCCAAAAGAAAACGAGACACTACTTCATC GCTGCTGTGGAGCGGCTCTGGGACTACGGCA TGTCCTCATCGCCCCACGTGCTTAGGAACCG GGCTCAATCCGGGAGCGTCCCTCAGTTCAAG	

	AAAGTGGTGTTTCAAGAATTCACCGATGGAAG CTTCACGCAGCCGTTGTACAGGGGCGAACTG AACGAGCACCTTGGCCTGCTGGGACCTTACA TCAGAGCAGAGGTCGAGGACAACATCATGGT GACCTTCCGGAACCAAGCCTCCCGGCCATAT TCATTCTACTCGAGCCTTATCTCATACGAGGA GGATCAGAGACAGGGGGCTGAACCTCGGAAG AACTTCGTCAAGCCGAACGAGACAAAGACCTA CTTTTGGAAGGTGCAGCACCACATGGCCCCG ACCAAGGATGAGTTCGACTGCAAGGCCTGGG CGTACTTCTCCGACGTGGATCTCGAAAAGGA CGTGCATTCCGGGCTGATCGGACCGCTGCTC GTCTGCCACACTAACACCCTCAATCCTGCTCA CGGCAGACAAGTGACCGTGCAGGAGTTCGCC CTGTTCTTCACCATCTTCGACGAAACTAAGTC ATGGTACTTTACCGAGAACATGGAGCGGAATT GTCGGGCCCCATGTAACATCCAGATGGAGGA CCCGACATTCAAGGAGAACTACCGGTTCCAC GCCATTAACGGATACATTATGGACACTCTTCC GGGACTCGTGATGGCACAGGACCAACGCATC AGATGGTATCTTCTGTCGATGGGGAGCAACG AAAACATCCATTCGATCCACTTTAGCGGTCAC GTGTTCACAGTGCGCAAGAAGGAAGAGTACA AGATGGCGCTGTACAACCTGTACCCTGGGGT GTTCGAGACTGTGGAAATGCTGCCGTCCAAG GCCGGAATTTGGCGCGTGGAATGTCTGATCG GTGAACATCTGCATGCCGGAATGTCCACCCT GTTCCTGGTGTACTCCAACAAGTGCCAAACCC CACTGGGAATGGCATCAGGACACATTAGAGA CTTCCAGATTACCGCGAGCGGACAGTACGGA CAATGGGCCCCCAAGTTGGCCAGGCTGCACT ACTCTGGAAGCATTAACGCCTGGAGCACCAA GGAGCCGTTCAGCTGGATCAAGGTGGACCTT CTGGCGCCAATGATCATCCACGGAATTAAGAC TCAGGGAGCCCGCCAGAAGTTCTCATCGCTC TACATCTCCCAGTTTATCATCATGTACTCACTG GATGGGAAGAAGTGGCAGACTTACCGGGGAA ATTCCACCGGTACTCTGATGGTGTTCTTCGGA AACGTGGACAGCTCCGGCATCAAGCACAATA TCTTTAACCCGCCTATCATCGCCCGATACATC CGGCTCCACCCGACTCACTACTCCATCCGGT CGACTCTGCGGATGGAACTCATGGGTTGCGA CCTCAACTCCTGCTCAATGCCACTGGGCATG GAGTCCAAGGCTATCTCGGACGCTCAGATTA CTGCATCGTCGTACTTTACCAACATGTTCGCT ACCTGGTCCCCGTCCAAAGCCCGGCTGCATC TCCAAGGCAGATCAAACGCGTGGAGGCCTCA GGTCAACAACCCGAAGGAATGGCTTCAGGTC GACTTCCAAAAGACCATGAAAGTCACCGGAGT GACCACCCAGGGCGTGAAATCGCTGCTGACC TCTATGTACGTGAAGGAATTCCTGATCTCATC AAGCCAGGACGGCCACCAGTGGACACTGTTC TTCCAAAATGGAAAGGTCAAGGTCTTTCAGGG AAATCAAGACTCCTTCACCCCCGTGGTGAACT	

	CCCTGGACCCCCCTCTGCTTACCCGCTACTT GCGCATTCATCCGCAATCCTGGGTGCACCAG ATCGCCCTGCGAATGGAAGTGCTGGGCTGTG AAGCGCAGGACCTGTACTAAAATAAAAGATCT TTATTTTCATTAGATCTGTGTGTTGGTTTTTTG TGTGCCGC	
88	AATTGAACTTTGAGTGTAGCAGAGAGGAACCA TTGCCACCTTCAGATTTTAATGTCTGACCTCTT CTCTTCCTCCCACAGTGGCCACCAGAAGATAC TACCTCGGAGCCGTCGAATTGAGCTGGGATT ACATGCAATCCGACCTGGGAGAACTGCCCGT GGATGCCAGGTTTCCTCCTCGGGTCCCCAAG TCCTTCCCGTTCAACACCTCAGTCGTCTACAA GAAAACCCTCTTCGTGGAGTTCACCGACCATC TGTTCAACATCGCCAAGCCAAGACCCCCGTG GATGGGACTCCTCGGTCCGACCATCCAAGCC GAAGTGTACGACACTGTGGTCATTACCCTGAA GAACATGGCCTCCCATCCTGTGTCCCTGCAT GCAGTGGGCGTGTCCTACTGGAAGGCTTCCG AAGGGGCCGAGTACGACGATCAAACCAGCCA GCGGGAAAAGGAGGATGACAAAGTGTTCCCG GGTGGTTCGCACACCTACGTGTGGCAAGTGC TCAAGGAGAACGGTCCTATGGCCTCTGATCC CCTGTGTCTGACCTACTCCTACCTGTCCCATG TCGACCTCGTGAAGGATCTGAACAGCGGGCT GATTGGCGCCCTGCTCGTGTGCCGGGAAGGC TCCCTGGCCAAGGAAAAGACCCAGACACTGC ACAAGTTCATCTTGCTGTTCGCCGTGTTTGAT GAGGGAAAGTCCTGGCATAGCGAGACTAAGA ACTCCCTTATGCAAGACCGGGATGCTGCCTC CGCTAGGGCTTGGCCTAAGATGCATACTGTG AACGGATACGTGAACAGATCCCTGCCTGGCC TTATCGGTTGCCACCGGAAGTCCGTGTATTGG CATGTGATCGGCATGGGAACCACTCCAGAGG TGCACTCCATTTTCTTGGAGGGGCATACCTTC TTGGTGCGCAACCACAGACAGGCCTCCCTGG AAATTTCTCCGATCACTTTCCTGACTGCCCAG ACCCTCCTTATGGACCTGGGTCAGTTCCTGCT GTTCTGCCACATTTCGTCCCACCAACACGATG GCATGGAAGCCTACGTGAAAGTGGACTCGTG CCCGGAAGAACCACAGCTGCGGATGAAGAAC AACGAAGAGGCAGAGGACTACGATGATGATC TTACCGATTCGGAAATGGATGTGGTCCGATTC GACGACGATAATAGCCCATCCTTCATCCAAAT TAGGAGCGTGGCCAAGAAGCACCCCAAAACT TGGGTGCATTACATTGCGGCCGAGGAAGAGG ATTGGGACTACGCACCCCTCGTGCTTGCACC CGATGATCGGTCCTACAAGTCCCAATACCTGA ACAACGGCCCGCAGAGGATCGGTCGGAAGTA TAAGAAAGTGCGCTTCATGGCCTACACCGAC GAGACTTTCAAGACCAGAGAGGCCATTCAGC ACGAAAGCGGCATTCTGGGGCCGCTGTTGTA CGGGGAGGTCGGAGATACACTGCTCATCATT	MAB8B

	TTCAAGAACCAGGCGTCCAGACCCTACAACAT CTACCCGCACGGAATCACTGACGTCCGCCCC CTGTACTCCCGGAGACTCCCGAAGGGAGTCA AGCACTTGAAAGACTTCCCCATCCTGCCTGG GGAAATCTTCAAGTACAAGTGGACCGTGACC GTCGAGGATGGGCCGACCAAGTCCGATCCAA GATGCCTCACTAGATACTACTCATCCTTCGTC AACATGGAACGGGACCTGGCCTCAGGACTGA TTGGCCCCCTGCTCATCTGCTACAAGGAGTC CGTGGATCAGCGCGGAAACCAGATCATGTCG GACAAACGCAACGTCATCCTCTTCTCCGTCTT TGACGAGAACCGCTCATGGTACCTTACGGAG AACATCCAGCGGTTCCTCCCCAACCCTGCCG GAGTGCAGCTCGAGGACCCGGAATTCCAGGC ATCAAACATTATGCACTCCATCAACGGTTACG TGTTCGACAGCCTCCAGCTTAGCGTGTGCCT CCATGAAGTCGCATATTGGTACATCCTGTCCA TTGGAGCACAAACCGACTTTCTCTCCGTGTTC TTCTCCGGATATACCTTCAAGCACAAGATGGT GTACGAGGATACCCTGACCCTCTTCCCCTTCT CCGGAGAGACTGTGTTTATGTCGATGGAAAAC CCAGGCCTGTGGATTTTGGGGTGCCACAACT CGGATTTCCGAAACCGGGGCATGACTGCCTT GCTCAAGGTGTCCTCCTGTGACAAGAACACG GGAGACTACTACGAGGACTCCTACGAGGATA TTTCCGCCTACCTCCTGTCCAAGAACAACGCC ATCGAACCCAGGTCCTTCAGCCAGAACCCTC CTGTCCTCAAGCGCCATCAGAGAGAAATCAC CCGCACGACCCTGCAGTCCGACCAGGAAGAG ATCGATTACGACGACACTATCTCCGTCGAAAT GAAGAAGGAGGACTTTGACATCTACGACGAA GATGAAAATCAGTCCCCTCGCTCGTTCCAAAA GAAAACGAGACACTACTTCATCGCTGCTGTGG AGCGGCTCTGGGACTACGGCATGTCCTCATC GCCCCACGTGCTTAGGAACCGGGCTCAATCC GGGAGCGTCCCTCAGTTCAAGAAAGTGGTGT TTCAAGAATTCACCGATGGAAGCTTCACGCAG CCGTTGTACAGGGGCGAACTGAACGAGCACC TTGGCCTGCTGGGACCTTACATCAGAGCAGA GGTCGAGGACAACATCATGGTGACCTTCCGG AACCAAGCCTCCCGGCCATATTCATTCTACTC GAGCCTTATCTCATACGAGGAGGATCAGAGA CAGGGGGCTGAACCTCGGAAGAACTTCGTCA AGCCGAACGAGACAAAGACCTACTTTTGGAA GGTGCAGCACCACATGGCCCCGACCAAGGAT GAGTTCGACTGCAAGGCCTGGGCGTACTTCT CCGACGTGGATCTCGAAAAGGACGTGCATTC CGGGCTGATCGGACCGCTGCTCGTCTGCCAC ACTAACACCCTCAATCCTGCTCACGGCAGACA AGTGACCGTGCAGGAGTTCGCCCTGTTCTTC ACCATCTTCGACGAAACTAAGTCATGGTACTT TACCGAGAACATGGAGCGGAATTGTCGGGCC CCATGTAACATCCAGATGGAGGACCCGACATT CAAGGAGAACTACCGGTTCCACGCCATTAAC	

	GGATACATTATGGACACTCTTCCGGGACTCGT GATGGCACAGGACCAACGCATCAGATGGTAT CTTCTGTCGATGGGGAGCAACGAAAACATCC ATTCGATCCACTTTAGCGGTCACGTGTTCACA GTGCGCAAGAAGGAAGAGTACAAGATGGCGC TGTACAACCTGTACCCTGGGGTGTTCGAGACT GTGGAAATGCTGCCGTCCAAGGCCGGAATTT GGCGCGTGGAATGTCTGATCGGTGAACATCT GCATGCCGGAATGTCCACCCTGTTCCTGGTG TACTCCAACAAGTGCCAAACCCCACTGGGAAT GGCATCAGGACACATTAGAGACTTCCAGATTA CCGCGAGCGGACAGTACGGACAATGGGCCC CCAAGTTGGCCAGGCTGCACTACTCTGGAAG CATTAACGCCTGGAGCACCAAGGAGCCGTTC AGCTGGATCAAGGTGGACCTTCTGGCGCCAA TGATCATCCACGGAATTAAGACTCAGGGAGC CCGCCAGAAGTTCTCATCGCTCTACATCTCCC AGTTTATCATCATGTACTCACTGGATGGGAAG AAGTGGCAGACTTACCGGGGAAATTCCACCG GTACTCTGATGGTGTTCTTCGGAAACGTGGAC AGCTCCGGCATCAAGCACAATATCTTTAACCC GCCTATCATCGCCCGATACATCCGGCTCCAC CCGACTCACTACTCCATCCGGTCGACTCTGC GGATGGAACTCATGGGTTGCGACCTCAACTC CTGCTCAATGCCACTGGGCATGGAGTCCAAG GCTATCTCGGACGCTCAGATTACTGCATCGTC GTACTTTACCAACATGTTCGCTACCTGGTCCC CGTCCAAAGCCCGGCTGCATCTCCAAGGCAG ATCAAACGCGTGGAGGCCTCAGGTCAACAAC CCGAAGGAATGGCTTCAGGTCGACTTCCAAA AGACCATGAAAGTCACCGGAGTGACCACCCA GGGCGTGAAATCGCTGCTGACCTCTATGTAC GTGAAGGAATTCCTGATCTCATCAAGCCAGGA CGGCCACCAGTGGACACTGTTCTTCCAAAATG GAAAGGTCAAGGTCTTTCAGGGAAATCAAGAC TCCTTCACCCCCGTGGTGAACTCCCTGGACC CCCCTCTGCTTACCCGCTACTTGCGCATTCAT CCGCAATCCTGGGTGCACCAGATCGCCCTGC GAATGGAAGTGCTGGGCTGTGAAGCGCAGGA CCTGTACTAAAATAAAAGATCTTTATTTTCATT AGATCTGTGTGTTGGTTTTTTGTGTGCGATCG GGAACTGGCATCTTCAGGGAGTAGCTTAGGT CAGTGAAGAGAAGCCGC	
89	gcggcctaaggcAATTGTGCCAGTTCCCGATCGTT ACAGGAACTTTGAGTGTAGCAGAGAGGAACC ATTGCCACCTTCAGATTTTAATGTCTGACCTCT TCTCTTCCTCCCACAGTGGCCACCAGAAGATA CTACCTCGGAGCCGTCGAATTGAGCTGGGAT TACATGCAATCCGACCTGGGAGAACTGCCCG TGGATGCCAGGTTTCCTCCTCGGGTCCCCAA GTCCTTCCCGTTCAACACCTCAGTCGTCTACA AGAAAACCCTCTTCGTGGAGTTCACCGACCAT CTGTTCAACATCGCCAAGCCAAGACCCCCGT	MAB8C

	GGATGGGACTCCTCGGTCCGACCATCCAAGC CGAAGTGTACGACACTGTGGTCATTACCCTGA AGAACATGGCCTCCCATCCTGTGTCCCTGCAT GCAGTGGGCGTGTCCTACTGGAAGGCTTCCG AAGGGGCCGAGTACGACGATCAAACCAGCCA GCGGGAAAAGGAGGATGACAAAGTGTTCCCG GGTGGTTCGCACACCTACGTGTGGCAAGTGC TCAAGGAGAACGGTCCTATGGCCTCTGATCC CCTGTGTCTGACCTACTCCTACCTGTCCCATG TCGACCTCGTGAAGGATCTGAACAGCGGGCT GATTGGCGCCCTGCTCGTGTGCCGGGAAGGC TCCCTGGCCAAGGAAAAGACCCAGACACTGC ACAAGTTCATCTTGCTGTTCGCCGTGTTTGAT GAGGGAAAGTCCTGGCATAGCGAGACTAAGA ACTCCCTTATGCAAGACCGGGATGCTGCCTC CGCTAGGGCTTGGCCTAAGATGCATACTGTG AACGGATACGTGAACAGATCCCTGCCTGGCC TTATCGGTTGCCACCGGAAGTCCGTGTATTGG CATGTGATCGGCATGGGAACCACTCCAGAGG TGCACTCCATTTTCTTGGAGGGGCATACCTTC TTGGTGCGCAACCACAGACAGGCCTCCCTGG AAATTTCTCCGATCACTTTCCTGACTGCCCAG ACCCTCCTTATGGACCTGGGTCAGTTCCTGCT GTTCTGCCACATTTCGTCCCACCAACACGATG GCATGGAAGCCTACGTGAAAGTGGACTCGTG CCCGGAAGAACCACAGCTGCGGATGAAGAAC AACGAAGAGGCAGAGGACTACGATGATGATC TTACCGATTCGGAAATGGATGTGGTCCGATTC GACGACGATAATAGCCCATCCTTCATCCAAAT TAGGAGCGTGGCCAAGAAGCACCCCAAAACT TGGGTGCATTACATTGCGGCCGAGGAAGAGG ATTGGGACTACGCACCCCTCGTGCTTGCACC CGATGATCGGTCCTACAAGTCCCAATACCTGA ACAACGGCCCGCAGAGGATCGGTCGGAAGTA TAAGAAAGTGCGCTTCATGGCCTACACCGAC GAGACTTTCAAGACCAGAGAGGCCATTCAGC ACGAAAGCGGCATTCTGGGGCCGCTGTTGTA CGGGGAGGTCGGAGATACACTGCTCATCATT TTCAAGAACCAGGCGTCCAGACCCTACAACAT CTACCCGCACGGAATCACTGACGTCCGCCCC CTGTACTCCCGGAGACTCCCGAAGGGAGTCA AGCACTTGAAAGACTTCCCCATCCTGCCTGG GGAAATCTTCAAGTACAAGTGGACCGTGACC GTCGAGGATGGGCCGACCAAGTCCGATCCAA GATGCCTCACTAGATACTACTCATCCTTCGTC AACATGGAACGGGACCTGGCCTCAGGACTGA TTGGCCCCCTGCTCATCTGCTACAAGGAGTC CGTGGATCAGCGCGGAAACCAGATCATGTCG GACAAACGCAACGTCATCCTCTTCTCCGTCTT TGACGAGAACCGCTCATGGTACCTTACGGAG AACATCCAGCGGTTCCTCCCCAACCCTGCCG GAGTGCAGCTCGAGGACCCGGAATTCCAGGC ATCAAACATTATGCACTCCATCAACGGTTACG TGTTCGACAGCCTCCAGCTTAGCGTGTGCCT	

	CCATGAAGTCGCATATTGGTACATCCTGTCCA TTGGAGCACAAACCGACTTTCTCTCCGTGTTC TTCTCCGGATATACCTTCAAGCACAAGATGGT GTACGAGGATACCCTGACCCTCTTCCCCTTCT CCGGAGAGACTGTGTTTATGTCGATGGAAAAC CCAGGCCTGTGGATTTTGGGGTGCCACAACT CGGATTTCCGAAACCGGGGCATGACTGCCTT GCTCAAGGTGTCCTCCTGTGACAAGAACACG GGAGACTACTACGAGGACTCCTACGAGGATA TTTCCGCCTACCTCCTGTCCAAGAACAACGCC ATCGAACCCAGGTCCTTCAGCCAGAACCCTC CTGTCCTCAAGCGCCATCAGAGAGAAATCAC CCGCACGACCCTGCAGTCCGACCAGGAAGAG ATCGATTACGACGACACTATCTCCGTCGAAAT GAAGAAGGAGGACTTTGACATCTACGACGAA GATGAAAATCAGTCCCCTCGCTCGTTCCAAAA GAAAACGAGACACTACTTCATCGCTGCTGTGG AGCGGCTCTGGGACTACGGCATGTCCTCATC GCCCCACGTGCTTAGGAACCGGGCTCAATCC GGGAGCGTCCCTCAGTTCAAGAAAGTGGTGT TTCAAGAATTCACCGATGGAAGCTTCACGCAG CCGTTGTACAGGGGCGAACTGAACGAGCACC TTGGCCTGCTGGGACCTTACATCAGAGCAGA GGTCGAGGACAACATCATGGTGACCTTCCGG AACCAAGCCTCCCGGCCATATTCATTCTACTC GAGCCTTATCTCATACGAGGAGGATCAGAGA CAGGGGGCTGAACCTCGGAAGAACTTCGTCA AGCCGAACGAGACAAAGACCTACTTTTGGAA GGTGCAGCACCACATGGCCCCGACCAAGGAT GAGTTCGACTGCAAGGCCTGGGCGTACTTCT CCGACGTGGATCTCGAAAAGGACGTGCATTC CGGGCTGATCGGACCGCTGCTCGTCTGCCAC ACTAACACCCTCAATCCTGCTCACGGCAGACA AGTGACCGTGCAGGAGTTCGCCCTGTTCTTC ACCATCTTCGACGAAACTAAGTCATGGTACTT TACCGAGAACATGGAGCGGAATTGTCGGGCC CCATGTAACATCCAGATGGAGGACCCGACATT CAAGGAGAACTACCGGTTCCACGCCATTAAC GGATACATTATGGACACTCTTCCGGGACTCGT GATGGCACAGGACCAACGCATCAGATGGTAT CTTCTGTCGATGGGGAGCAACGAAAACATCC ATTCGATCCACTTTAGCGGTCACGTGTTCACA GTGCGCAAGAAGGAAGAGTACAAGATGGCGC TGTACAACCTGTACCCTGGGGTGTTCGAGACT GTGGAAATGCTGCCGTCCAAGGCCGGAATTT GGCGCGTGGAATGTCTGATCGGTGAACATCT GCATGCCGGAATGTCCACCCTGTTCCTGGTG TACTCCAACAAGTGCCAAACCCCACTGGGAAT GGCATCAGGACACATTAGAGACTTCCAGATTA CCGCGAGCGGACAGTACGGACAATGGGCCC CCAAGTTGGCCAGGCTGCACTACTCTGGAAG CATTAACGCCTGGAGCACCAAGGAGCCGTTC AGCTGGATCAAGGTGGACCTTCTGGCGCCAA TGATCATCCACGGAATTAAGACTCAGGGAGC	

	CCGCCAGAAGTTCTCATCGCTCTACATCTCCC AGTTTATCATCATGTACTCACTGGATGGGAAG AAGTGGCAGACTTACCGGGGAAATTCCACCG GTACTCTGATGGTGTTCTTCGGAAACGTGGAC AGCTCCGGCATCAAGCACAATATCTTTAACCC GCCTATCATCGCCCGATACATCCGGCTCCAC CCGACTCACTACTCCATCCGGTCGACTCTGC GGATGGAACTCATGGGTTGCGACCTCAACTC CTGCTCAATGCCACTGGGCATGGAGTCCAAG GCTATCTCGGACGCTCAGATTACTGCATCGTC GTACTTTACCAACATGTTCGCTACCTGGTCCC CGTCCAAAGCCCGGCTGCATCTCCAAGGCAG ATCAAACGCGTGGAGGCCTCAGGTCAACAAC CCGAAGGAATGGCTTCAGGTCGACTTCCAAA AGACCATGAAAGTCACCGGAGTGACCACCCA GGGCGTGAAATCGCTGCTGACCTCTATGTAC GTGAAGGAATTCCTGATCTCATCAAGCCAGGA CGGCCACCAGTGGACACTGTTCTTCCAAAATG GAAAGGTCAAGGTCTTTCAGGGAAATCAAGAC TCCTTCACCCCCGTGGTGAACTCCCTGGACC CCCCTCTGCTTACCCGCTACTTGCGCATTCAT CCGCAATCCTGGGTGCACCAGATCGCCCTGC GAATGGAAGTGCTGGGCTGTGAAGCGCAGGA CCTGTACTAAAATAAAAGATCTTTATTTTCATT AGATCTGTGTGTTGGTTTTTTGTGTGCGATCG GGAACTGGCATCTTCAGGGAGTAGCTTAGGT CAGTGAAGAGAAGTGCCAGTTCCCGATCGTT ACAGGCCGCgggccgc	
90	gcggcctaaggcAATTGTGCCAGTTCCCGATCGTT ACAGGAACTTTGAGTGTAGCAGAGAGGAACC ATTGCCACCTTCAGATTTTAATGTCTGACCTCT TCTCTTCCTCCCACAGTGGCCACCAGAAGGTA CTACCTAGGAGCCGTGGAACTGAGCTGGGAC TACATGCAGTCTGACCTGGGAGAGCTGCCCG TGGACGCTAGATTTCCTCCAAGAGTGCCCAA GAGCTTCCCCTTCAACACCTCCGTGGTGTACA AGAAAACCCTGTTCGTGGAATTCACCGACCAC CTGTTCAATATCGCCAAGCCTAGACCTCCTTG GATGGGCCTGCTGGGCCCTACAATTCAGGCC GAGGTGTACGACACCGTGGTCATCACCCTGA AGAACATGGCCAGCCATCCTGTGTCTCTGCA CGCCGTGGGAGTGTCTTACTGGAAGGCTTCT GAGGGCGCCGAGTACGACGACCAGACAAGC CAGAGAGAGAAAGAGGACGACAAGGTTTTCC CTGGCGGCAGCCACACCTATGTCTGGCAGGT CCTGAAAGAAAACGGCCCTATGGCCTCCGAT CCTCTGTGCCTGACATACAGCTACCTGAGCCA TGTGGACCTGGTCAAGGACCTGAACTCTGGC CTGATCGGCGCTCTGCTCGTGTGTAGAGAAG GCAGCCTGGCCAAAGAAAAGACCCAGACACT GCACAAGTTCATCCTGCTGTTCGCCGTGTTCG ACGAGGGCAAGAGCTGGCACAGCGAGACAAA GAACAGCCTGATGCAGGACAGAGATGCCGCC	MAB8D

	TCTGCTAGAGCTTGGCCCAAGATGCACACCG TGAACGGCTACGTGAACAGAAGCCTGCCTGG ACTGATCGGATGCCACAGAAAGTCCGTGTACT GGCATGTGATCGGCATGGGCACCACACCTGA GGTGCACAGCATCTTTCTGGAAGGACACACC TTCCTCGTGCGGAACCACAGACAGGCCAGCC TGGAAATCAGCCCTATCACCTTCCTGACCGCT CAGACCCTGCTGATGGATCTGGGCCAGTTTC TGCTGTTCTGCCACATCAGCAGCCACCAGCA CGATGGCATGGAAGCCTACGTGAAGGTGGAC AGCTGCCCCGAAGAACCCCAGCTGAGAATGA AGAACAACGAGGAAGCCGAGGACTACGACGA CGACCTGACCGACTCTGAGATGGACGTCGTC AGATTCGACGACGATAACAGCCCCAGCTTCAT CCAGATCAGAAGCGTGGCCAAGAAGCACCCC AAGACCTGGGTGCACTATATCGCCGCCGAGG AAGAGGACTGGGATTACGCTCCTCTGGTGCT GGCCCCTGACGACAGAAGCTACAAGAGCCAG TACCTGAACAACGGCCCTCAGAGAATCGGCC GGAAGTATAAGAAAGTGCGGTTCATGGCCTA CACCGACGAGACATTCAAGACCAGAGAGGCT ATCCAGCACGAGAGCGGCATTCTGGGACCTC TGCTGTATGGCGAAGTGGGCGACACACTGCT GATCATCTTCAAGAACCAGGCCAGCAGACCC TACAACATCTACCCTCACGGCATCACCGATGT GCGGCCTCTGTACTCTAGAAGGCTGCCCAAG GGCGTGAAGCACCTGAAGGACTTCCCTATCC TGCCTGGCGAGATCTTCAAGTACAAGTGGAC CGTGACCGTCGAGGACGGCCCTACCAAGAGC GATCCTAGATGCCTGACACGGTACTACAGCA GCTTCGTGAACATGGAACGCGACCTGGCCAG CGGCCTGATTGGTCCTCTGCTGATCTGCTACA AAGAAAGCGTGGACCAGAGGGGCAACCAGAT CATGAGCGACAAGAGAAACGTGATCCTGTTCT CCGTCTTTGACGAGAACAGGTCCTGGTATCTG ACCGAGAACATCCAGCGGTTTCTGCCCAATC CTGCTGGCGTGCAGCTGGAAGATCCTGAGTT CCAGGCCTCCAACATCATGCACTCCATCAACG GCTATGTGTTCGACAGCCTGCAGCTGAGCGT GTGCCTGCACGAAGTGGCCTACTGGTACATC CTGTCTATCGGCGCCCAGACCGACTTCCTGT CCGTGTTCTTTAGCGGCTACACCTTCAAGCAC AAGATGGTGTACGAGGATACCCTGACACTGTT CCCATTCAGCGGCGAGACAGTGTTCATGAGC ATGGAAAACCCCGGCCTGTGGATCCTGGGCT GTCACAACAGCGACTTCAGAAACAGAGGCAT GACAGCCCTGCTGAAGGTGTCCAGCTGCGAC AAGAACACCGGCGACTACTACGAGGACTCTT ACGAGGACATCAGCGCCTACCTGCTGAGCAA GAACAATGCCATCGAGCCTCGGAGCTTCTCT CAGAACCCTCCTGTGCTGAAGAGACACCAGC GCGAGATCACCAGAACCACACTGCAGAGCGA CCAAGAGGAAATCGATTACGACGACACCATCA GCGTCGAGATGAAGAAAGAAGATTTCGACATC	

	TACGACGAGGACGAGAATCAGAGCCCCAGAT CTTTCCAGAAGAAAACGCGGCACTACTTCATT GCCGCCGTGGAAAGACTGTGGGACTACGGCA TGAGCAGCAGCCCACATGTGCTGAGAAACAG GGCCCAGAGCGGAAGCGTGCCCCAGTTCAAG AAAGTGGTGTTCCAAGAGTTCACCGACGGCA GCTTCACCCAGCCTCTGTATAGAGGCGAGCT GAACGAGCACCTGGGACTGCTGGGACCTTAC ATCAGAGCTGAGGTCGAGGATAACATCATGG TCACCTTTAGAAACCAGGCCTCTAGGCCCTAC TCCTTCTACAGCTCCCTGATCAGCTACGAAGA GGACCAGAGACAGGGCGCTGAGCCCAGAAA GAACTTCGTGAAGCCCAACGAGACTAAGACC TACTTTTGGAAGGTGCAGCACCACATGGCCC CTACAAAGGACGAGTTCGACTGCAAGGCCTG GGCCTACTTCTCTGACGTGGACCTCGAGAAG GATGTGCACAGCGGACTCATCGGACCCCTGC TTGTGTGCCACACCAACACACTGAATCCCGCT CACGGCAGGCAAGTGACCGTGCAAGAGTTCG CCCTGTTCTTCACCATCTTCGATGAGACAAAG TCCTGGTACTTCACCGAAAACATGGAAAGAAA CTGCAGGGCCCCTTGCAACATCCAGATGGAA GATCCCACCTTCAAAGAGAACTACCGGTTCCA CGCCATCAATGGCTACATCATGGACACTCTGC CCGGCCTGGTTATGGCACAGGATCAGAGGAT CAGATGGTATCTGCTGTCCATGGGCTCCAAC GAGAATATCCACAGCATCCACTTCAGCGGCC ATGTGTTCACCGTGCGGAAAAAAGAAGAGTAC AAGATGGCCCTGTACAATCTGTACCCCGGCG TGTTCGAGACTGTGGAAATGCTGCCTAGCAA GGCCGGAATCTGGCGCGTGGAATGTCTGATC GGAGAGCATCTGCATGCCGGAATGTCTACCC TGTTCCTGGTGTACAGCAACAAGTGTCAGACC CCTCTCGGCATGGCCTCTGGACACATCAGAG ACTTCCAGATCACCGCCTCTGGCCAGTACGG ACAGTGGGCTCCTAAACTGGCTAGACTGCAC TACAGCGGCAGCATCAACGCCTGGTCCACCA AAGAGCCCTTCAGCTGGATCAAGGTGGACCT GCTGGCTCCCATGATCATCCACGGAATCAAG ACCCAGGGCGCCAGACAGAAGTTCAGCAGCC TGTACATCAGCCAGTTCATCATCATGTACAGC CTGGACGGCAAGAAGTGGCAGACCTACAGAG GCAACAGCACCGGCACACTCATGGTGTTCTT CGGCAACGTGGACTCCAGCGGCATTAAGCAC AACATCTTCAACCCTCCAATCATTGCCCGGTA CATCCGGCTGCACCCCACACACTACAGCATC AGATCTACCCTGAGGATGGAACTGATGGGCT GCGACCTGAACAGCTGCTCTATGCCCCTCGG AATGGAAAGCAAGGCCATCAGCGACGCCCAG ATCACAGCCAGCAGCTACTTCACCAACATGTT CGCCACATGGTCCCCATCTAAGGCCCGGCTG CATCTGCAGGGCAGATCTAACGCTTGGAGGC CCCAAGTGAACAACCCCAAAGAGTGGCTGCA GGTCGACTTTCAGAAAACCATGAAAGTGACCG	

	GCGTGACCACACAGGGCGTCAAGTCTCTGCT GACCTCTATGTACGTGAAAGAGTTCCTGATCT CCAGCAGCCAGGACGGCCACCAGTGGACCCT GTTTTTCCAGAACGGCAAAGTCAAGGTGTTCC AGGGAAACCAGGACAGCTTCACACCCGTGGT CAACTCCCTGGATCCTCCACTGCTGACCAGAT ACCTGAGAATTCACCCTCAGTCTTGGGTGCAC CAGATCGCTCTGAGAATGGAAGTGCTGGGAT GTGAAGCTCAGGACCTCTACTAAAATAAAAGA TCTTTATTTTCATTAGATCTGTGTGTTGGTTTT TTGTGTGCGATCGGGAACTGGCATCTTCAGG GAGTAGCTTAGGTCAGTGAAGAGAAGTGCCA GTTCCCGATCGTTACAGGCCGCgggccgc	
91	AATTGCTGACCTCTTCTCTTCCTCCCACAGTG GCCACCAGAAGATACTACCTGGGAGCTGTGG AATTGAGCTGGGATTACATGCAATCTGACCTG GGAGAACTGCCTGTGGATGCCAGGTTTCCTC CTAGGGTCCCCAAGTCCTTCCCATTCAACACC TCAGTGGTCTACAAGAAAACCCTCTTTGTGGA GTTCACAGACCATCTGTTCAACATTGCCAAGC CAAGACCCCCATGGATGGGACTCCTGGGTCC AACCATCCAAGCTGAAGTGTATGACACTGTGG TCATTACCCTGAAGAACATGGCCTCCCATCCT GTGTCCCTGCATGCAGTGGGAGTGTCCTACT GGAAGGCTTCTGAAGGGGCTGAGTATGATGA TCAAACCAGCCAGAGAGAAAAGGAGGATGAC AAAGTGTTCCCAGGTGGTAGTCACACCTATGT GTGGCAAGTGCTCAAGGAGAATGGTCCTATG GCCTCTGATCCCCTGTGTCTGACCTACTCCTA CCTGTCCCATGTGGACCTGGTGAAGGATCTG AACTCTGGGCTGATTGGAGCCCTGCTGGTGT GCAGAGAAGGCTCCCTGGCCAAGGAAAAGAC CCAGACACTGCACAAGTTCATCTTGCTGTTTG CTGTGTTTGATGAGGGAAAGTCCTGGCATTCT GAGACTAAGAACTCCCTTATGCAAGACAGAGA TGCTGCCTCAGCTAGGGCTTGGCCTAAGATG CATACTGTGAATGGATATGTGAACAGATCCCT GCCTGGCCTTATTGGTTGCCACAGGAAGTCT GTGTATTGGCATGTGATTGGCATGGGAACCA CTCCAGAGGTGCACTCCATTTTCTTGGAGGG GCATACCTTCTTGGTGAGGAACCACAGACAG GCCTCCCTGGAAATTTCTCCAATCACTTTCCT GACTGCCCAGACCCTCCTTATGGACCTGGGT CAGTTCCTGCTGTTCTGCCACATTTCATCCCA CCAACATGATGGCATGGAAGCCTATGTGAAA GTGGACTCATGCCCAGAAGAACCACAGCTGA GAATGAAGAACAATGAAGAGGCAGAGGACTA TGATGATGATCTTACAGATTCAGAAATGGATG TGGTCAGATTTGATGATGATAATAGCCCATCC TTCATCCAAATTAGGAGTGTGGCCAAGAAGCA CCCCAAAACTTGGGTGCATTACATTGCAGCTG AGGAAGAGGATTGGGACTATGCACCCTTGGT GCTTGCACCAGATGATAGGTCCTACAAGTCCC	MAB8E

	AATACCTGAACAATGGCCCACAGAGGATTGGT AGAAAGTATAAGAAAGTGAGATTCATGGCCTA CACAGATGAGACTTTCAAGACCAGAGAGGCC ATTCAGCATGAATCTGGCATTCTGGGGCCACT GTTGTATGGGGAGGTTGGAGATACACTGCTC ATCATTTTCAAGAACCAGGCCTCCAGACCCTA CAACATCTACCCTCATGGAATCACTGATGTCA GACCCCTGTACTCCAGAAGACTCCCAAAGGG AGTCAAGCACTTGAAAGACTTCCCCATCCTGC CTGGGGAAATCTTCAAGTACAAGTGGACAGT GACAGTGGAGGATGGGCCAACCAAGTCTGAT CCAAGATGCCTCACTAGATACTACTCATCCTT TGTCAACATGGAAAGAGACCTGGCCTCAGGA CTGATTGGCCCCCTGCTCATCTGCTACAAGGA GTCTGTGGATCAGAGAGGAAACCAGATCATG TCTGACAAAAGGAATGTCATCCTCTTCTCTGT CTTTGATGAGAACAGATCATGGTACCTTACAG AGAACATCCAGAGGTTCCTCCCCAACCCTGCT GGAGTGCAGCTGGAGGACCCAGAATTCCAGG CATCAAACATTATGCACTCCATCAATGGTTAT GTGTTTGACAGCCTCCAGCTTTCTGTGTGCCT CCATGAAGTGGCATATTGGTACATCCTGTCCA TTGGAGCACAAACAGACTTTCTCTCTGTGTTC TTCTCTGGATATACCTTCAAGCACAAGATGGT GTATGAGGATACCCTGACCCTCTTCCCCTTCT CTGGAGAGACTGTGTTTATGTCAATGGAAAAC CCAGGCCTGTGGATTTTGGGGTGCCACAACT CAGATTTCAGAAACAGGGGCATGACTGCCTT GCTCAAGGTGTCCTCCTGTGACAAGAACACA GGAGACTACTATGAGGACTCCTATGAGGATAT TTCTGCCTACCTCCTGTCCAAGAACAATGCCA TTGAACCCAGGTCCTTCAGCCAGAACCCTCCT GTCCTCAAGAGGCATCAGAGAGAAATCACCA GAACTACCCTGCAGTCTGACCAGGAAGAGAT TGATTATGATGACACTATCTCAGTGGAAATGA AGAAGGAGGACTTTGACATCTATGATGAAGAT GAAAATCAGTCCCCTAGGTCCTTCCAAAAGAA AACAAGACACTACTTCATTGCTGCTGTGGAGA GACTCTGGGACTATGGCATGTCCTCATCACCC CATGTGCTTAGGAACAGGGCTCAATCTGGGT CTGTCCCTCAGTTCAAGAAAGTGGTGTTTCAA GAATTCACAGATGGAAGCTTCACACAGCCATT GTACAGGGGAGAACTGAATGAGCACCTTGGC CTGCTGGGACCTTACATCAGAGCAGAGGTGG AGGACAACATCATGGTGACCTTCAGAAACCAA GCCTCCAGGCCATATTCATTCTACTCCAGCCT TATCTCATATGAGGAGGATCAGAGACAGGGG GCTGAACCTAGGAAGAACTTTGTCAAGCCAAA TGAGACAAAGACCTACTTTTGGAAGGTGCAGC ACCACATGGCCCCTACCAAGGATGAGTTTGA CTGCAAGGCCTGGGCTTACTTCTCTGATGTG GATCTGGAAAAGGATGTGCATTCTGGGCTGAT TGGACCTCTGCTGGTCTGCCACACTAACACC CTCAATCCTGCTCATGGCAGACAAGTGACAGT	

	GCAGGAGTTTGCCCTGTTCTTCACCATCTTTG ATGAAACTAAGTCATGGTACTTTACAGAGAAC ATGGAGAGAAATTGTAGGGCCCCATGTAACAT CCAGATGGAGGACCCAACATTCAAGGAGAAC TACAGATTCCATGCCATTAATGGATACATTATG GACACTCTTCCAGGACTGGTGATGGCACAGG ACCAAAGAATCAGATGGTATCTTCTGAGCATG GGGAGCAATGAAAACATCCATTCCATCCACTT TTCAGGTCATGTGTTCACAGTGAGGAAGAAG GAAGAGTACAAGATGGCTCTGTACAACCTGTA CCCTGGGGTGTTTGAGACTGTGGAAATGCTG CCATCCAAGGCTGGAATTTGGAGGGTGGAAT GTCTGATTGGTGAACATCTGCATGCTGGAATG TCCACCCTGTTCCTGGTGTACTCCAACAAGTG CCAAACCCCACTGGGAATGGCATCAGGACAC ATTAGAGACTTCCAGATTACAGCATCTGGACA GTATGGACAATGGGCCCCCAAGTTGGCCAGG CTGCACTACTCTGGAAGCATTAATGCCTGGAG CACCAAGGAGCCATTCAGCTGGATCAAGGTG GACCTTCTGGCTCCAATGATCATCCATGGAAT TAAGACTCAGGGAGCCAGACAGAAGTTCTCAT CCCTCTACATCTCCCAGTTTATCATCATGTACT CACTGGATGGGAAGAAGTGGCAGACTTACAG GGGAAATTCCACAGGTACTCTGATGGTGTTCT TTGGAAATGTGGACAGCTCTGGCATCAAGCA CAATATCTTTAACCCTCCTATCATTGCCAGGTA CATCAGACTCCACCCAACTCACTACTCCATCA GGTCCACTCTGAGGATGGAACTCATGGGTTG TGACCTCAACTCCTGCTCAATGCCACTGGGCA TGGAGTCCAAGGCTATCTCAGATGCTCAGATT ACTGCATCCTCTTACTTTACCAACATGTTTGCT ACCTGGTCCCCCTCCAAAGCCAGACTGCATC TCCAAGGCAGATCAAATGCCTGGAGGCCTCA GGTCAACAACCCAAAGGAATGGCTTCAGGTG GACTTCCAAAAGACCATGAAAGTCACAGGAGT GACCACCCAGGGAGTGAAATCCCTGCTGACC TCTATGTATGTGAAGGAATTCCTGATCTCATC AAGCCAGGATGGCCACCAGTGGACACTGTTC TTCCAAAATGGAAAGGTCAAGGTCTTTCAGGG AAATCAAGACTCCTTCACCCCTGTGGTGAACT CCCTGGACCCCCCTCTGCTTACCAGGTACTT GAGAATTCATCCACAATCCTGGGTGCACCAGA TTGCCCTGAGGATGGAAGTGCTGGGCTGTGA AGCCCAGGACCTGTACTAAAATAAAAGATCTT TATTTTCATTAGATCTGTGTGTTGGTTTTTTGT GTGCCGC	
92	GCTAGCCAATTGctgacctcttctcttcctcccacagtggcca ccagaaggtactacctgggagctgtggaactgagctgggactaca tgcagtctgacctgggagagctgccTgtggatgctagatttcctcca agagtgcccaagagcttccccttcaacacctctgtggtgtacaaga aaaccctgtttgtggaattcacagaccacctgttcaatattgccaag cctagacctccttggatgggcctgctgggccctacaattcaggctga ggtgtatgacacagtggtcatcaccctgaagaacatggccagcca	MAB8F (GA co/CpG livre)

	tcctgtgtctctgcatgctgtgggagtgtcttactggaaggcttctgag ggggctgagtatgatgaccagacaagccagagagagaaagag gatgacaaggttttccctgggggcagccacacctatgtctggcagg tcctgaaagaaaatggccctatggcctctgatcctctgtgcctgaca tacagctacctgagccatgtggacctggtcaaggacctgaactctg gcctgattggggctctgctggtgtgtagagaaggcagcctggccaa agaaaagacccagacactgcacaagttcatcctgctgtttgctgtgt ttgatgagggcaagagctggcactctgagacaaagaacagcctg atgcaggacagagatgctgcctctgctagagcttggcccaagatg cacacagtgaatggctatgtgaacagaagcctgcctggactgattg gatgccacagaaagtctgtgtactggcatgtgattggcatgggcac cacacctgaggtgcacagcatctttctggaaggacacaccttcctg gtgaggaaccacagacaggccagcctggaaatcagccctatca ccttcctgacagctcagaccctgctgatggatctgggccagtttctgc tgttctgccacatcagcagccaccagcatgatggcatggaagcct atgtgaaggtggacagctgccctgaagaaccccagctgagaatg aagaacaatgaggaagctgaggactatgatgatgacctgacaga ctctgagatggatgtggtcagatttgatgatgataacagccccagct tcatccagatcagatctgtggccaagaagcaccccaagacctgg gtgcactatattgctgctgaggaagaggactgggattatgctcctct ggtgctggcccctgatgacagaagctacaagagccagtacctga acaatggccctcagagaattggcaggaagtataagaaagtgagg ttcatggcctacacagatgagacattcaagaccagagaggctatc cagcatgagtctggcattctgggacctctgctgtatggggaagtgg gggacacactgctgatcatcttcaagaaccaggccagcagaccct acaacatctaccctcatggcatcacagatgtgaggcctctgtactct agaaggctgcccaaggggtgaagcacctgaaggacttccctat cctgcctggggagatcttcaagtacaagtggacagtgacagtgga ggatggccctaccaagtctgatcctagatgcctgacaaggtactac agcagctttgtgaacatggaaagggacctggcctctggcctgattg gtcctctgctgatctgctacaaagaatctgtggaccagaggggcaa ccagatcatgagtgacaagagaaatgtgatcctgttctctgtctttga tgagaacaggtcctggtatctgacagagaacatccagaggtttctg cccaatcctgctggggtgcagctggaagatcctgagttccaggcct ccaacatcatgcactccatcaatggctatgtgtttgacagcctgcag ctgtctgtgtgcctgcatgaagtggcctactggtacatcctgtctattg gggcccagacagacttcctgtctgtgttcttttctggctacaccttcaa gcacaagatggtgtatgaggataccctgacactgttcccattctctg gggagacagtgttcatgagcatggaaaaccctggcctgtggatcct gggctgtcacaacagtgacttcagaaacagaggcatgacagccc tgctgaaggtgtccagctgtgacaagaacactggggactactatg aggactcttatgaggacatctctgcctacctgctgagcaagaacaa tgccattgagcctaggagcttctctcagaaccctcctgtgctgaaga gacaccagagggagatcaccagaaccacactgcagtctgacca agaggaaattgattatgatgacaccatctctgtggagatgaagaaa gaagattttgacatctatgatgaggatgagaatcagagccccagat ctttccagaagaaaacaaggcactacttcattgctgctgtggaaag actgtgggactatggcatgagcagcagcccccatgtgctgagaaa cagggcccagtctggaagtgtgccccagttcaagaaagtggtgttc caagagttcacagatggcagcttcacccagcctctgtatagaggg gagctgaatgagcacctgggactgctgggaccttacatcagagct gaggtggaggataacatcatggtcacctttagaaaccaggcctcta ggccctactccttctacagctccctgatcagctatgaagaggacca gagacaggggctgagcccagaaagaactttgtgaagcccaat	

	gagactaagacctacttttggaaggtgcagcaccacatggcccct acaaaggatgagtttgactgcaaggcctgggcctacttctctgatgt ggacctggagaaggatgtgcactctggactcattggacccctgctt gtgtgccacaccaacacactgaatcctgctcatggcaggcaagtg acagtgcaagagtttgccctgttcttcaccatctttgatgagacaaag tcctggtacttcacagaaaacatggaaagaaactgcagggcccct tgcaacatccagatggaagatcccaccttcaaagagaactacag gttccatgccatcaatggctacatcatggacactctgcctggcctggt tatggcacaggatcagaggatcagatggtatctgctgtccatgggc tccaatgagaatatccacagcatccacttctctggccatgtgttcaca gtgaggaaaaaagaagagtacaagatggccctgtacaatctgta ccctggggtgtttgagactgtggaaatgctgcctagcaaggctgga atctggagggtggaatgtctgattggagagcatctgcatgctggaat gtctaccctgttcctggtgtacagcaacaagtgtcagacccctctgg gcatggcctctggacacatcagagacttccagatcacagcctctgg ccagtatggacagtgggctcctaaactggctagactgcactactct ggcagcatcaatgcctggtccaccaaagagcccttcagctggatc aaggtggacctgctggctcccatgatcatccatggaatcaagacc caggggccagacagaagttcagcagcctgtacatcagccagtt catcatcatgtacagcctggatggcaagaagtggcagacctacag aggcaacagcacaggcacactcatggtgttctttggcaatgtggac tcttctggcattaagcacaacatcttcaaccctccaatcattgccagg tacatcaggctgcaccccacacactacagcatcagatctaccctg aggatggaactgatgggctgtgacctgaacagctgctctatgcccc tgggaatggaaagcaaggccatctctgatgcccagatcacagcc agcagctacttcaccaacatgtttgccacatggtccccatctaaggc caggctgcatctgcagggcagatctaatgcttggaggccccaagt gaacaaccccaaagagtggctgcaggtggactttcagaaaacca tgaaagtgacaggagtgaccacacaggggtcaagtctctgctg acctctatgtatgtgaaagagttcctgatctccagcagccaggatgg ccaccagtggaccctgtttttccagaatggcaaagtcaaggtgttcc agggaaaccaggacagcttcacacctgtggtcaactccctggatc ctccactgctgaccagatacctgagaattcaccctcagtcttgggtg caccagattgctctgagaatggaagtgctgggatgtgaagctcag gacctctacTAAAATAAAAGATCTTTATTTTCATTA GATCTGTGTGTTGGTTTTTTGTGTGCCGCGGG TTTAAAC	
93	GCCACCAGAAGATACTACCTCGGAGCCGTCG AATTGAGCTGGGATTACATGCAATCCGACCTG GGAGAACTGCCCGTGGATGCCAGGTTTCCTC CTCGGGTCCCCAAGTCCTTCCCGTTCAACAC CTCAGTCGTCTACAAGAAAACCCTCTTCGTGG AGTTCACCGACCATCTGTTCAACATCGCCAAG CCAAGACCCCCGTGGATGGGACTCCTCGGTC CGACCATCCAAGCCGAAGTGTACGACACTGT GGTCATTACCCTGAAGAACATGGCCTCCCATC CTGTGTCCCTGCATGCAGTGGGCGTGTCCTA CTGGAAGGCTTCCGAAGGGGCCGAGTACGAC GATCAAACCAGCCAGCGGGAAAAGGAGGATG ACAAAGTGTTCCCGGGTGGTTCGCACACCTA CGTGTGGCAAGTGCTCAAGGAGAACGGTCCT ATGGCCTCTGATCCCCTGTGTCTGACCTACTC CTACCTGTCCCATGTCGACCTCGTGAAGGATC	F8-BDD1, otimizado quanto a códon

	TGAACAGCGGGCTGATTGGCGCCCTGCTCGT GTGCCGGGAAGGCTCCCTGGCCAAGGAAAAG ACCCAGACACTGCACAAGTTCATCTTGCTGTT CGCCGTGTTTGATGAGGGAAAGTCCTGGCAT AGCGAGACTAAGAACTCCCTTATGCAAGACC GGGATGCTGCCTCCGCTAGGGCTTGGCCTAA GATGCATACTGTGAACGGATACGTGAACAGAT CCCTGCCTGGCCTTATCGGTTGCCACCGGAA GTCCGTGTATTGGCATGTGATCGGCATGGGA ACCACTCCAGAGGTGCACTCCATTTTCTTGGA GGGGCATACCTTCTTGGTGCGCAACCACAGA CAGGCCTCCCTGGAAATTTCTCCGATCACTTT CCTGACTGCCCAGACCCTCCTTATGGACCTG GGTCAGTTCCTGCTGTTCTGCCACATTTCGTC CCACCAACACGATGGCATGGAAGCCTACGTG AAAGTGGACTCGTGCCCGGAAGAACCACAGC TGCGGATGAAGAACAACGAAGAGGCAGAGGA CTACGATGATGATCTTACCGATTCGGAAATGG ATGTGGTCCGATTCGACGACGATAATAGCCCA TCCTTCATCCAAATTAGGAGCGTGGCCAAGAA GCACCCCAAAACTTGGGTGCATTACATTGCG GCCGAGGAAGAGGATTGGGACTACGCACCCC TCGTGCTTGCACCCGATGATCGGTCCTACAA GTCCCAATACCTGAACAACGGCCCGCAGAGG ATCGGTCGGAAGTATAAGAAAGTGCGCTTCAT GGCCTACACCGACGAGACTTTCAAGACCAGA GAGGCCATTCAGCACGAAAGCGGCATTCTGG GGCCGCTGTTGTACGGGGAGGTCGGAGATAC ACTGCTCATCATTTTCAAGAACCAGGCGTCCA GACCCTACAACATCTACCCGCACGGAATCACT GACGTCCGCCCCCTGTACTCCCGGAGACTCC CGAAGGGAGTCAAGCACTTGAAAGACTTCCC CATCCTGCCTGGGGAAATCTTCAAGTACAAGT GGACCGTGACCGTCGAGGATGGGCCGACCA AGTCCGATCCAAGATGCCTCACTAGATACTAC TCATCCTTCGTCAACATGGAACGGGACCTGG CCTCAGGACTGATTGGCCCCCTGCTCATCTG CTACAAGGAGTCCGTGGATCAGCGCGGAAAC CAGATCATGTCGGACAAACGCAACGTCATCCT CTTCTCCGTCTTTGACGAGAACCGCTCATGGT ACCTTACGGAGAACATCCAGCGGTTCCTCCC CAACCCTGCCGGAGTGCAGCTCGAGGACCCG GAATTCCAGGCATCAAACATTATGCACTCCAT CAACGGTTACGTGTTCGACAGCCTCCAGCTTA GCGTGTGCCTCCATGAAGTCGCATATTGGTAC ATCCTGTCCATTGGAGCACAAACCGACTTTCT CTCCGTGTTCTTCTCCGGATATACCTTCAAGC ACAAGATGGTGTACGAGGATACCCTGACCCT CTTCCCCTTCTCCGGAGAGACTGTGTTTATGT CGATGGAAAACCCAGGCCTGTGGATTTTGGG GTGCCACAACTCGGATTTCCGAAACCGGGGC ATGACTGCCTTGCTCAAGGTGTCCTCCTGTGA CAAGAACACGGGAGACTACTACGAGGACTCC TACGAGGATATTTCCGCCTACCTCCTGTCCAA	

	GAACAACGCCATCGAACCCAGGTCCTTCAGC CAGAACCCTCCTGTCCTCAAGCGCCATCAGA GAGAAATCACCCGCACGACCCTGCAGTCCGA CCAGGAAGAGATCGATTACGACGACACTATCT CCGTCGAAATGAAGAAGGAGGACTTTGACAT CTACGACGAAGATGAAAATCAGTCCCCTCGCT CGTTCCAAAAGAAAACGAGACACTACTTCATC GCTGCTGTGGAGCGGCTCTGGGACTACGGCA TGTCCTCATCGCCCCACGTGCTTAGGAACCG GGCTCAATCCGGGAGCGTCCCTCAGTTCAAG AAAGTGGTGTTTCAAGAATTCACCGATGGAAG CTTCACGCAGCCGTTGTACAGGGGCGAACTG AACGAGCACCTTGGCCTGCTGGGACCTTACA TCAGAGCAGAGGTCGAGGACAACATCATGGT GACCTTCCGGAACCAAGCCTCCCGGCCATAT TCATTCTACTCGAGCCTTATCTCATACGAGGA GGATCAGAGACAGGGGGCTGAACCTCGGAAG AACTTCGTCAAGCCGAACGAGACAAAGACCTA CTTTTGGAAGGTGCAGCACCACATGGCCCCG ACCAAGGATGAGTTCGACTGCAAGGCCTGGG CGTACTTCTCCGACGTGGATCTCGAAAAGGA CGTGCATTCCGGGCTGATCGGACCGCTGCTC GTCTGCCACACTAACACCCTCAATCCTGCTCA CGGCAGACAAGTGACCGTGCAGGAGTTCGCC CTGTTCTTCACCATCTTCGACGAAACTAAGTC ATGGTACTTTACCGAGAACATGGAGCGGAATT GTCGGGCCCCATGTAACATCCAGATGGAGGA CCCGACATTCAAGGAGAACTACCGGTTCCAC GCCATTAACGGATACATTATGGACACTCTTCC GGGACTCGTGATGGCACAGGACCAACGCATC AGATGGTATCTTCTGTCGATGGGGAGCAACG AAAACATCCATTCGATCCACTTTAGCGGTCAC GTGTTCACAGTGCGCAAGAAGGAAGAGTACA AGATGGCGCTGTACAACCTGTACCCTGGGGT GTTCGAGACTGTGGAAATGCTGCCGTCCAAG GCCGGAATTTGGCGCGTGGAATGTCTGATCG GTGAACATCTGCATGCCGGAATGTCCACCCT GTTCCTGGTGTACTCCAACAAGTGCCAAACCC CACTGGGAATGGCATCAGGACACATTAGAGA CTTCCAGATTACCGCGAGCGGACAGTACGGA CAATGGGCCCCCAAGTTGGCCAGGCTGCACT ACTCTGGAAGCATTAACGCCTGGAGCACCAA GGAGCCGTTCAGCTGGATCAAGGTGGACCTT CTGGCGCCAATGATCATCCACGGAATTAAGAC TCAGGGAGCCCGCCAGAAGTTCTCATCGCTC TACATCTCCCAGTTTATCATCATGTACTCACTG GATGGGAAGAAGTGGCAGACTTACCGGGGAA ATTCCACCGGTACTCTGATGGTGTTCTTCGGA AACGTGGACAGCTCCGGCATCAAGCACAATA TCTTTAACCCGCCTATCATCGCCCGATACATC CGGCTCCACCCGACTCACTACTCCATCCGGT CGACTCTGCGGATGGAACTCATGGGTTGCGA CCTCAACTCCTGCTCAATGCCACTGGGCATG GAGTCCAAGGCTATCTCGGACGCTCAGATTA	

	CTGCATCGTCGTACTTTACCAACATGTTCGCT ACCTGGTCCCCGTCCAAAGCCCGGCTGCATC TCCAAGGCAGATCAAACGCGTGGAGGCCTCA GGTCAACAACCCGAAGGAATGGCTTCAGGTC GACTTCCAAAAGACCATGAAAGTCACCGGAGT GACCACCCAGGGCGTGAAATCGCTGCTGACC TCTATGTACGTGAAGGAATTCCTGATCTCATC AAGCCAGGACGGCCACCAGTGGACACTGTTC TTCCAAAATGGAAAGGTCAAGGTCTTTCAGGG AAATCAAGACTCCTTCACCCCCGTGGTGAACT CCCTGGACCCCCCTCTGCTTACCCGCTACTT GCGCATTCATCCGCAATCCTGGGTGCACCAG ATCGCCCTGCGAATGGAAGTGCTGGGCTGTG AAGCGCAGGACCTGTAC	
94	GCCACCAGAAGGTACTACCTAGGAGCCGTGG AACTGAGCTGGGACTACATGCAGTCTGACCT GGGAGAGCTGCCCGTGGACGCTAGATTTCCT CCAAGAGTGCCCAAGAGCTTCCCCTTCAACA CCTCCGTGGTGTACAAGAAAACCCTGTTCGTG GAATTCACCGACCACCTGTTCAATATCGCCAA GCCTAGACCTCCTTGGATGGGCCTGCTGGGC CCTACAATTCAGGCCGAGGTGTACGACACCG TGGTCATCACCCTGAAGAACATGGCCAGCCA TCCTGTGTCTCTGCACGCCGTGGGAGTGTCT TACTGGAAGGCTTCTGAGGGCGCCGAGTACG ACGACCAGACAAGCCAGAGAGAGAAAGAGGA CGACAAGGTTTTCCCTGGCGGCAGCCACACC TATGTCTGGCAGGTCCTGAAAGAAAACGGCC CTATGGCCTCCGATCCTCTGTGCCTGACATAC AGCTACCTGAGCCATGTGGACCTGGTCAAGG ACCTGAACTCTGGCCTGATCGGCGCTCTGCT CGTGTGTAGAGAAGGCAGCCTGGCCAAAGAA AAGACCCAGACACTGCACAAGTTCATCCTGCT GTTCGCCGTGTTCGACGAGGGCAAGAGCTGG CACAGCGAGACAAAGAACAGCCTGATGCAGG ACAGAGATGCCGCCTCTGCTAGAGCTTGGCC CAAGATGCACACCGTGAACGGCTACGTGAAC AGAAGCCTGCCTGGACTGATCGGATGCCACA GAAAGTCCGTGTACTGGCATGTGATCGGCAT GGGCACCACACCTGAGGTGCACAGCATCTTT CTGGAAGGACACACCTTCCTCGTGCGGAACC ACAGACAGGCCAGCCTGGAAATCAGCCCTAT CACCTTCCTGACCGCTCAGACCCTGCTGATG GATCTGGGCCAGTTTCTGCTGTTCTGCCACAT CAGCAGCCACCAGCACGATGGCATGGAAGCC TACGTGAAGGTGGACAGCTGCCCCGAAGAAC CCCAGCTGAGAATGAAGAACAACGAGGAAGC CGAGGACTACGACGACGACCTGACCGACTCT GAGATGGACGTCGTCAGATTCGACGACGATA ACAGCCCCAGCTTCATCCAGATCAGAAGCGT GGCCAAGAAGCACCCCAAGACCTGGGTGCAC TATATCGCCGCCGAGGAAGAGGACTGGGATT ACGCTCCTCTGGTGCTGGCCCCTGACGACAG	F8-BDD2, otimizado quanto a códon

	AAGCTACAAGAGCCAGTACCTGAACAACGGC CCTCAGAGAATCGGCCGGAAGTATAAGAAAG TGCGGTTCATGGCCTACACCGACGAGACATT CAAGACCAGAGAGGCTATCCAGCACGAGAGC GGCATTCTGGGACCTCTGCTGTATGGCGAAG TGGGCGACACACTGCTGATCATCTTCAAGAAC CAGGCCAGCAGACCCTACAACATCTACCCTC ACGGCATCACCGATGTGCGGCCTCTGTACTC TAGAAGGCTGCCCAAGGGCGTGAAGCACCTG AAGGACTTCCCTATCCTGCCTGGCGAGATCTT CAAGTACAAGTGGACCGTGACCGTCGAGGAC GGCCCTACCAAGAGCGATCCTAGATGCCTGA CACGGTACTACAGCAGCTTCGTGAACATGGA ACGCGACCTGGCCAGCGGCCTGATTGGTCCT CTGCTGATCTGCTACAAAGAAAGCGTGGACC AGAGGGGCAACCAGATCATGAGCGACAAGAG AAACGTGATCCTGTTCTCCGTCTTTGACGAGA ACAGGTCCTGGTATCTGACCGAGAACATCCA GCGGTTTCTGCCCAATCCTGCTGGCGTGCAG CTGGAAGATCCTGAGTTCCAGGCCTCCAACAT CATGCACTCCATCAACGGCTATGTGTTCGACA GCCTGCAGCTGAGCGTGTGCCTGCACGAAGT GGCCTACTGGTACATCCTGTCTATCGGCGCC CAGACCGACTTCCTGTCCGTGTTCTTTAGCGG CTACACCTTCAAGCACAAGATGGTGTACGAG GATACCCTGACACTGTTCCCATTCAGCGGCG AGACAGTGTTCATGAGCATGGAAAACCCCGG CCTGTGGATCCTGGGCTGTCACAACAGCGAC TTCAGAAACAGAGGCATGACAGCCCTGCTGA AGGTGTCCAGCTGCGACAAGAACACCGGCGA CTACTACGAGGACTCTTACGAGGACATCAGC GCCTACCTGCTGAGCAAGAACAATGCCATCG AGCCTCGGAGCTTCTCTCAGAACCCTCCTGT GCTGAAGAGACACCAGCGCGAGATCACCAGA ACCACACTGCAGAGCGACCAAGAGGAAATCG ATTACGACGACACCATCAGCGTCGAGATGAA GAAAGAAGATTTCGACATCTACGACGAGGAC GAGAATCAGAGCCCCAGATCTTTCCAGAAGAA AACGCGGCACTACTTCATTGCCGCCGTGGAA AGACTGTGGGACTACGGCATGAGCAGCAGCC CACATGTGCTGAGAAACAGGGCCCAGAGCGG AAGCGTGCCCCAGTTCAAGAAAGTGGTGTTC CAAGAGTTCACCGACGGCAGCTTCACCCAGC CTCTGTATAGAGGCGAGCTGAACGAGCACCT GGGACTGCTGGGACCTTACATCAGAGCTGAG GTCGAGGATAACATCATGGTCACCTTTAGAAA CCAGGCCTCTAGGCCCTACTCCTTCTACAGCT CCCTGATCAGCTACGAAGAGGACCAGAGACA GGGCGCTGAGCCCAGAAAGAACTTCGTGAAG CCCAACGAGACTAAGACCTACTTTTGGAAGGT GCAGCACCACATGGCCCCTACAAAGGACGAG TTCGACTGCAAGGCCTGGGCCTACTTCTCTGA CGTGGACCTCGAGAAGGATGTGCACAGCGGA CTCATCGGACCCCTGCTTGTGTGCCACACCA	

	ACACACTGAATCCCGCTCACGGCAGGCAAGT GACCGTGCAAGAGTTCGCCCTGTTCTTCACCA TCTTCGATGAGACAAAGTCCTGGTACTTCACC GAAAACATGGAAAGAAACTGCAGGGCCCCTT GCAACATCCAGATGGAAGATCCCACCTTCAAA GAGAACTACCGGTTCCACGCCATCAATGGCT ACATCATGGACACTCTGCCCGGCCTGGTTAT GGCACAGGATCAGAGGATCAGATGGTATCTG CTGTCCATGGGCTCCAACGAGAATATCCACA GCATCCACTTCAGCGGCCATGTGTTCACCGT GCGGAAAAAAGAAGAGTACAAGATGGCCCTG TACAATCTGTACCCCGGCGTGTTCGAGACTGT GGAAATGCTGCCTAGCAAGGCCGGAATCTGG CGCGTGGAATGTCTGATCGGAGAGCATCTGC ATGCCGGAATGTCTACCCTGTTCCTGGTGTAC AGCAACAAGTGTCAGACCCCTCTCGGCATGG CCTCTGGACACATCAGAGACTTCCAGATCACC GCCTCTGGCCAGTACGGACAGTGGGCTCCTA AACTGGCTAGACTGCACTACAGCGGCAGCAT CAACGCCTGGTCCACCAAAGAGCCCTTCAGC TGGATCAAGGTGGACCTGCTGGCTCCCATGA TCATCCACGGAATCAAGACCCAGGGCGCCAG ACAGAAGTTCAGCAGCCTGTACATCAGCCAGT TCATCATCATGTACAGCCTGGACGGCAAGAA GTGGCAGACCTACAGAGGCAACAGCACCGGC ACACTCATGGTGTTCTTCGGCAACGTGGACTC CAGCGGCATTAAGCACAACATCTTCAACCCTC CAATCATTGCCCGGTACATCCGGCTGCACCC CACACACTACAGCATCAGATCTACCCTGAGGA TGGAACTGATGGGCTGCGACCTGAACAGCTG CTCTATGCCCCTCGGAATGGAAAGCAAGGCC ATCAGCGACGCCCAGATCACAGCCAGCAGCT ACTTCACCAACATGTTCGCCACATGGTCCCCA TCTAAGGCCCGGCTGCATCTGCAGGGCAGAT CTAACGCTTGGAGGCCCCAAGTGAACAACCC CAAAGAGTGGCTGCAGGTCGACTTTCAGAAA ACCATGAAAGTGACCGGCGTGACCACACAGG GCGTCAAGTCTCTGCTGACCTCTATGTACGTG AAAGAGTTCCTGATCTCCAGCAGCCAGGACG GCCACCAGTGGACCCTGTTTTTCCAGAACGG CAAAGTCAAGGTGTTCCAGGGAAACCAGGAC AGCTTCACACCCGTGGTCAACTCCCTGGATC CTCCACTGCTGACCAGATACCTGAGAATTCAC CCTCAGTCTTGGGTGCACCAGATCGCTCTGA GAATGGAAGTGCTGGGATGTGAAGCTCAGGA CCTCTACTAAAATAAAAGATCTTTATTTTCATT AGATCTGTGTGTTGGTTTTTTGTGTGCGATCG GGAACTGGCATCTTCAGGGAGTAGCTTAGGT CAGTGAAGAGAAGTGCCAGTTCCCGATCGTT ACAGGCCGC	
95	GCCACCAGAAGATACTACCTGGGAGCTGTGG AATTGAGCTGGGATTACATGCAATCTGACCTG GGAGAACTGCCTGTGGATGCCAGGTTTCCTC	F8-BDD3, otimizado quanto a códon

	CTAGGGTCCCCAAGTCCTTCCCATTCAACACC TCAGTGGTCTACAAGAAAACCCTCTTTGTGGA GTTCACAGACCATCTGTTCAACATTGCCAAGC CAAGACCCCCATGGATGGGACTCCTGGGTCC AACCATCCAAGCTGAAGTGTATGACACTGTGG TCATTACCCTGAAGAACATGGCCTCCCATCCT GTGTCCCTGCATGCAGTGGGAGTGTCCTACT GGAAGGCTTCTGAAGGGGCTGAGTATGATGA TCAAACCAGCCAGAGAGAAAAGGAGGATGAC AAAGTGTTCCCAGGTGGTAGTCACACCTATGT GTGGCAAGTGCTCAAGGAGAATGGTCCTATG GCCTCTGATCCCCTGTGTCTGACCTACTCCTA CCTGTCCCATGTGGACCTGGTGAAGGATCTG AACTCTGGGCTGATTGGAGCCCTGCTGGTGT GCAGAGAAGGCTCCCTGGCCAAGGAAAAGAC CCAGACACTGCACAAGTTCATCTTGCTGTTTG CTGTGTTTGATGAGGGAAAGTCCTGGCATTCT GAGACTAAGAACTCCCTTATGCAAGACAGAGA TGCTGCCTCAGCTAGGGCTTGGCCTAAGATG CATACTGTGAATGGATATGTGAACAGATCCCT GCCTGGCCTTATTGGTTGCCACAGGAAGTCT GTGTATTGGCATGTGATTGGCATGGGAACCA CTCCAGAGGTGCACTCCATTTTCTTGGAGGG GCATACCTTCTTGGTGAGGAACCACAGACAG GCCTCCCTGGAAATTTCTCCAATCACTTTCCT GACTGCCCAGACCCTCCTTATGGACCTGGGT CAGTTCCTGCTGTTCTGCCACATTTCATCCCA CCAACATGATGGCATGGAAGCCTATGTGAAA GTGGACTCATGCCCAGAAGAACCACAGCTGA GAATGAAGAACAATGAAGAGGCAGAGGACTA TGATGATGATCTTACAGATTCAGAAATGGATG TGGTCAGATTTGATGATGATAATAGCCCATCC TTCATCCAAATTAGGAGTGTGGCCAAGAAGCA CCCCAAAACTTGGGTGCATTACATTGCAGCTG AGGAAGAGGATTGGGACTATGCACCCTTGGT GCTTGCACCAGATGATAGGTCCTACAAGTCCC AATACCTGAACAATGGCCCACAGAGGATTGGT AGAAAGTATAAGAAAGTGAGATTCATGGCCTA CACAGATGAGACTTTCAAGACCAGAGAGGCC ATTCAGCATGAATCTGGCATTCTGGGGCCACT GTTGTATGGGGAGGTTGGAGATACACTGCTC ATCATTTTCAAGAACCAGGCCTCCAGACCCTA CAACATCTACCCTCATGGAATCACTGATGTCA GACCCCTGTACTCCAGAAGACTCCCAAAGGG AGTCAAGCACTTGAAAGACTTCCCCATCCTGC CTGGGGAAATCTTCAAGTACAAGTGGACAGT GACAGTGGAGGATGGGCCAACCAAGTCTGAT CCAAGATGCCTCACTAGATACTACTCATCCTT TGTCAACATGGAAAGAGACCTGGCCTCAGGA CTGATTGGCCCCCTGCTCATCTGCTACAAGGA GTCTGTGGATCAGAGAGGAAACCAGATCATG TCTGACAAAAGGAATGTCATCCTCTTCTCTGT CTTTGATGAGAACAGATCATGGTACCTTACAG AGAACATCCAGAGGTTCCTCCCCAACCCTGCT	

	GGAGTGCAGCTGGAGGACCCAGAATTCCAGG CATCAAACATTATGCACTCCATCAATGGTTAT GTGTTTGACAGCCTCCAGCTTTCTGTGTGCCT CCATGAAGTGGCATATTGGTACATCCTGTCCA TTGGAGCACAAACAGACTTTCTCTCTGTGTTC TTCTCTGGATATACCTTCAAGCACAAGATGGT GTATGAGGATACCCTGACCCTCTTCCCCTTCT CTGGAGAGACTGTGTTTATGTCAATGGAAAAC CCAGGCCTGTGGATTTTGGGGTGCCACAACT CAGATTTCAGAAACAGGGGCATGACTGCCTT GCTCAAGGTGTCCTCCTGTGACAAGAACACA GGAGACTACTATGAGGACTCCTATGAGGATAT TTCTGCCTACCTCCTGTCCAAGAACAATGCCA TTGAACCCAGGTCCTTCAGCCAGAACCCTCCT GTCCTCAAGAGGCATCAGAGAGAAATCACCA GAACTACCCTGCAGTCTGACCAGGAAGAGAT TGATTATGATGACACTATCTCAGTGGAAATGA AGAAGGAGGACTTTGACATCTATGATGAAGAT GAAAATCAGTCCCCTAGGTCCTTCCAAAAGAA AACAAGACACTACTTCATTGCTGCTGTGGAGA GACTCTGGGACTATGGCATGTCCTCATCACCC CATGTGCTTAGGAACAGGGCTCAATCTGGGT CTGTCCCTCAGTTCAAGAAAGTGGTGTTTCAA GAATTCACAGATGGAAGCTTCACACAGCCATT GTACAGGGGAGAACTGAATGAGCACCTTGGC CTGCTGGGACCTTACATCAGAGCAGAGGTGG AGGACAACATCATGGTGACCTTCAGAAACCAA GCCTCCAGGCCATATTCATTCTACTCCAGCCT TATCTCATATGAGGAGGATCAGAGACAGGGG GCTGAACCTAGGAAGAACTTTGTCAAGCCAAA TGAGACAAAGACCTACTTTTGGAAGGTGCAGC ACCACATGGCCCCTACCAAGGATGAGTTTGA CTGCAAGGCCTGGGCTTACTTCTCTGATGTG GATCTGGAAAAGGATGTGCATTCTGGGCTGAT TGGACCTCTGCTGGTCTGCCACACTAACACC CTCAATCCTGCTCATGGCAGACAAGTGACAGT GCAGGAGTTTGCCCTGTTCTTCACCATCTTTG ATGAAACTAAGTCATGGTACTTTACAGAGAAC ATGGAGAGAAATTGTAGGGCCCCATGTAACAT CCAGATGGAGGACCCAACATTCAAGGAGAAC TACAGATTCCATGCCATTAATGGATACATTATG GACACTCTTCCAGGACTGGTGATGGCACAGG ACCAAAGAATCAGATGGTATCTTCTGAGCATG GGGAGCAATGAAAACATCCATTCCATCCACTT TTCAGGTCATGTGTTCACAGTGAGGAAGAAG GAAGAGTACAAGATGGCTCTGTACAACCTGTA CCCTGGGGTGTTTGAGACTGTGGAAATGCTG CCATCCAAGGCTGGAATTTGGAGGGTGGAAT GTCTGATTGGTGAACATCTGCATGCTGGAATG TCCACCCTGTTCCTGGTGTACTCCAACAAGTG CCAAACCCCACTGGGAATGGCATCAGGACAC ATTAGAGACTTCCAGATTACAGCATCTGGACA GTATGGACAATGGGCCCCCAAGTTGGCCAGG CTGCACTACTCTGGAAGCATTAATGCCTGGAG	

	CACCAAGGAGCCATTCAGCTGGATCAAGGTG GACCTTCTGGCTCCAATGATCATCCATGGAAT TAAGACTCAGGGAGCCAGACAGAAGTTCTCAT CCCTCTACATCTCCCAGTTTATCATCATGTACT CACTGGATGGGAAGAAGTGGCAGACTTACAG GGGAAATTCCACAGGTACTCTGATGGTGTTCT TTGGAAATGTGGACAGCTCTGGCATCAAGCA CAATATCTTTAACCCTCCTATCATTGCCAGGTA CATCAGACTCCACCCAACTCACTACTCCATCA GGTCCACTCTGAGGATGGAACTCATGGGTTG TGACCTCAACTCCTGCTCAATGCCACTGGGCA TGGAGTCCAAGGCTATCTCAGATGCTCAGATT ACTGCATCCTCTTACTTTACCAACATGTTTGCT ACCTGGTCCCCCTCCAAAGCCAGACTGCATC TCCAAGGCAGATCAAATGCCTGGAGGCCTCA GGTCAACAACCCAAAGGAATGGCTTCAGGTG GACTTCCAAAAGACCATGAAAGTCACAGGAGT GACCACCCAGGGAGTGAAATCCCTGCTGACC TCTATGTATGTGAAGGAATTCCTGATCTCATC AAGCCAGGATGGCCACCAGTGGACACTGTTC TTCCAAAATGGAAAGGTCAAGGTCTTTCAGGG AAATCAAGACTCCTTCACCCCTGTGGTGAACT CCCTGGACCCCCCTCTGCTTACCAGGTACTT GAGAATTCATCCACAATCCTGGGTGCACCAGA TTGCCCTGAGGATGGAAGTGCTGGGCTGTGA AGCCCAGGACCTGTACTAA	
96	gccaccagaaggtactacctgggagctgtggaactgagctggga ctacatgcagtctgacctgggagagctgccTgtggatgctagatttc ctccaagagtgcccaagagcttccccttcaacacctctgtggtgtac aagaaaaccctgtttgtggaattcacagaccacctgttcaatattgc caagcctagacctccttggatgggcctgctgggccctacaattcag gctgaggtgtatgacacagtggtcatcaccctgaagaacatggcc agccatcctgtgtctctgcatgctgtgggagtgtcttactggaaggctt ctgaggggctgagtatgatgaccagacaagccagagagagaa agaggatgacaaggttttccctgggggcagccacacctatgtctgg caggtcctgaaagaaaatggccctatggcctctgatcctctgtgcct gacatacagctacctgagccatgtggacctggtcaaggacctgaa ctctggcctgattggggctctgctggtgtgtagagaaggcagcctgg ccaaagaaaagacccagacactgcacaagttcatcctgctgtttg ctgtgtttgatgagggcaagagctggcactctgagacaaagaaca gcctgatgcaggacagagatgctgcctctgctagagcttggccca agatgcacacagtgaatggctatgtgaacagaagcctgcctggac tgattggatgccacagaaagtctgtgtactggcatgtgattggcatg ggcaccacacctgaggtgcacagcatctttctggaaggacacacc ttcctggtgaggaaccacagacaggccagcctggaaatcagccc tatcaccttcctgacagctcagaccctgctgatggatctgggccagtt tctgctgttctgccacatcagcagccaccagcatgatggcatggaa gcctatgtgaaggtggacagctgccctgaagaaccccagctgag aatgaagaacaatgaggaagctgaggactatgatgatgacctga cagactctgagatggatgtggtcagatttgatgatgataacagccc cagcttcatccagatcagatctgtggccaagaagcaccccaagac ctgggtgcactatattgctgctgaggaagaggactgggattatgctc ctctggtgctggcccctgatgacagaagctacaagagccagtacc	F8-BDD4, otimizado quanto a códon

	tgaacaatggccctcagagaattggcaggaagtataagaaagtg aggttcatggcctacacagatgagacattcaagaccagagaggct atccagcatgagtctggcattctgggacctctgctgtatggggaagt gggggacacactgctgatcatcttcaagaaccaggccagcagac cctacaacatctaccctcatggcatcacagatgtgaggcctctgtac tctagaaggctgcccaaggggtgaagcacctgaaggacttccct atcctgcctggggagatcttcaagtacaagtggacagtgacagtg gaggatggccctaccaagtctgatcctagatgcctgacaaggtact acagcagctttgtgaacatggaaagggacctggcctctggcctgat tggtcctctgctgatctgctacaaagaatctgtggaccagaggggc aaccagatcatgagtgacaagagaaatgtgatcctgttctctgtcttt gatgagaacaggtcctggtatctgacagagaacatccagaggttt ctgcccaatcctgctggggtgcagctggaagatcctgagttccagg cctccaacatcatgcactccatcaatggctatgtgtttgacagcctgc agctgtctgtgtgcctgcatgaagtggcctactggtacatcctgtctat tggggcccagacagacttcctgtctgtgttcttttctggctacaccttca agcacaagatggtgtatgaggataccctgacactgttcccattctct ggggagacagtgttcatgagcatggaaaaccctggcctgtggatc ctgggctgtcacaacagtgacttcagaaacagaggcatgacagc cctgctgaaggtgtccagctgtgacaagaacactggggactactat gaggactcttatgaggacatctctgcctacctgctgagcaagaaca atgccattgagcctaggagcttctctcagaaccctcctgtgctgaag agacaccagagggagatcaccagaaccacactgcagtctgacc aagaggaaattgattatgatgacaccatctctgtggagatgaagaa agaagattttgacatctatgatgaggatgagaatcagagccccag atctttccagaagaaaacaaggcactacttcattgctgctgtggaaa gactgtgggactatggcatgagcagcagcccccatgtgctgagaa acagggcccagtctggaagtgtgccccagttcaagaaagtggtgtt ccaagagttcacagatggcagcttcacccagcctctgtatagagg ggagctgaatgagcacctgggactgctgggaccttacatcagagc tgaggtggaggataacatcatggtcacctttagaaaccaggcctct aggccctactccttctacagctccctgatcagctatgaagaggacc agagacaggggctgagcccagaaagaactttgtgaagcccaa tgagactaagacctacttttggaaggtgcagcaccacatggcccct acaaaggatgagtttgactgcaaggcctgggcctacttctctgatgt ggacctggagaaggatgtgcactctggactcattggacccctgctt gtgtgccacaccaacacactgaatcctgctcatggcaggcaagtg acagtgcaagagtttgccctgttcttcaccatctttgatgagacaaag tcctggtacttcacagaaaacatggaaagaaactgcagggcccct tgcaacatccagatggaagatcccaccttcaaagagaactacag gttccatgccatcaatggctacatcatggacactctgcctggcctggt tatggcacaggatcagaggatcagatggtatctgctgtccatgggc tccaatgagaatatccacagcatccacttctctggccatgtgttcaca gtgaggaaaaaagaagagtacaagatggccctgtacaatctgta ccctggggtgtttgagactgtggaaatgctgcctagcaaggctgga atctggagggtggaatgtctgattggagagcatctgcatgctggaat gtctaccctgttcctggtgtacagcaacaagtgtcagacccctctgg gcatggcctctggacacatcagagacttccagatcacagcctctgg ccagtatggacagtgggctcctaaactggctagactgcactactct ggcagcatcaatgcctggtccaccaaagagcccttcagctggatc aaggtggacctgctggctcccatgatcatccatggaatcaagacc caggggggccagacagaagttcagcagcctgtacatcagccagtt catcatcatgtacagcctggatggcaagaagtggcagacctacag aggcaacagcacaggcacactcatggtgttctttggcaatgtggac	

	tcttctggcattaagcacaacatcttcaaccctccaatcattgccagg tacatcaggctgcaccccacacactacagcatcagatctaccctg aggatggaactgatgggctgtgacctgaacagctgctctatgcccc tgggaatggaaagcaaggccatctctgatgcccagatcacagcc agcagctacttcaccaacatgtttgccacatggtccccatctaaggc caggctgcatctgcagggcagatctaatgcttggaggccccaagt gaacaaccccaaagagtggctgcaggtggactttcagaaaacca tgaaagtgacaggagtgaccacacaggggtcaagtctctgctg acctctatgtatgtgaaagagttcctgatctccagcagccaggatgg ccaccagtggaccctgtttttccagaatggcaaagtcaaggtgttcc agggaaaccaggacagcttcacacctgtggtcaactccctggatc ctccactgctgaccagatacctgagaattcaccctcagtcttgggtg caccagattgctctgagaatggaagtgctgggatgtgaagctcag gacctctac	
97	gccaccagaagatactacctgggtgcagtggaactgtcatgggac tatatgcaaagtgatctcggtgagctgcctgtggacgcaagatttcct cctagagtgccaaaatcttttccattcaacacctcagtcgtgtacaa aaagactctgtttgtagaattcacggttcacctttcaacatcgctaag ccaaggccaccctggatgggtctgctaggtcctaccatccaggctg aggtttatgatacagtggtcattacacttaagaacatggcttcccatc ctgtcagtcttcatgctgttggtgtatcctactggaaagcttctgaggg agctgaatatgatgatcagaccagtcaaagggagaaagaagatg ataaagtcttccctggtggaagccatacatatgtctggcaggtcctg aaagagaatggtccaatggcctctgacccactgtgccttacctactc atatctttctcatgtggacctggtaaaagacttgaattcaggcctcatt ggagccctactagtatgtagagaagggagtctggccaaggaaaa gacacagaccttgcacaaatttatactacttttgctgtatttgatgaa gggaaaagttggcactcagaaacaaagaactccttgatgcagga tagggatgctgcatctgctcgggcctggcctaaaatgcacacagtc aatggttatgtaaacaggtctctgccaggtctgattggatgccacag gaaatcagtctattggcatgtgattggaatgggcaccactcctgaa gtgcactcaatattcctcgaaggtcacacatttcttgtgaggaaccat cgccaggcgtccttggaaatctcgccaataactttccttactgctcaa acactcttgatggaccttggacagtttctactgttttgtcatatctcttcc caccaacatgatggcatggaagcttatgtcaaagtagacagctgtc cagaggaaccccaactacgaatgaaaaataatgaagaagcgg aagactatgatgatgatcttactgattctgaaatggatgtggtcaggtt tgatgatgacaactctccttcctttatccaaattcgctcagttgccaag aagcatcctaaaacttgggtacattacattgctgctgaagaggagg actgggactatgctcccttagtcctcgcccccgatgacagaagttat aaaagtcaatatttgaacaatggccctcagcggattggtaggaagt acaaaaaagtccgatttatggcatacacagatgaaaccttaaga ctcgtgaagctattcagcatgaatcaggaatcttgggacctttacttta tggggaagttggagacacactgttgattatatttaagaatcaagcaa gcagaccatataacatctaccctcacggaatcactgatgtccgtcct ttgtattcaaggagattaccaaaaggtgtaaaacatttgaaggatttt ccaattctgccaggagaaatattcaaatataaatggacagtgactg tagaagatgggccaactaaatcagatcctcggtgcctgacccgct attactctagtttcgttaatatggagagagatctagcttcaggactcat tggccctctcctcatctgctacaaagaatctgtagatcaaagagga aaccagataatgtcagacaagaggaatgtcatcctgttttctgtattt gatgagaaccgaagctggtacctcacagagaatatacaacgcttt ctccccaatccagctggagtgcagcttgaggatccagagttccaa	FVIII-BDD nativa madura CDS

	gcctccaacatcatgcacagcatcaatggctatgtttttgatagtttgc agttgtcagtttgtttgcatgaggtggcatactggtacattctaagcatt ggagcacagactgacttccttttctgtcttcttctctggatataccttcaa acacaaaatggtctatgaagacacactcaccctattcccattctcag gagaaactgtcttcatgtcgatggaaaacccaggtctatggattctg gggtgccacaactcagactttcggaacagaggcatgaccgcctta ctgaaggtttctagttgtgacaagaacactggtgattattacgagga cagttatgaagatatttcagcatacttgctgagtaaaaacaatgcca ttgaaccaagaagcttctcccagaatccaccagtcttgaaacgcca tcaacgggaaataactcgtactactcttcagtcagatcaagaggaa attgactatgatgataccatatcagttgaaatgaagaaggaagattt tgacatttatgatgaggatgaaaatcagagcccccgcagctttcaa aagaaaacacgacactattttattgctgcagtggagaggctctggg attatgggatgagtagctccccacatgttctaagaaacagggctca gagtggcagtgtccctcagttcaagaaagttgttttccaggaatttac tgatggctcctttactcagcccttataccgtggagaactaaatgaac atttgggactcctggggccatatataagagcagaagttgaagataa tatcatggtaactttcagaaatcaggcctctcgtccctattccttctattc tagccttatttcttatgaggaagatcagaggcaaggagcagaacct agaaaaaactttgtcaagcctaatgaaaccaaaacttacttttgga aagtgcaacatcatatggcacccactaaagatgagtttgactgcaa agcctgggcttatttctctgatgttgacctggaaaaagatgtgcactc aggcctgattggacccttctggtctgccacactaacacactgaac cctgctcatgggagacaagtgacagtacaggaatttgctctgtttttc accatctttgatgagaccaaaagctggtacttcactgaaaatatgg aaagaaactgcagggctccctgcaatatccagatggaagatccc acttttaaagagaattatcgcttccatgcaatcaatggctacataatg gatacactacctggcttagtaatggctcaggatcaaaggattcgat ggtatctgctcagcatgggcagcaatgaaaacatccattctattcatt tcagtggacatgtgttcactgtacgaaaaaaagaggagtataaaat ggcactgtacaatctctatccaggtgtttttgagacagtggaaatgtt accatccaaagctggaatttggcgggtggaatgccttattggcgag catctacatgctgggatgagcacacttttctggtgtacagcaataag tgtcagactcccctgggaatggcttctggacacattagagattttcag attacagcttcaggacaatatggacagtgggccccaaagctggcc agacttcattattccggatcaatcaatgcctggagcaccaaggagc ccttttcttggatcaaggtggatctgttggcaccaatgattattcacgg catcaagacccagggtgcccgtcagaagttctccagcctctacatc tctcagtttatcatcatgtatagtcttgatgggaagaagtggcagactt atcgaggaaattccactggaaccttaatggtcttctttggcaatgtgg attcatctgggataaaacacaatatttttaaccctccaattattgctcg atacatccgtttgcacccaactcattatagcattcgcagcactcttcg catggagttgatgggctgtgatttaaatagttgcagcatgccattggg aatggagagtaaagcaatatcagatgcacagattactgcttcatcc tactttaccaatatgtttgccacctggtctccttcaaaagctcgacttca cctccaagggaggagtaatgcctggagacctcaggtgaataatcc aaaagagtggctgcaagtggacttccagaagacaatgaaagtca caggagtaactactcagggagtaaaatctctgcttaccagcatgtat gtgaaggagttcctcatctccagcagtcaagatggccatcagtgga ctctctttttttcagaatggcaaagtaaaggtttttcagggaaatcaag actccttcacacctgtggtgaactctctagacccaccgttactgactc gctaccttcgaattcacccccagagttgggtgcaccagattgccctg aggatggaggttctgggctgcgaggcacaggacctctac	

98	DAHATRRYY	
99	T/CNC/TT/CA/GAC/T N is any nucleotide	Sequência consenso do local ramo
100	NNNNNNNNNNNNNNNNNNNNRG N is any nucleotide R is G ou A	sequência de ácido nucleico alvo
101	NNNNGATT N is any nucleotide	PAM de Neisseria meningitidis
102	NNNNNGTTT N is any nucleotide	PAM de Neisseria meningitidis
103	NNNNGCTT N is any nucleotide	PAM de Neisseria meningitidis
104	ACCTTTTTTTTTTTTTACCTAGG	Albumina Humana Íntron-1_T25
105	UAAUUUUCUUUUGCGCACUA	Espaçador de gRNA exemplificativo
106	TAAAGCATAGTGCAATGGAT	
107	GAAAGCATGGTGCAATGGAT	
108	TATTGCACAGTGCAATGGAT	
109	TGATGCATATTGCAATGGAT	
110	TAATGAATAGGGCAATGGAT	
111	TAAGGCACAGTGTAATGGAT	
112	AAAAGCATAGACCAATGGAT	
113	TAGAGTATAGTGCAGTGGAT	
114	CAAAGCAAAGTGCAATTGAT	
115	GGAAGCATAGTGCAATGGTT	
116	TAAAGGATAGAGCAATGTAT	
117	TAGAGTATAGTGCAATGGAG	
118	TAAAGAATAGTGAAATGGTT	
119	ATTTATGAGATCAACAGCAC	
120	ATTTATGATATCATCAGCAC	
121	AAATATGACATCAACAGCAC	
122	ATCTTTGAGATCATCAGCAC	
123	ATGTATCAGATCATCAGCAC	
124	AATTATGAGATTCACAGCAC	
125	ATTTATGTGTTCAACCGCAC	
126	ATATATGACATCAACAGAAC	
127	ACTTATGATATCAACAGCAT	
128	TTAAATAAAGCATAGTGCAA	
129	TAAAATAAAGCATAGTGCAA	
130	TTAAATAAAGGATATTGCAA	

131	TTAAATAAAGCATTGAGCAA	
132	TACTATAAAGCATAGTGCAA	
133	TACTATAAAGCATAGTGCAA	
134	TTAAGGAAACCATAGTGCAA	
135	ATAAATATATCATAGTGCAA	
136	CTAAATAGAGAATAGTGCAA	
137	TTAAAGAAATTATAGTGCAA	
138	TTAAATATATAATAGTGCAA	
139	TTAAAAACAGCACAGTGCAA	
140	TTAAAATAAGCATGGTGCAA	
141	TTTGATAAAGCATAGGGCAA	
142	TTTTATAAAGCATAGTCCAA	
143	TTAAATGAAGAATATTGCAA	
144	ATAAATAAAGAATAGAGCAA	
145	TTGAATAAAGCAGAGTGGAA	
146	TTAATTAATGCATAGTGCCA	
147	TTAGATAAAGCTTAGTGCTA	
148	TTAGATAAAGCATACTGGAA	
149	TTAAAGAAAGCATGGTGCAG	
150	TTACATAAAGCATACTGCAT	
151	TTATATAAAGCATAGAGCAG	
152	TTAAATGAAGCATAGTGAAG	
153	TAATAAAATTCAAACATCCT	
154	GAATAAAATTCTAACATCCT	
155	TAATATAATTCCAACATCCT	
156	AAATAAAATTCAAACTTCCT	
157	GAGTTAAATTCAAACATCCT	
158	TTTTAAAAATCAAACATCCT	
159	AAATGAAAGTCAAACATCCT	
160	GATTAAAATTTAAACATCCT	
161	TCTTAAAATTCCAACATCCT	
162	AAAAAAAATTCCAACATCCT	
163	TACTGAAATTCTAACATCCT	
164	TACAAAAATTCACACATCCT	
165	TATTAGAATTCAGACATCCT	
166	TAATAAAGCCCAAACATCCT	
167	TATTTAAATTCAAATATCCT	
168	AAATAAAGTTCAAAGATCCT	
169	TACAAAAATTCAAACTTCCT	

170	GAATAAAATTTAAATATCCT	
171	TATAAAAATTCAAACAGCCT	
172	TACTAAAATTTAAACTTCCT	
173	TAATAACCTTCAAACATTCT	
174	TAGTAAAATTCAAATGTCCT	
175	CAATAAAATTCAACCATCAT	
176	GAATAAAATTCAAACTTCTT	
177	AAATAAAATTCAAAAATCCC	
178	CCCTCCGTTTGTCCTAGCTTTTC	Iniciador AlbF
179	CCAGATACAGAATATCTTCCTCAACGCAGA	Iniciador AlbR
180	CCTTTGGCACAATGAAGTGG	Iniciador Direto
181	GAATCTGAACCCTGATGACAAG	Iniciador Rev
182	TAAAGCATAGTGCCAATGGAT	
183	GAAAGCATAATAGCAATGGAT	
184	CTCACCATGGGGCGCCTGCAACTGGTT	
185	TAAAGGATCGTTCAACTCTGTGAGT	
186	CCAAGCATAGGTAATGGAT	
187	AAAAGCATAGTGAATGAAT	
188	CTTCGTATAGAAACAAGACAG	
189	GAGAGAGAGAGAAAGAGACAG	
190	ATTCTAGAGGCATAGAGAGTTCAACCT	
191	TTTTCTGCCATTGACCTTAAAGCGC	
192	ATCTGTGGGATTATGACTGAAC	
193	CTTCTCATAAAACCTAGACAG	
194	ATGTGGAGATCATTGAGCA	
195	ATTAATATGGTATCATGGGAGCAGGAC	
196	TTAAAATAAAGCATAGTGCAA	
197	TTCCAACGAAGGCCTCAA	
198	TTACTATAAAGCATAGTGCAA	
199	AAAAAAAAAGAAAAGAAAAGAAA	
200	TTACCTAAAACAATTTCACA	
201	TTAATACTGGGCCCTGAAGCCAAATACAGTT	
202	GGCAACAACACATCATCAGTAGGGTAA	
203	TTCAGAAATAGAAAAGCTGATCCTCAA	
204	TTAAAGCGCTTGAAATCTACACTTGCAA	
205	TGATAGGATGTGTGTGTAGAAGACTCC	
206	CCATAGAAGATACCAGGACTTCTT	
207	ATGTAGAAGTTGGTCACGGTCCGCATCGGCT	
208	AAATAGAATACCTCAGCATTTCT	

209	AGATGAAAATCTATCAATGGCACCAGCGCCT	
210	TAAAAAAGGGCTGAGCATAGTGGCTCACACCT	
211	TATTCAACTCACAGAGTTGAACGATCCT	
212	tcgcg	Espaçador

REIVINDICAÇÕES

1. Sistema, caracterizado pelo fato de que compreende:

uma endonuclease de ácido desoxirribonucleico (DNA) ou ácido nucleico que codifica a referida endonuclease de DNA;

um RNA guia (gRNA) compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104; e

um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional.

2. Sistema, de acordo com a reivindicação 1, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

3. Sistema, de acordo com a reivindicação 2, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

4. Sistema, de acordo com a reivindicação 2, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

5. Sistema, de acordo com a reivindicação 2, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

6. Sistema, de acordo com a reivindicação 2, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

7. Sistema, de acordo com qualquer uma das reivindicações 1 a 6, caracterizado pelo fato de que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1,

Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1 ou um seu derivado funcional.

8. Sistema, de acordo com qualquer uma das reivindicações 1 a 7, caracterizado pelo fato de que a referida endonuclease de DNA é Cas9.

9. Sistema, de acordo com qualquer uma das reivindicações 1 a 8, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão em uma célula hospedeira.

10. Sistema, de acordo com qualquer uma das reivindicações 1 a 9, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão em uma célula hospedeira.

11. Sistema, de acordo com qualquer uma das reivindicações 1 a 10, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

12. Sistema, de acordo com qualquer uma das reivindicações 1 a 10, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

13. Sistema, de acordo com a reivindicação 12, caracterizado pelo fato de que o RNA que codifica a referida endonuclease de DNA é um mRNA.

14. Sistema, de acordo com qualquer uma das reivindicações 1 a 13, caracterizado pelo fato de que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

15. Sistema, de acordo com a reivindicação 14, caracterizado pelo fato de que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA.

16. Sistema, de acordo com a reivindicação 15, caracterizado pelo fato de que o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

17. Sistema, de acordo com a reivindicação 15 ou 16, caracterizado pelo fato de que o local alvo de gRNA é um local alvo para um gRNA no sistema.

18. Sistema, de acordo com a reivindicação 17, caracterizado pelo fato de que o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA genômico para um gRNA no sistema.

19. Sistema, de acordo com qualquer uma das reivindicações 1 a 18, caracterizado pelo fato de que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulado em um lipossomo ou nanopartícula lipídica.

20. Sistema, de acordo com a reivindicação 19, caracterizado pelo fato de que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

21. Sistema, de acordo com qualquer uma das reivindicações 1a 20, caracterizado pelo fato de que compreende a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

22. Método de edição de um genoma em uma célula, caracterizado pelo fato de que compreende:

fornecer o seguinte para a célula:

(a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104;

(b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e

(c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

23. Método, de acordo com a reivindicação 22, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

24. Método, de acordo com a reivindicação 23, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

25. Método, de acordo com a reivindicação 23, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

26. Método, de acordo com a reivindicação 23, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

27. Método, de acordo com a reivindicação 23, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

28. Método, de acordo com qualquer uma das reivindicações 22 a 27, caracterizado pelo fato de que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17,

Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1; ou um seu derivado funcional.

29. Método, de acordo com qualquer uma das reivindicações 22 a 28, caracterizado pelo fato de que a referida endonuclease de DNA é Cas9.

30. Método, de acordo com qualquer uma das reivindicações 22 a 29, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão na célula.

31. Método, de acordo com qualquer uma das reivindicações 22 a 30, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão na célula.

32. Método, de acordo com qualquer uma das reivindicações 22 a 31, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

33. Método, de acordo com qualquer uma das reivindicações 22 a 31, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

34. Método, de acordo com a reivindicação 33, caracterizado pelo fato de que o RNA que codifica a referida endonuclease de DNA é um mRNA.

35. Método, de acordo com qualquer uma das reivindicações 22 a 34, caracterizado pelo fato de que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

36. Método, de acordo com qualquer uma das reivindicações 22 a 35, caracterizado pelo fato de que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local

alvo de gRNA.

37. Método, de acordo com a reivindicação 36, caracterizado pelo fato de que o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

38. Método, de acordo com a reivindicação 36 ou 37, caracterizado pelo fato de que o local alvo de gRNA é um local alvo para o gRNA de (a).

39. Método, de acordo com a reivindicação 38, caracterizado pelo fato de que o local alvo de gRNA do molde doador é o complemento reverso de um local alvo de gRNA no genoma da célula para o gRNA de (a).

40. Método, de acordo com qualquer uma das reivindicações 22 a 39, caracterizado pelo fato de que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

41. Método, de acordo com a reivindicação 40, caracterizado pelo fato de que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

42. Método, de acordo com qualquer uma das reivindicações 22 a 41, caracterizado pelo fato de que compreende fornecer à célula a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

43. Método, de acordo com qualquer uma das reivindicações 22 a 42, caracterizado pelo fato de que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser fornecido à célula.

44. Método, de acordo com qualquer uma das reivindicações 22 a 43, caracterizado pelo fato de que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b)

são fornecidos à célula pelo menos 14 dias após (c) ser fornecido à célula.

45. Método, de acordo com a reivindicação 43 ou 44, caracterizado pelo fato de que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b).

46. Método, de acordo com a reivindicação 45, caracterizado pelo fato de que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

47. Método, de acordo com qualquer uma das reivindicações 22 a 46, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

48. Método, de acordo com qualquer uma das reivindicações 22 a 47, caracterizado pelo fato de que a referida célula é um hepatócito.

49. Célula geneticamente modificada, caracterizada pelo fato de que o genoma da célula é editado pelo método, como definido em qualquer uma das reivindicações 22 a 48.

50. Célula geneticamente modificada, de acordo com a reivindicação 49, caracterizada pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado

funcional é expressa sob o controle do promotor endógeno da albumina.

51. Célula geneticamente modificada, de acordo com a reivindicação 49 ou 50, caracterizada pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é otimizada quanto a códon para expressão na célula.

52. Célula geneticamente modificada, de acordo com qualquer uma das reivindicações 49 a 51, caracterizada pelo fato de que a referida célula é um hepatócito.

53. Método de tratamento da Hemofilia A em um sujeito, caracterizado pelo fato de que compreende:

fornecer o seguinte a uma célula no sujeito:

(a) um gRNA compreendendo uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28, 30, 18-20, 23-27, 29, 31-44 e 104;

(b) uma endonuclease de DNA ou ácido nucleico que codifica a referida endonuclease de DNA; e

(c) um molde doador compreendendo uma sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional.

54. Método, de acordo com a reivindicação 53, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de qualquer uma das SEQ ID NOs: 22, 21, 28 e 30.

55. Método, de acordo com a reivindicação 54, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 22.

56. Método, de acordo com a reivindicação 54, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 21.

57. Método, de acordo com a reivindicação 54, caracterizado

pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 28.

58. Método, de acordo com a reivindicação 54, caracterizado pelo fato de que o gRNA compreende uma sequência espaçadora de SEQ ID NO: 30.

59. Método, de acordo com qualquer uma das reivindicações 53 a 58, caracterizado pelo fato de que o referido sujeito é um paciente com ou é suspeito de ter Hemofilia A.

60. Método, de acordo com qualquer uma das reivindicações 53 a 58, caracterizado pelo fato de que o referido sujeito é diagnosticado com um risco de Hemofilia A.

61. Método, de acordo com qualquer uma das reivindicações 53 a 60, caracterizado pelo fato de que a referida endonuclease de DNA é selecionada a partir do grupo que consiste em uma endonuclease Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecida como Csn1 e Csx12), Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 ou Cpf1; ou um seu derivado funcional.

62. Método, de acordo com qualquer uma das reivindicações 53 a 61, caracterizado pelo fato de que a referida endonuclease de DNA é Cas9.

63. Método, de acordo com qualquer uma das reivindicações 53 a 62, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é otimizado quanto a códon para expressão na célula.

64. Método, de acordo com qualquer uma das reivindicações 53 a 63, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou seu derivado funcional é

otimizada quanto a códon para expressão na célula.

65. Método, de acordo com qualquer uma das reivindicações 53 a 64, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido desoxirribonucleico (DNA).

66. Método, de acordo com qualquer uma das reivindicações 53 a 64, caracterizado pelo fato de que o ácido nucleico que codifica a referida endonuclease de DNA é um ácido ribonucleico (RNA).

67. Método, de acordo com a reivindicação 66, caracterizado pelo fato de que o RNA que codifica a referida endonuclease de DNA é um mRNA.

68. Método, de acordo com qualquer uma das reivindicações 53 a 67, caracterizado pelo fato de que um ou mais do gRNA de (a), a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) e o molde doador de (c) são formulados em um lipossomo ou nanopartícula lipídica.

69. Método, de acordo com qualquer uma das reivindicações 53 a 68, caracterizado pelo fato de que o molde doador é codificado em um vetor de Vírus Adeno-Associado (AAV).

70. Método, de acordo com qualquer uma das reivindicações 53 a 69, caracterizado pelo fato de que o molde doador compreende um cassete doador compreendendo a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional, e em que o cassete doador é flanqueado em um ou nos dois lados por um local alvo de gRNA.

71. Método, de acordo com a reivindicação 70, caracterizado pelo fato de que o cassete doador é flanqueado em ambos os lados por um local alvo de gRNA.

72. Método, de acordo com a reivindicação 70 ou 71, caracterizado pelo fato de que o local alvo de gRNA é um local alvo para o gRNA de (a).

73. Método, de acordo com a reivindicação 72, caracterizado pelo fato de que o local alvo de gRNA do molde doador é o complemento reverso do local alvo de gRNA no genoma da célula para o gRNA de (a).

74. Método, de acordo com qualquer uma das reivindicações 53 a 73, caracterizado pelo fato de que fornecer o molde doador à célula compreende administrar o molde doador ao sujeito.

75. Método, de acordo com a reivindicação 74, caracterizado pelo fato de que a administração é por via intravenosa.

76. Método, de acordo com qualquer uma das reivindicações 53 a 75, caracterizado pelo fato de que a referida endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA é formulada em um lipossomo ou nanopartícula lipídica.

77. Método, de acordo com a reivindicação 76, caracterizado pelo fato de que o referido lipossomo ou nanopartícula lipídica também compreende o gRNA.

78. Método, de acordo com a reivindicação 77, caracterizado pelo fato de que fornecer o gRNA e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA à célula compreende administrar o lipossomo ou nanopartícula lipídica ao sujeito.

79. Método, de acordo com a reivindicação 78, caracterizado pelo fato de que a administração é por via intravenosa.

80. Método, de acordo com qualquer uma das reivindicações 53 a 79, caracterizado pelo fato de que compreende proporcionar à célula a endonuclease de DNA pré-complexada com o gRNA, formando um complexo de Ribonucleoproteína (RNP).

81. Método, de acordo com qualquer uma das reivindicações 53 a 80, caracterizado pelo fato de que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula mais de 4 dias após o molde doador de (c) ser

fornecido à célula.

82. Método, de acordo com qualquer uma das reivindicações 53 a 81, caracterizado pelo fato de que o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidos à célula pelo menos 14 dias após o molde doador de (c) ser fornecido à célula.

83. Método, de acordo com a reivindicação 81 ou 82, caracterizado pelo fato de que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b).

84. Método, de acordo com a reivindicação 83, caracterizado pelo fato de que uma ou mais doses adicionais do gRNA de (a) e da endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) são fornecidas à célula após a primeira dose do gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) até um nível alvo de integração direcionada da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional e/ou um nível alvo de expressão da sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional ser alcançado.

85. Método, de acordo com qualquer uma das reivindicações 81 a 84, caracterizado pelo fato de que fornecer o gRNA de (a) e a endonuclease de DNA ou ácido nucleico que codifica a endonuclease de DNA de (b) à célula compreende administrar ao sujeito uma nanopartícula lipídica compreendendo o ácido nucleico que codifica a endonuclease de DNA e o gRNA.

86. Método, de acordo com qualquer uma das reivindicações 81 a 85, caracterizado pelo fato de que fornecer o molde doador de (c)

à célula compreende administrar ao sujeito o molde doador codificado em um vetor AAV.

87. Método, de acordo com qualquer uma das reivindicações 53 a 86, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa sob o controle do promotor endógeno da albumina.

88. Método, de acordo com qualquer uma das reivindicações 53 a 87, caracterizado pelo fato de que a referida célula é um hepatócito.

89. Método, de acordo com qualquer uma das reivindicações 53 a 88, caracterizado pelo fato de que a sequência de ácido nucleico que codifica uma proteína Fator VIII (FVIII) ou derivado funcional é expressa no fígado do sujeito.

90. Método de tratamento da Hemofilia A em um sujeito caracterizado pelo fato de que compreende:

administrar a célula geneticamente modificada, como definido em qualquer uma das reivindicações 49 a 52, ao sujeito.

91. Método, de acordo com a reivindicação 90, caracterizado pelo fato de que a referida célula geneticamente modificada é autóloga ao sujeito.

92. Método, de acordo com a reivindicação 90 ou 91, caracterizado pelo fato de que compreende adicionalmente:

obter uma amostra biológica do sujeito em que a amostra biológica compreende uma célula de hepatócito, em que a célula geneticamente modificada é preparada a partir do hepatócito.

93. Kit, caracterizado pelo fato de que compreende um ou mais elementos do sistema, como definido em qualquer uma das reivindicações 1 a 21, e compreendendo adicionalmente instruções de uso.

FIG. 1
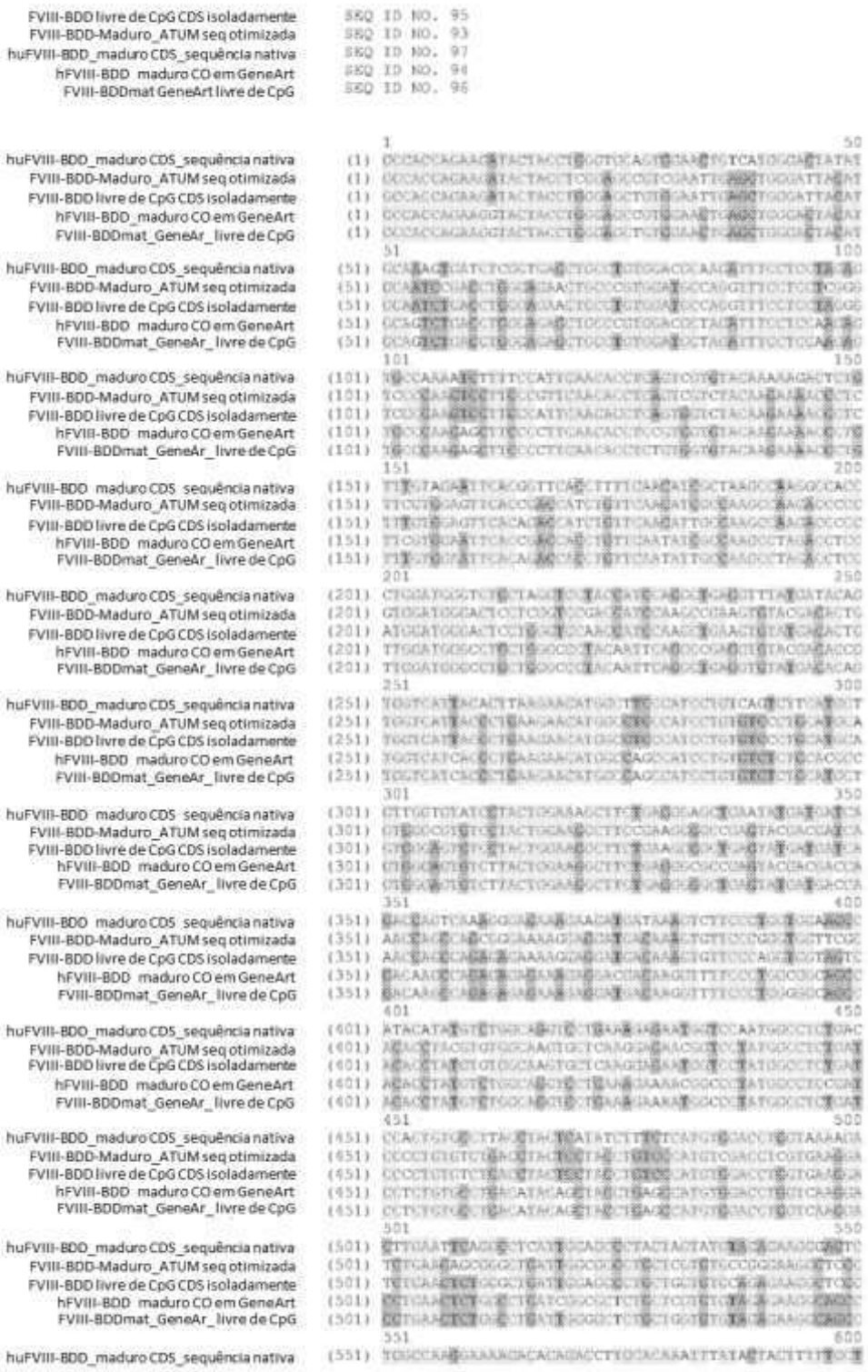

FVIII-BDD livre de CpG CDS isoladamente SEQ ID NO. 95
FVIII-BDD-Maduro_ATUM seq otimizada SEQ ID NO. 93
huFVIII-BDD_maduro CDS_sequência nativa SEQ ID NO. 97
hFVIII-BDD maduro CO em GeneArt SEQ ID NO. 94
FVIII-BDDmat GeneArt livre de CpG SEQ ID NO. 96
1 50
51 100
101 150
151 200
201 250
251 300
301 350
351 400
401 450
451 500
501 550
551 600

FIG. 1 (CONTINUAÇÃO)

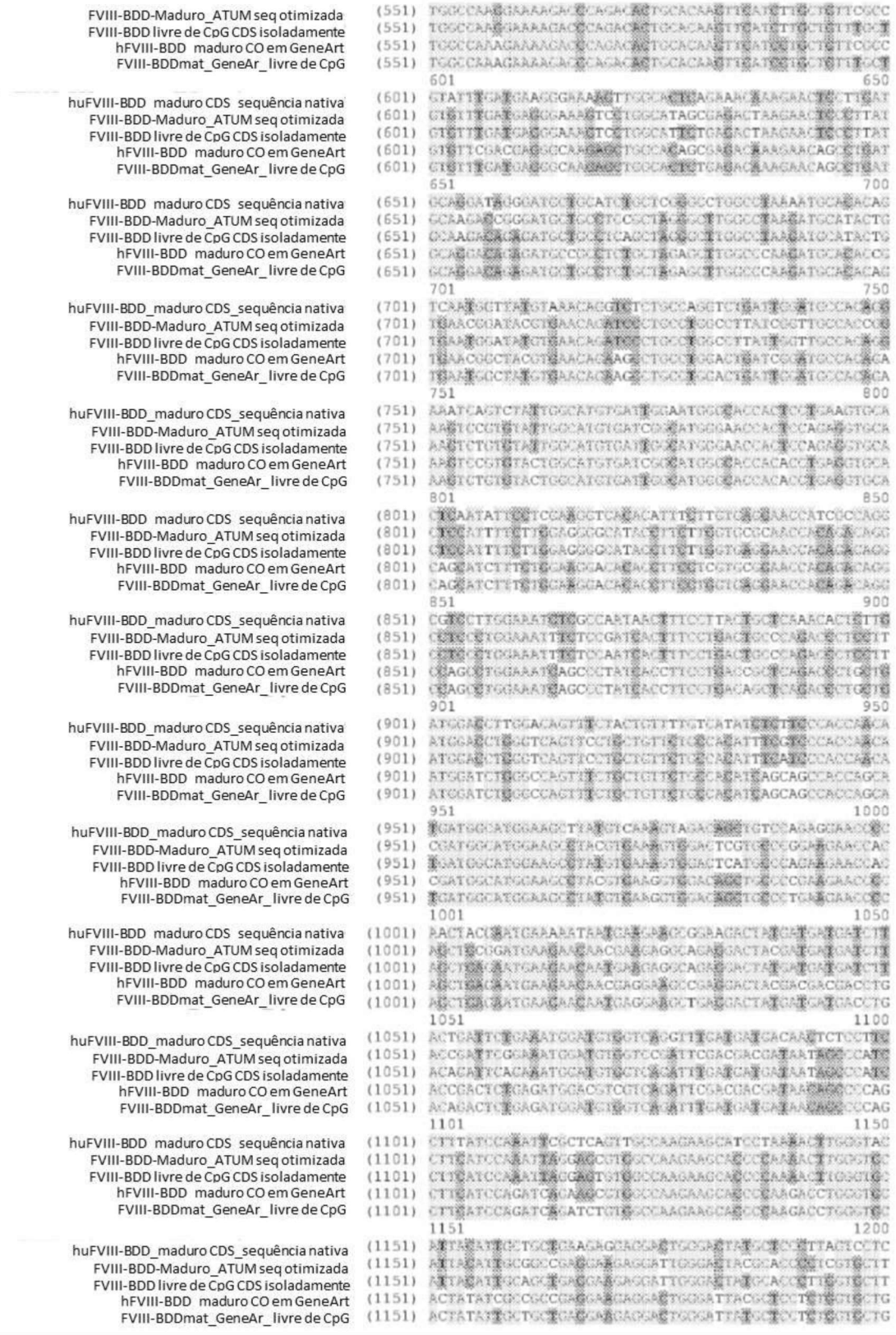


FIG. 1 (CONTINUAÇÃO)

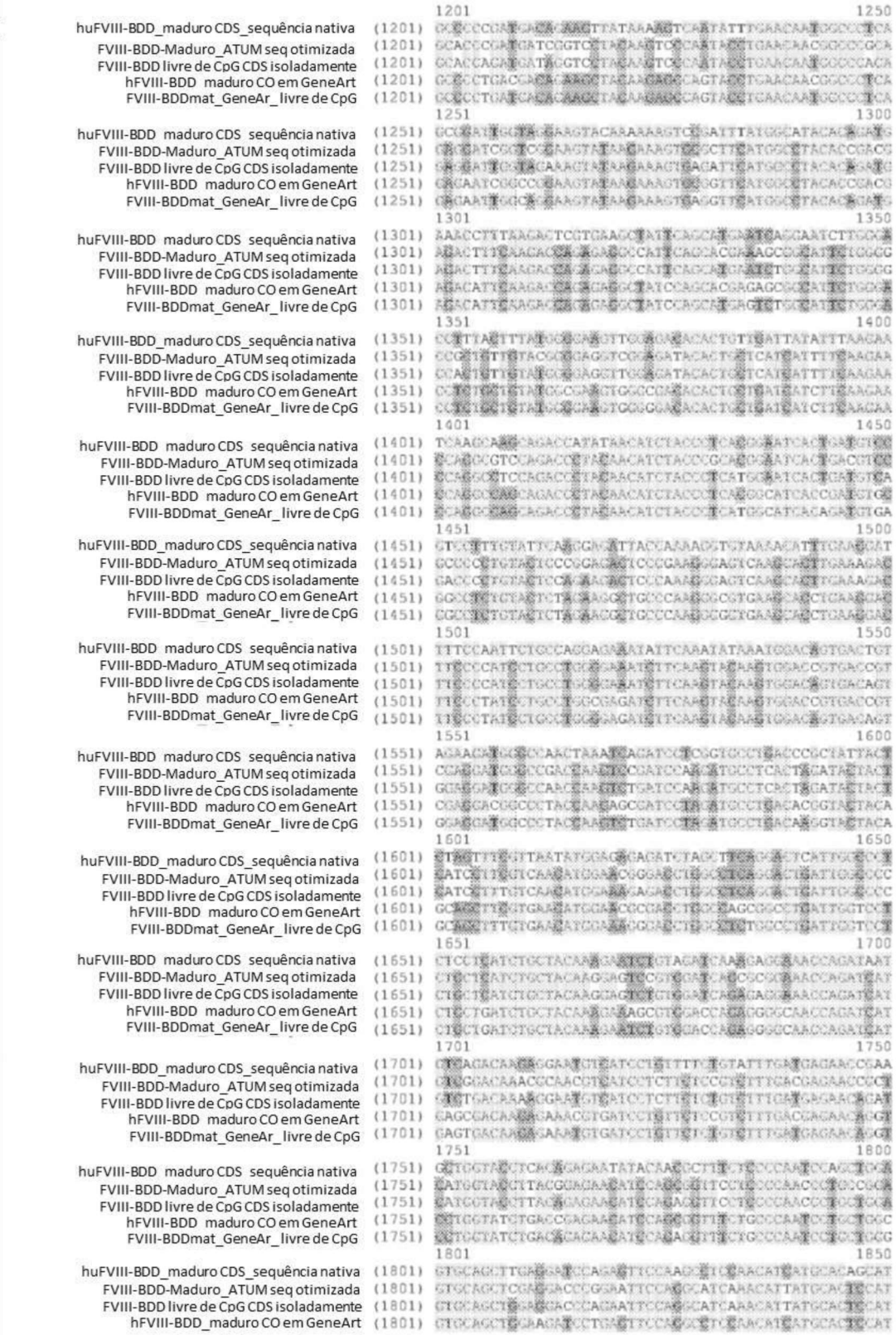


FIG. 1 (CONTINUAÇÃO)

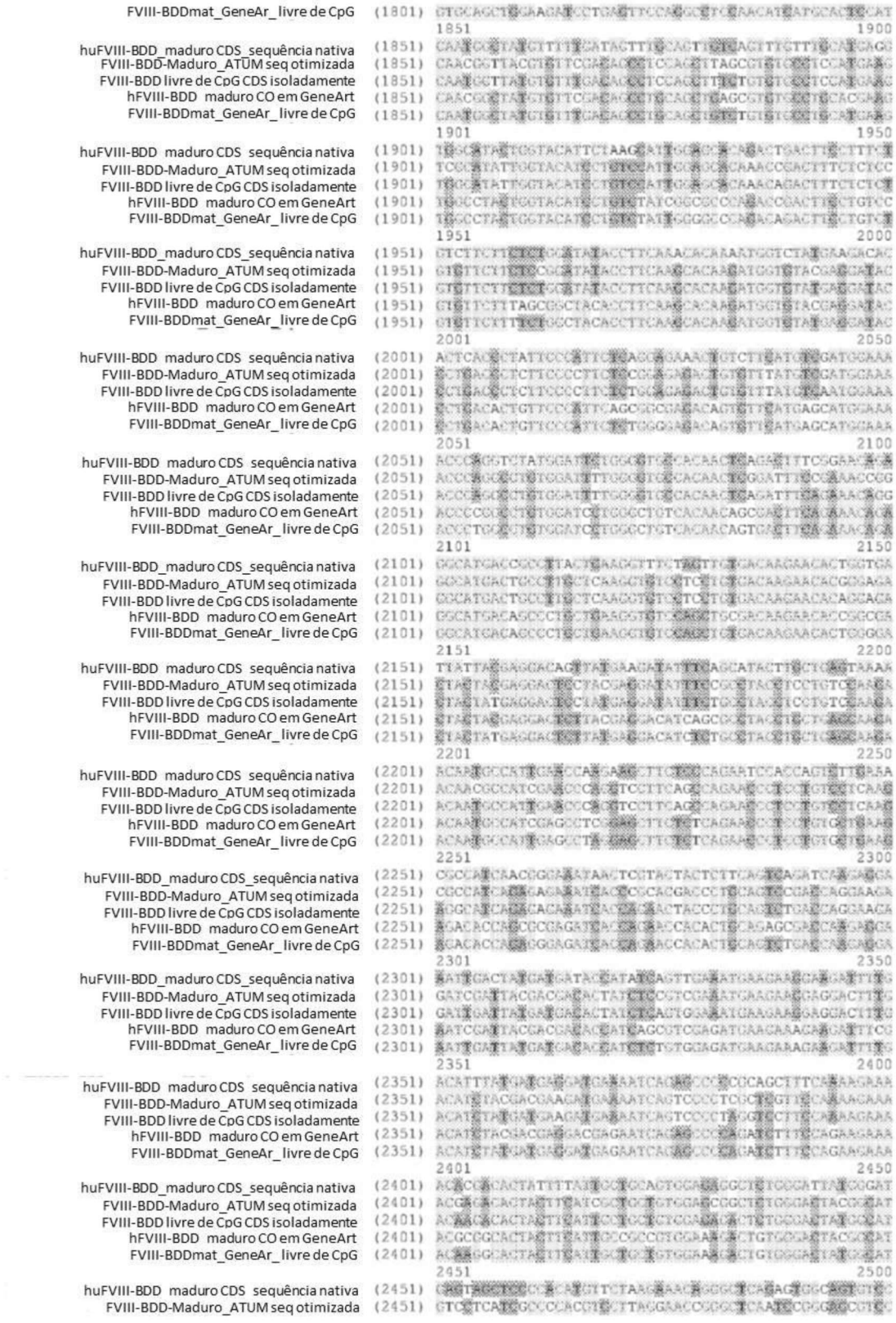


FIG. 1 (CONTINUAÇÃO)

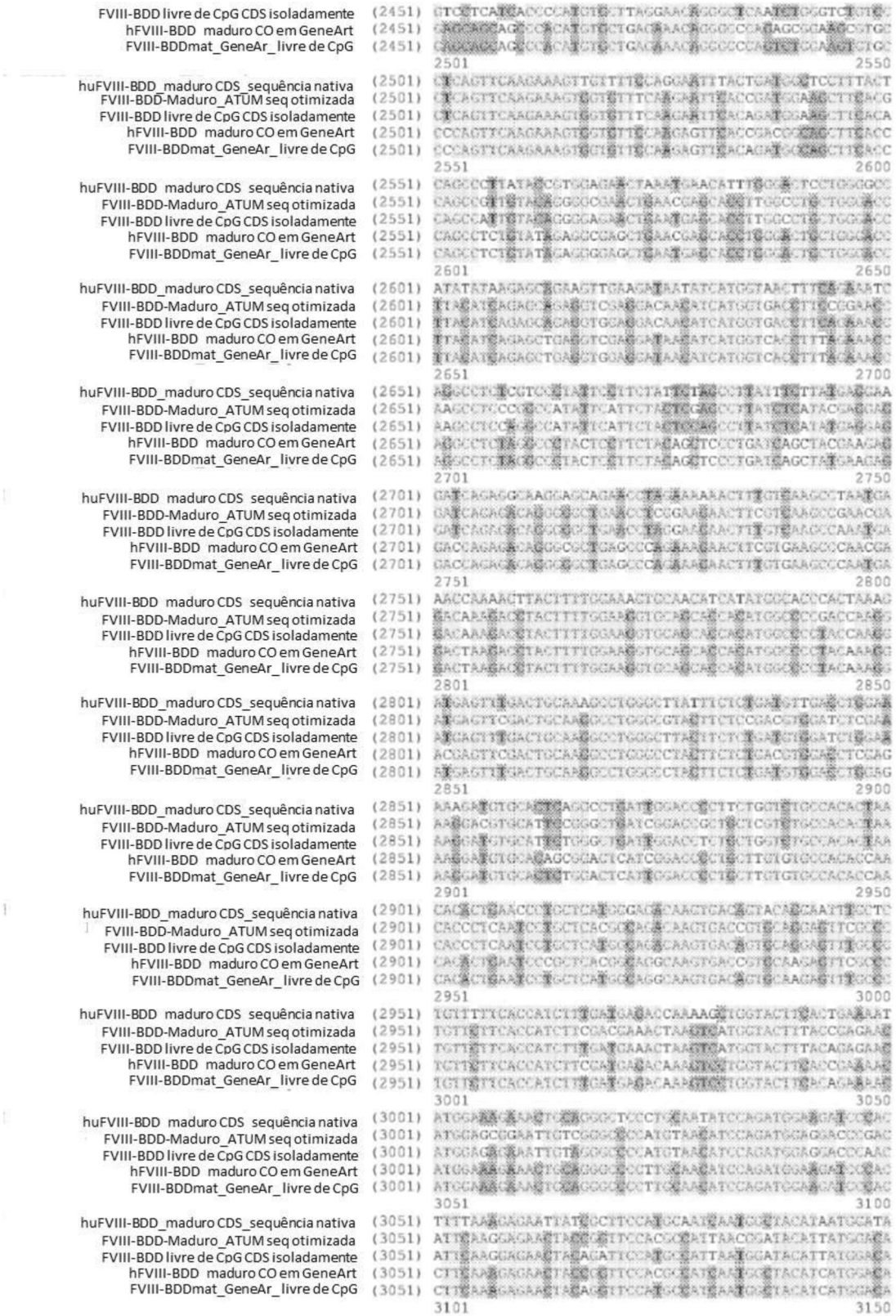


FIG. 1 (CONTINUAÇÃO)

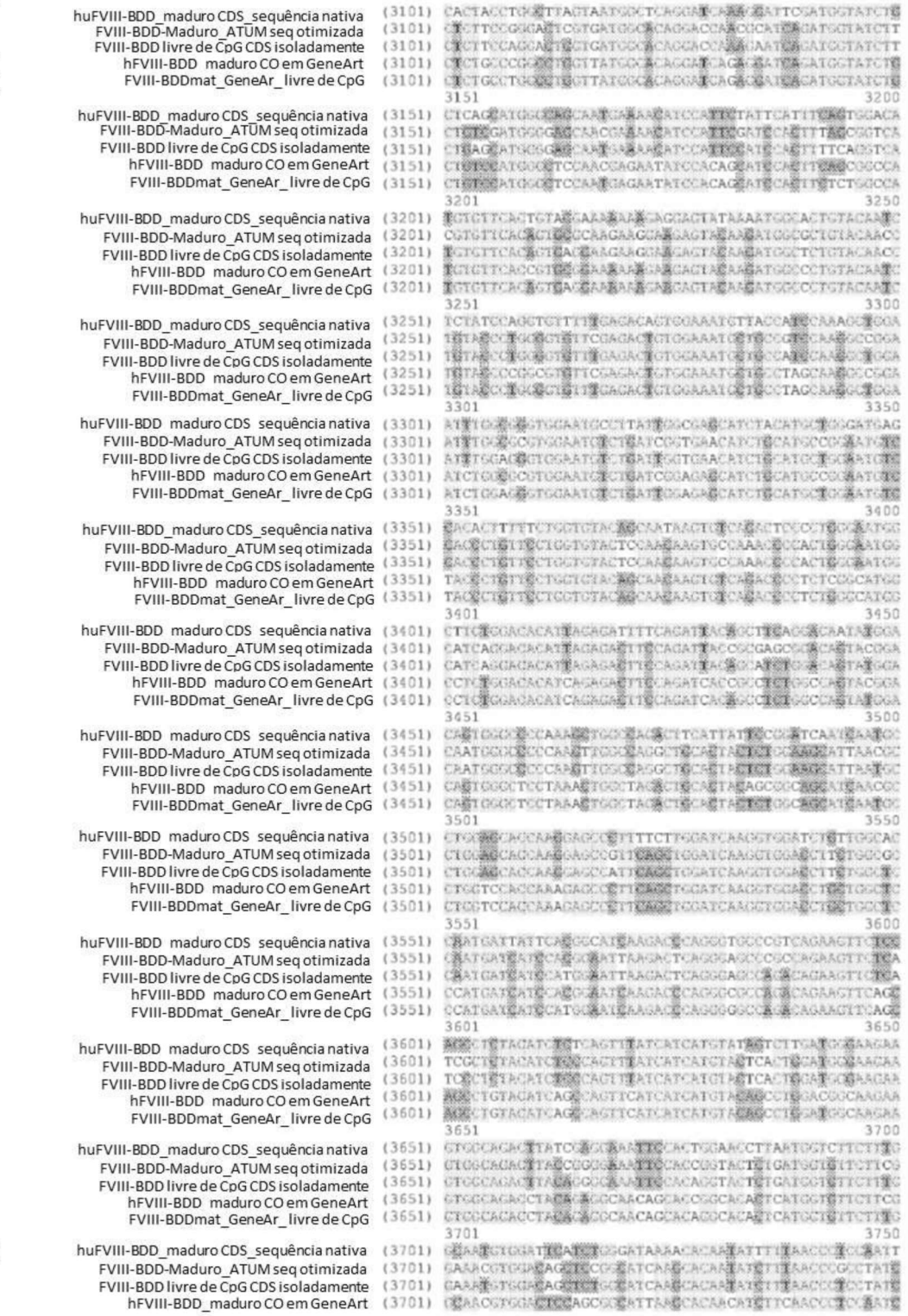
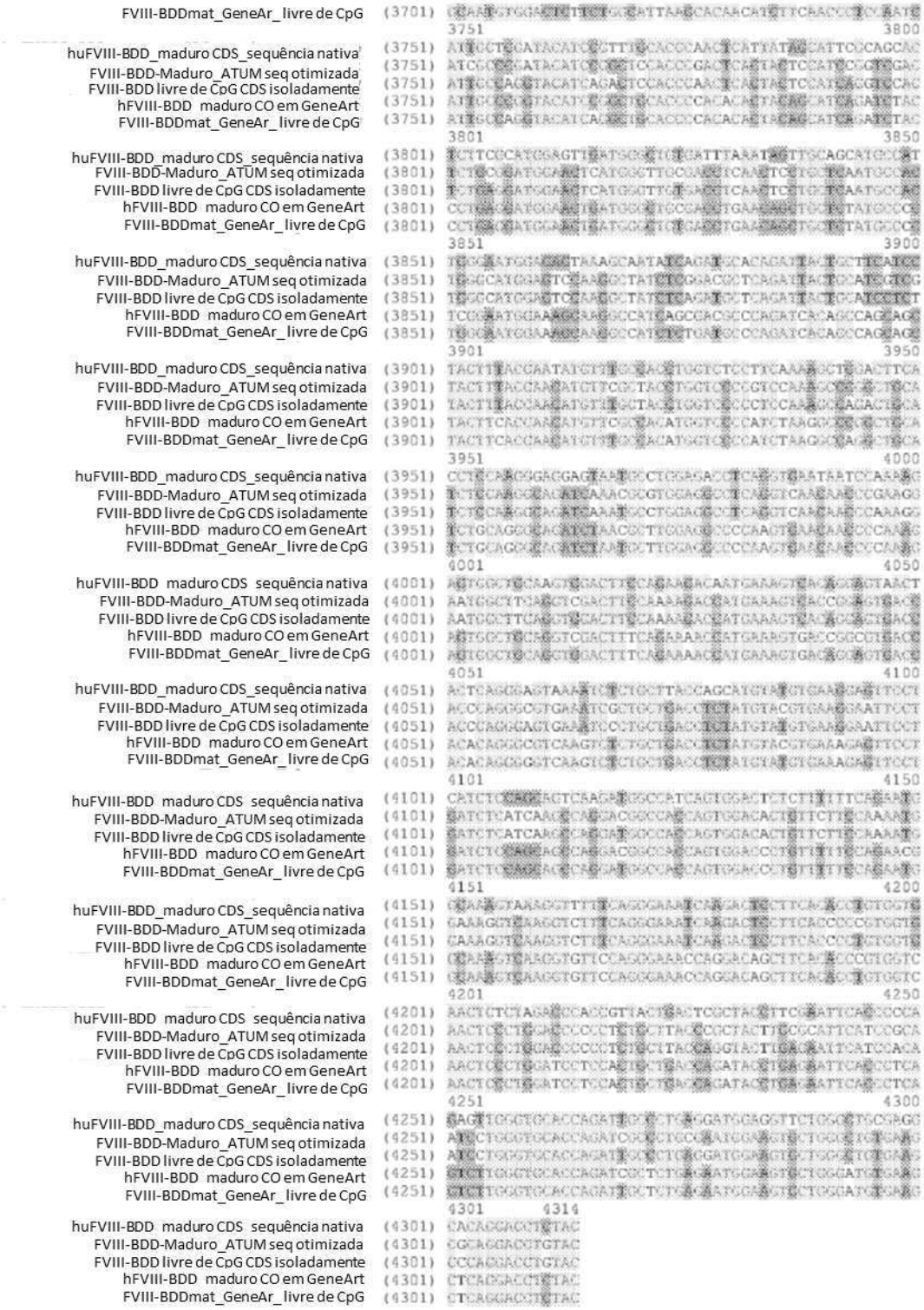


FIG. 1 (CONTINUAÇÃO)



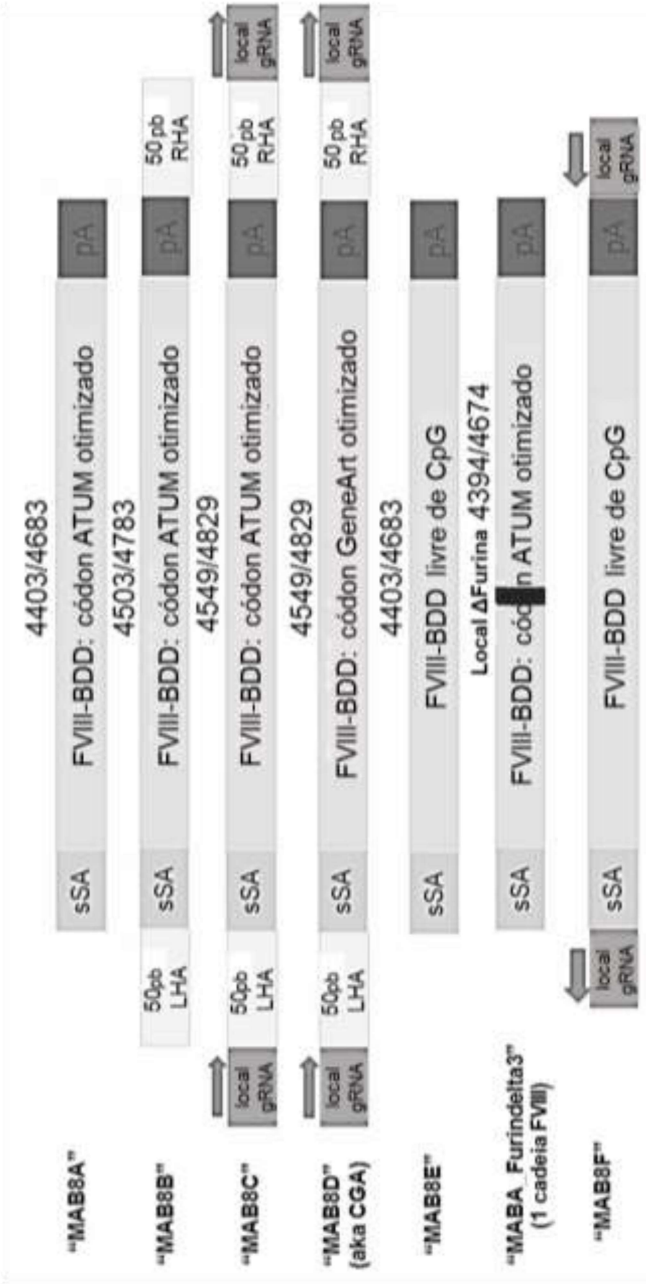

"MAB8A"
4403/4683
sSA
FVIII-BDD: códon ATUM otimizado
pA
"MAB8B"
4503/4783
50pb LHA
sSA
FVIII-BDD: códon ATUM otimizado
pA
50 pb RHA
local gRNA
"MAB8C"
4549/4829
local gRNA
50pb LHA
sSA
FVIII-BDD: códon ATUM otimizado
pA
50 pb RHA
local gRNA
"MAB8D" (aka CGA)
4549/4829
local gRNA
50pb LHA
sSA
FVIII-BDD: códon GeneArt otimizado
pA
50 pb RHA
local gRNA
"MAB8E"
4403/4683
sSA
FVIII-BDD livre de CpG
pA
"MABA_Furindelta3" (1 cadeia FVIII)
Local ΔFurina 4394/4674
sSA
FVIII-BDD: cód n ATUM otimizado
pA
local gRNA
"MAB8F"
local gRNA
sSA
FVIII-BDD livre de CpG
pA
local gRNA

FIG. 2

FIG. 3

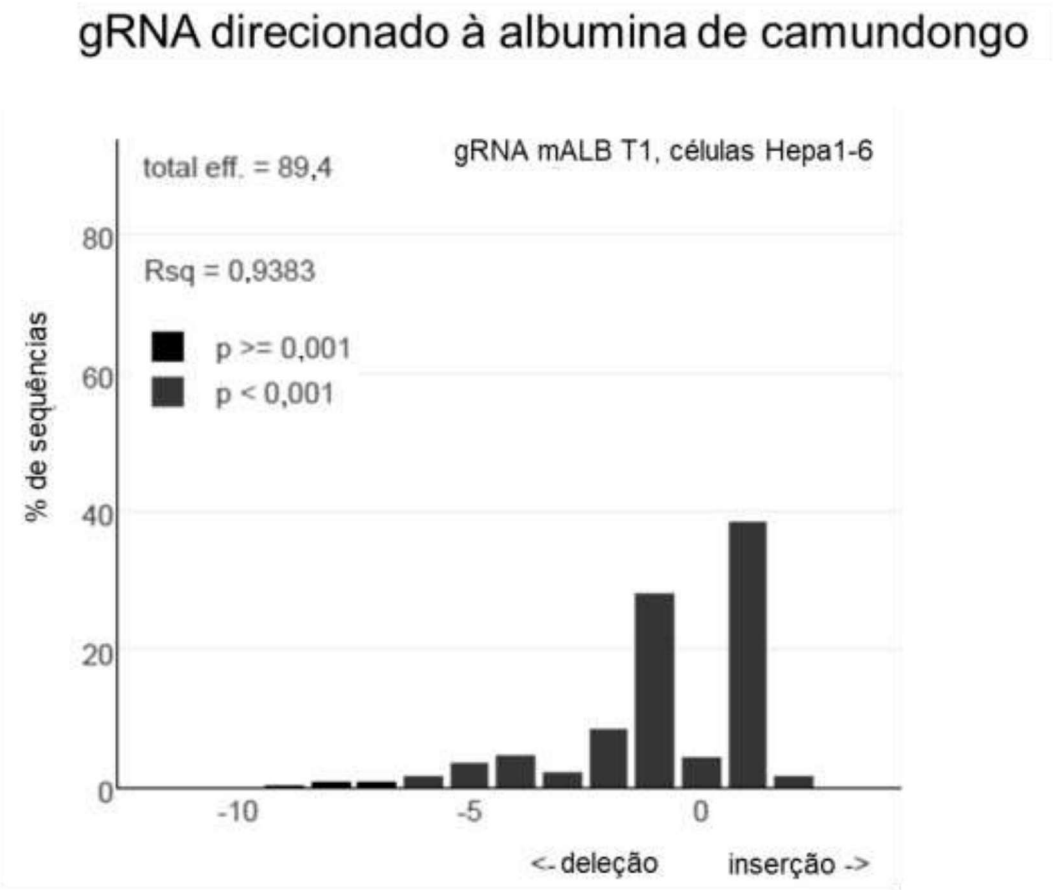
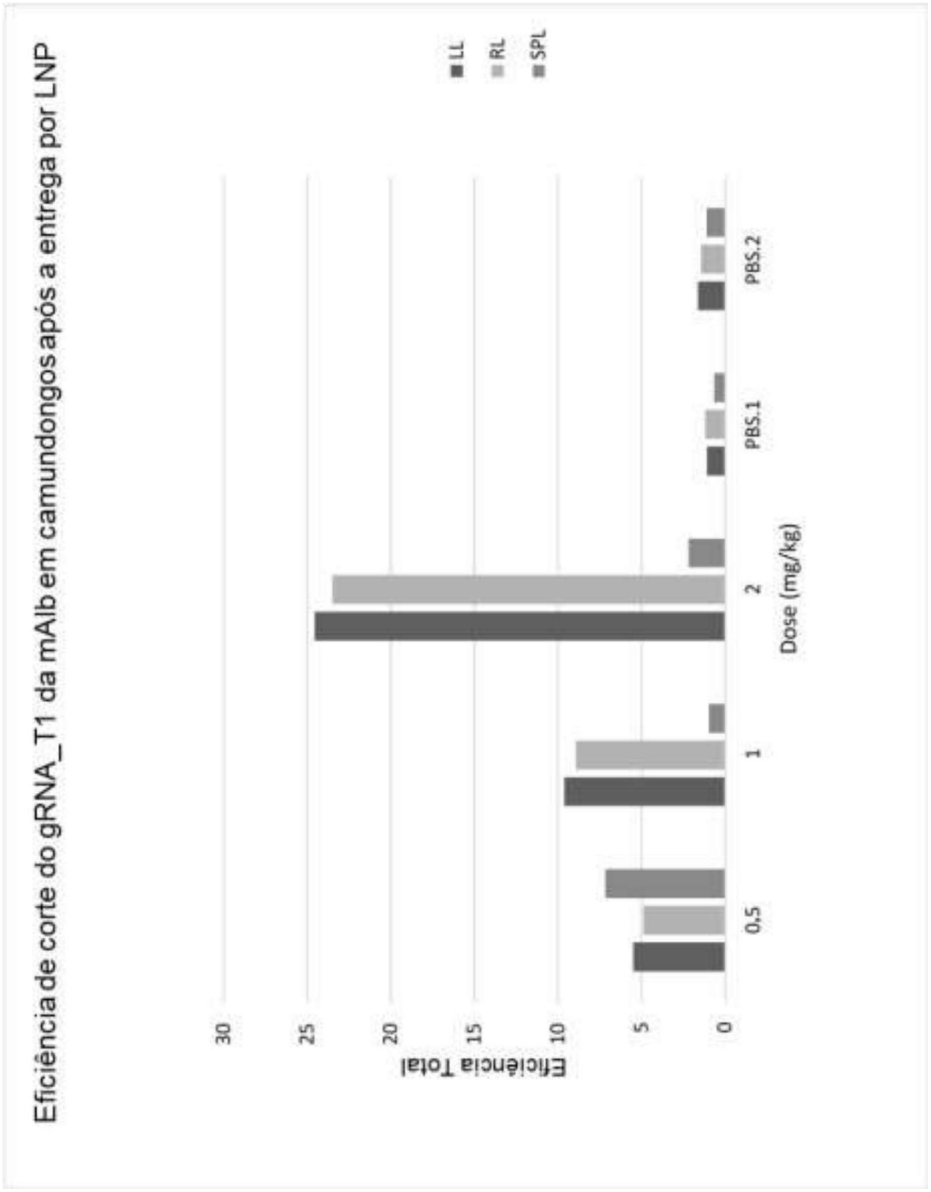

gRNA direcionado à albumina de camundongo
gRNA mALB T1, células Hepa1-6
total eff. = 89,4
Rsq = 0,9383
p >= 0,001
p < 0,001
% de sequências
80
60
40
20
0
-10
-5
0
<- deleção
inserção ->

FIG. 4


Eficiência de corte do gRNA_T1 da mAlb em camundongos após a entrega por LNP
LL
RL
SPL
Eficiência Total
30
25
20
15
10
5
0
0,5
1
2
PBS.1
PBS.2
Dose (mg/kg)

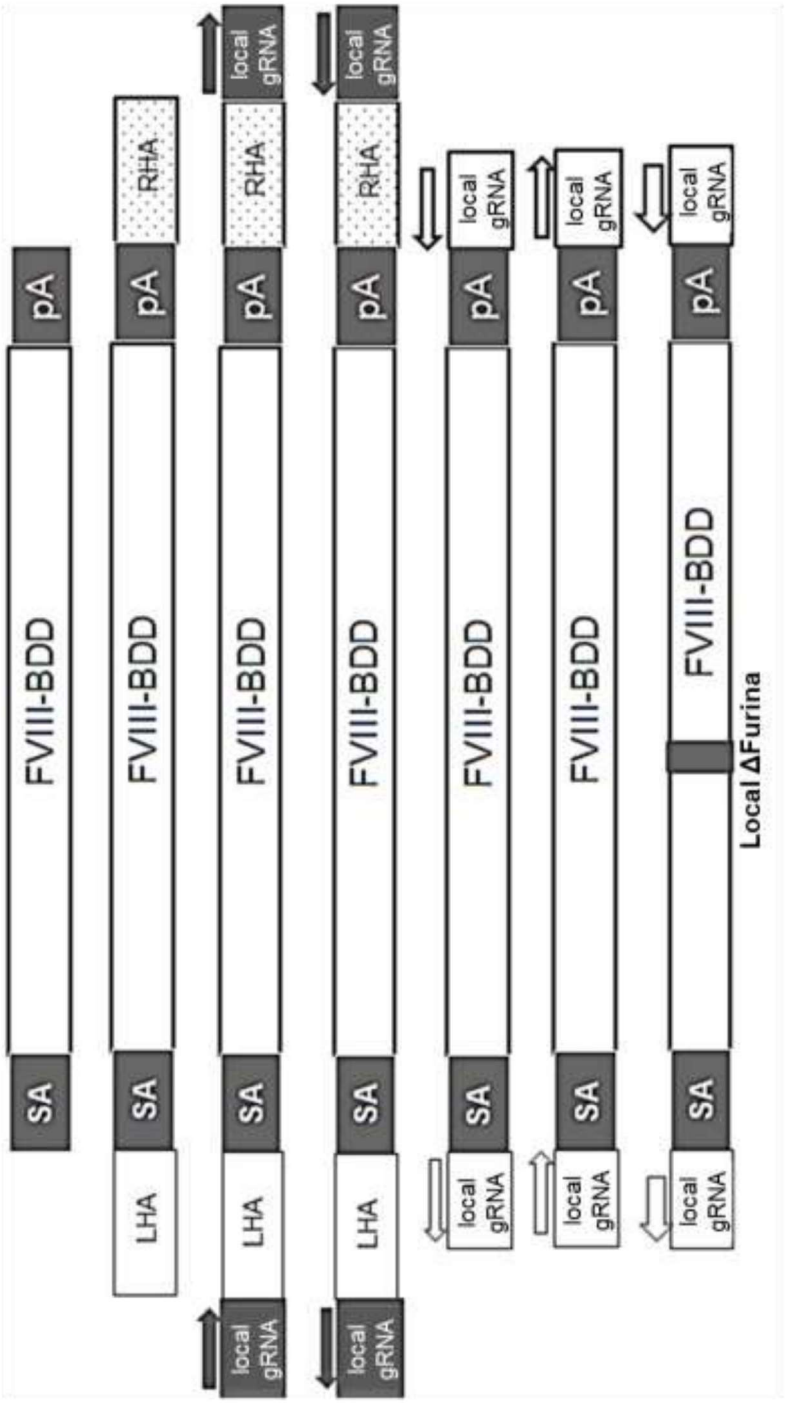


FIG. 5

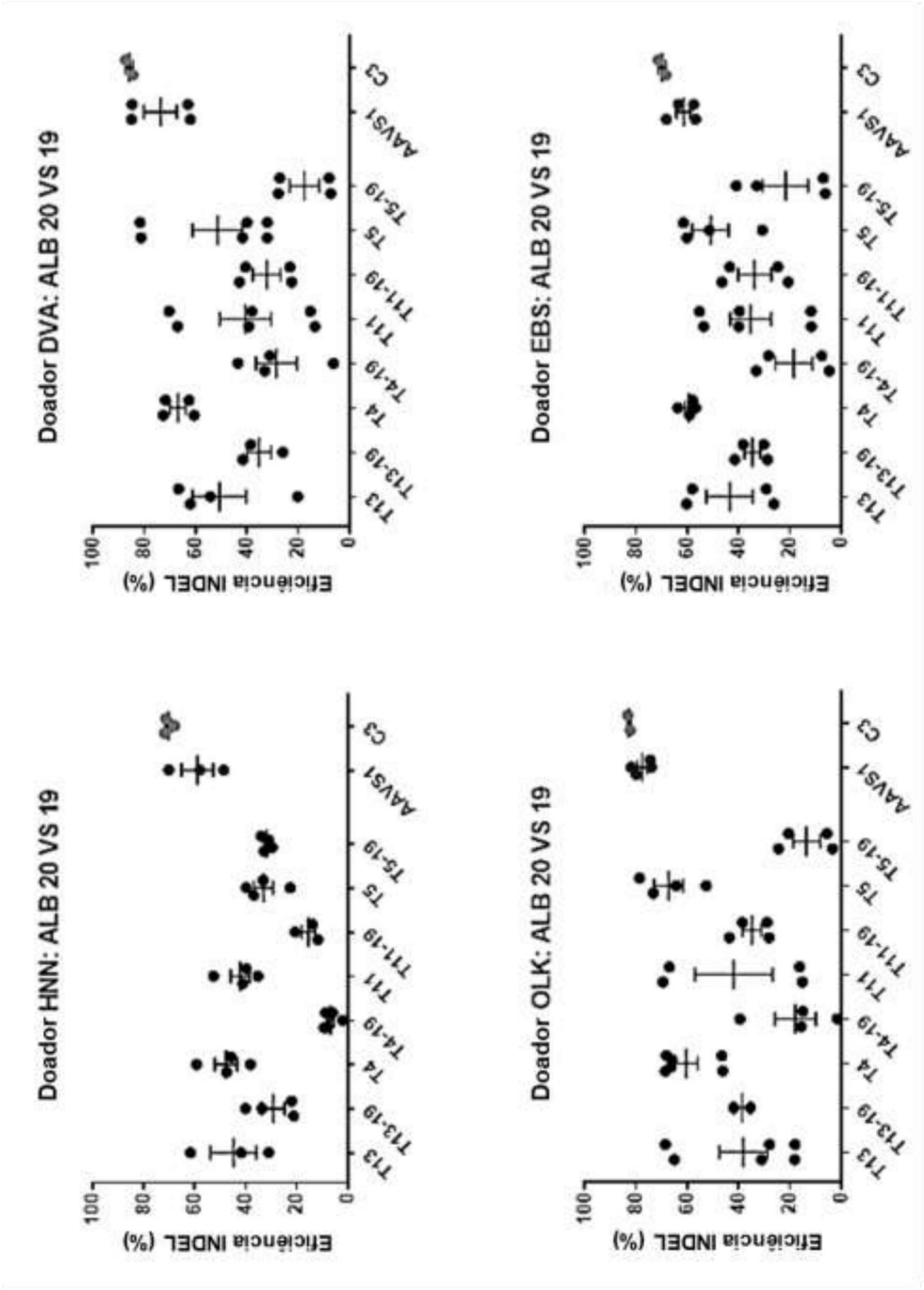

Doador DVA: ALB 20 VS 19
Doador HNN: ALB 20 VS 19
Doador EBS: ALB 20 VS 19
Doador OLK: ALB 20 VS 19
Eficiência INDEL (%)
100
80
60
40
20
0
T13
T13-19
T4
T4-19
T11
T11-19
T5
T5-19
AAVS1
C3

FIG. 6

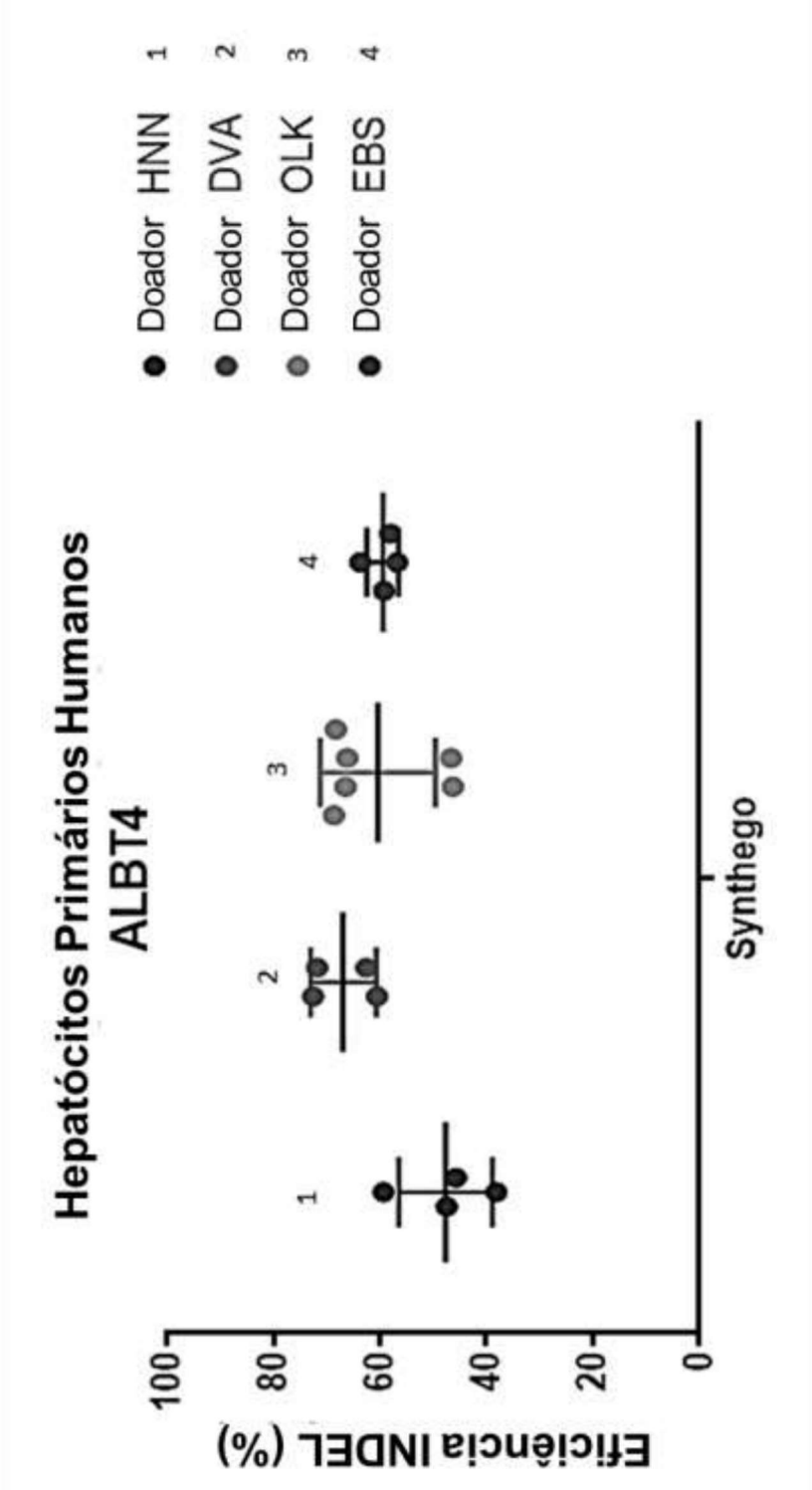

Hepatócitos Primários Humanos
ALBT4
Doador HNN 1
Doador DVA 2
Doador OLK 3
Doador EBS 4
Eficiência INDEL (%)
100
80
60
40
20
0
1
2
3
4
Synthego

FIG. 7

FIG. 7 (CONTINUAÇÃO)

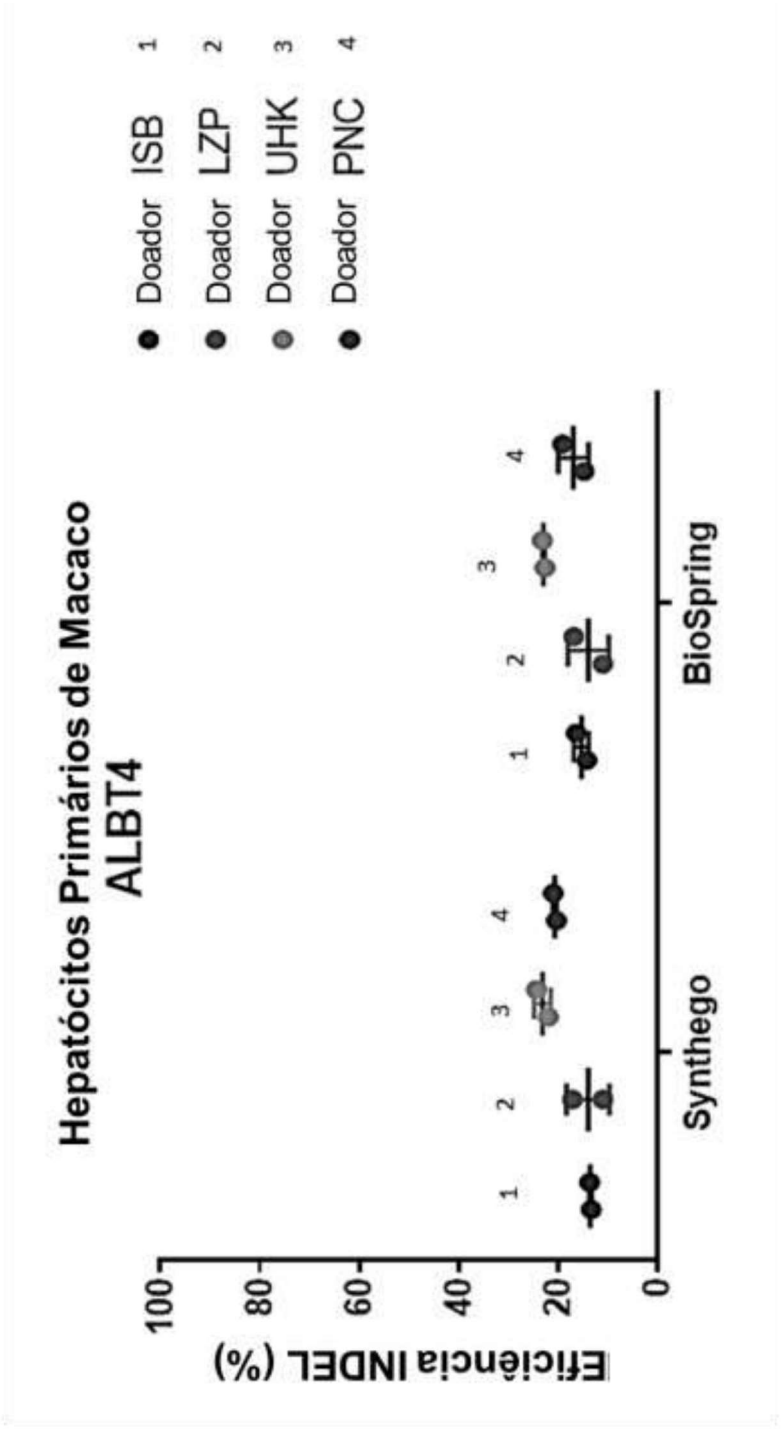
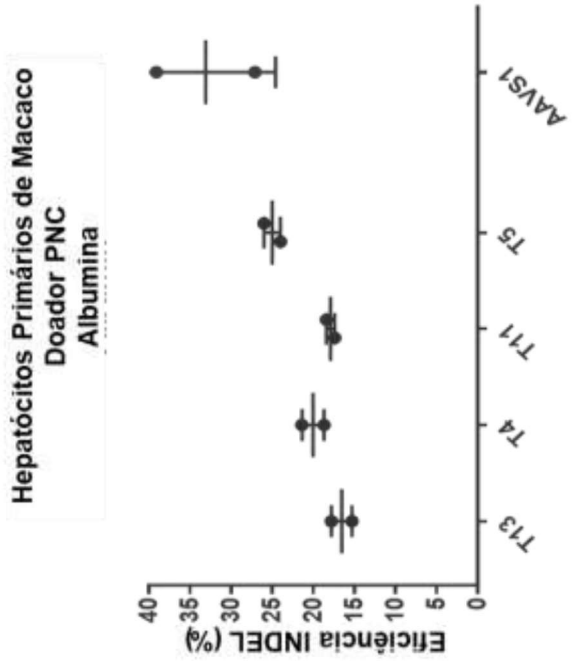


FIG. 7 (CONTINUAÇÃO)



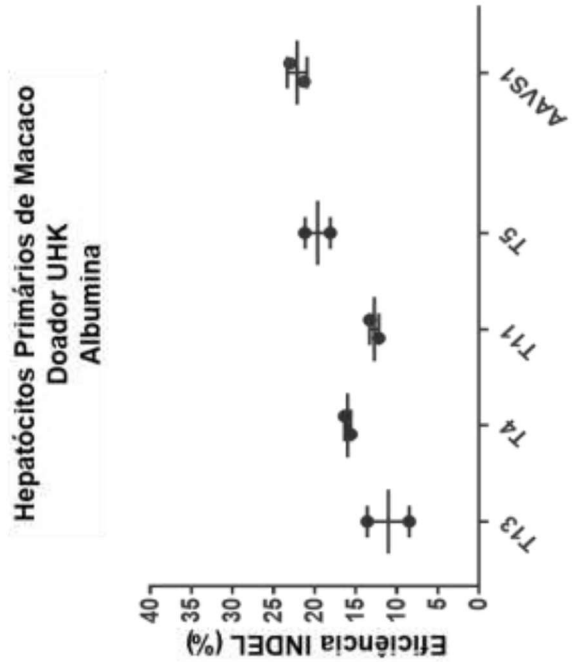


FIG. 8

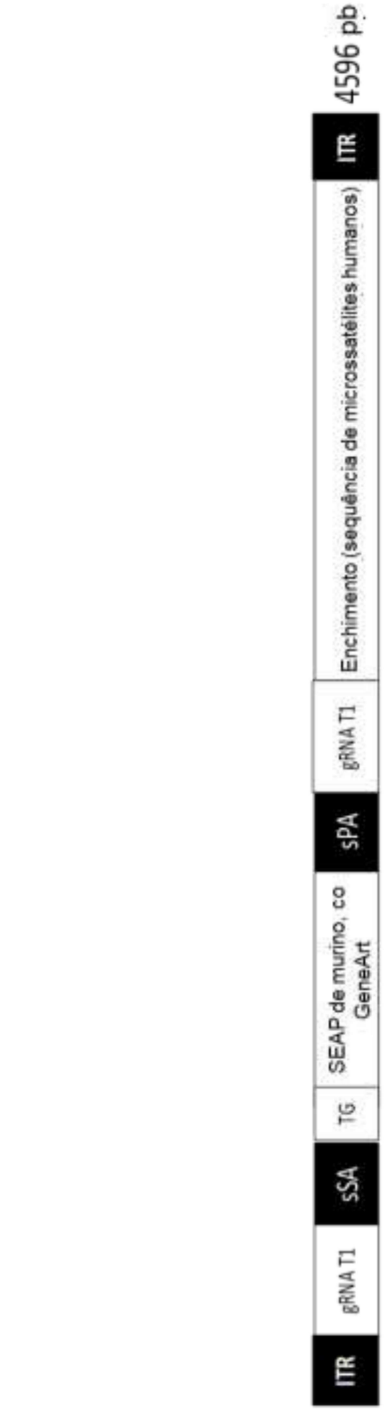
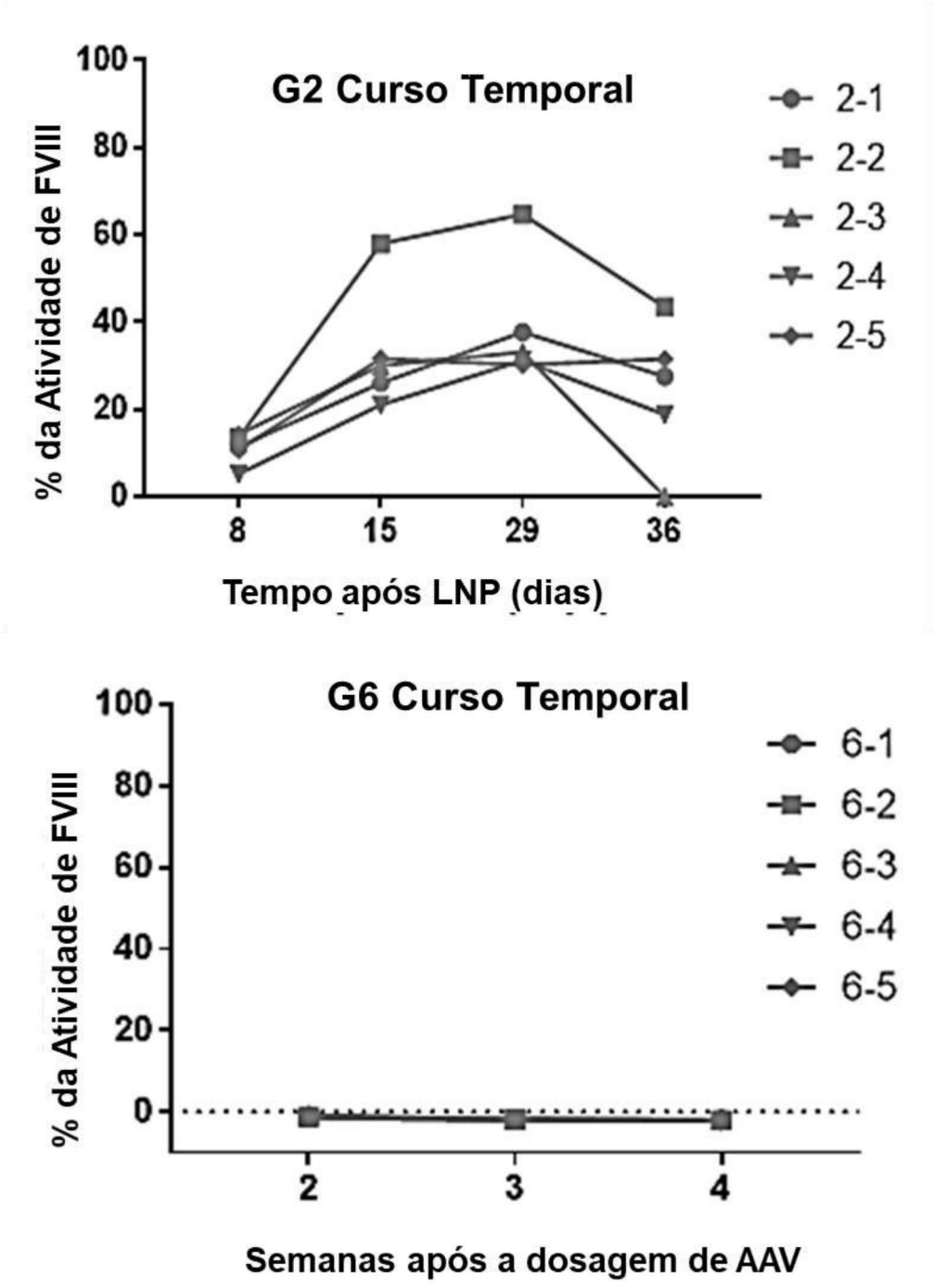

pCB0047
ITR
gRNA T1
sSA
TG
SEAP de murino, co GeneArt
sPA
gRNA T1
Enchimento (sequência de microssatélites humanos)
ITR
4596 pb

FIG. 9


pCB056
ITR
gRNA T1
mAlb SA
FVIII com domínio B deletado
UTR
sPA
gRNA T1
ITR

FIG. 10


G2 Curso Temporal
% da Atividade de FVIII
100
80
60
40
20
0
8
15
29
36
Tempo após LNP (dias)
2-1
2-2
2-3
2-4
2-5
G6 Curso Temporal
% da Atividade de FVIII
100
80
60
40
20
0
2
3
4
Semanas após a dosagem de AAV
6-1
6-2
6-3
6-4
6-5

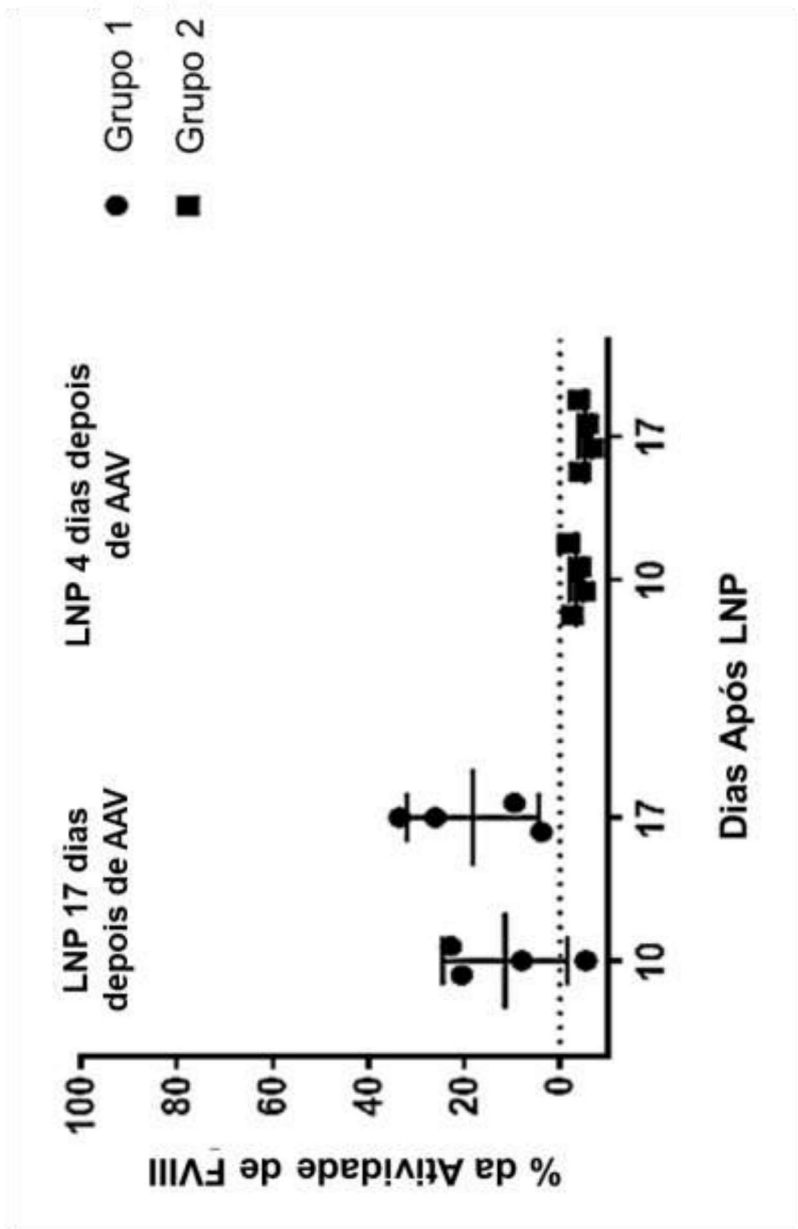

Grupo 1
Grupo 2
LNP 4 dias depois de AAV
LNP 17 dias depois de AAV
% da Atividade de FVIII
100
80
60
40
20
0
10
17
10
17
Dias Após LNP

FIG. 11

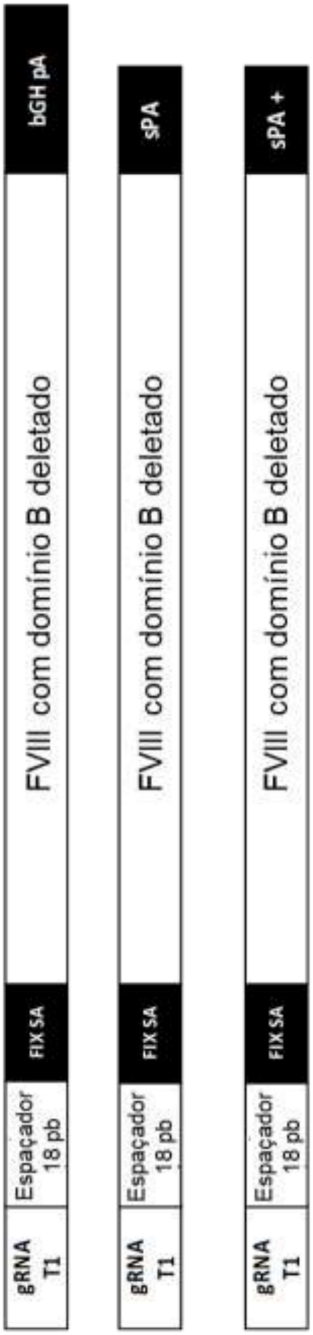
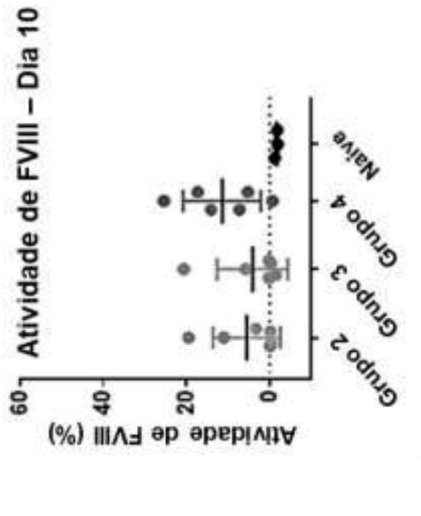
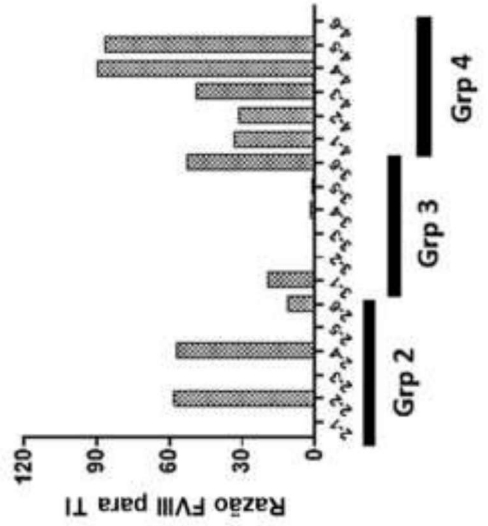

pCB065
gRNA T1
Espaçador 18 pb
FIX SA
FVIII com domínio B deletado
bGH pA
pCB076
gRNA T1
Espaçador 18 pb
FIX SA
FVIII com domínio B deletado
sPA
pCB077
gRNA T1
Espaçador 18 pb
FIX SA
FVIII com domínio B deletado
sPA +

FIG. 12

FIG. 13





Grupo		Amostra	Média do valor FVIII	Média do valor TI	Razão
Grupo 2	bGH	2-1	0.000	0.108	0.00
		2-2	11.034	0.190	58.11
		2-3	0.000	0.090	0.00
		2-4	19.468	0.339	57.40
		2-5	0.000	0.060	0.00
		2-6	3.390	0.295	11.48
Grupo 3	sPA	3-1	5.710	0.291	19.65
		3-2	0.000	0.043	0.00
		3-3	0.000	0.134	0.00
		3-4	0.175	0.097	1.81
		3-5	0.175	0.123	1.42
		3-6	20.680	0.392	52.82
Grupo 4	sPA+Espaçador	4-1	5.235	0.158	33.22
		4-2	7.238	0.230	31.48
		4-3	25.425	0.521	48.82
		4-4	17.307	0.192	89.98
		4-5	14.038	0.162	86.57
		4-6	0.000	0.125	0.00

FIG. 14


pCB0107
ITR
gRNA hT4
SSA
SEAP de murino, co GeneArt
sPA
gRNA hT4
Enchimento (sequência de microssatélites humanos)
ITR
4596 pb
pCB0156
ITR
gRNA hT4
Espaçador 18 pb
FIX SA
FVIII com domínio B deletado
sPA
gRNA hT4
ITR

FIG. 15

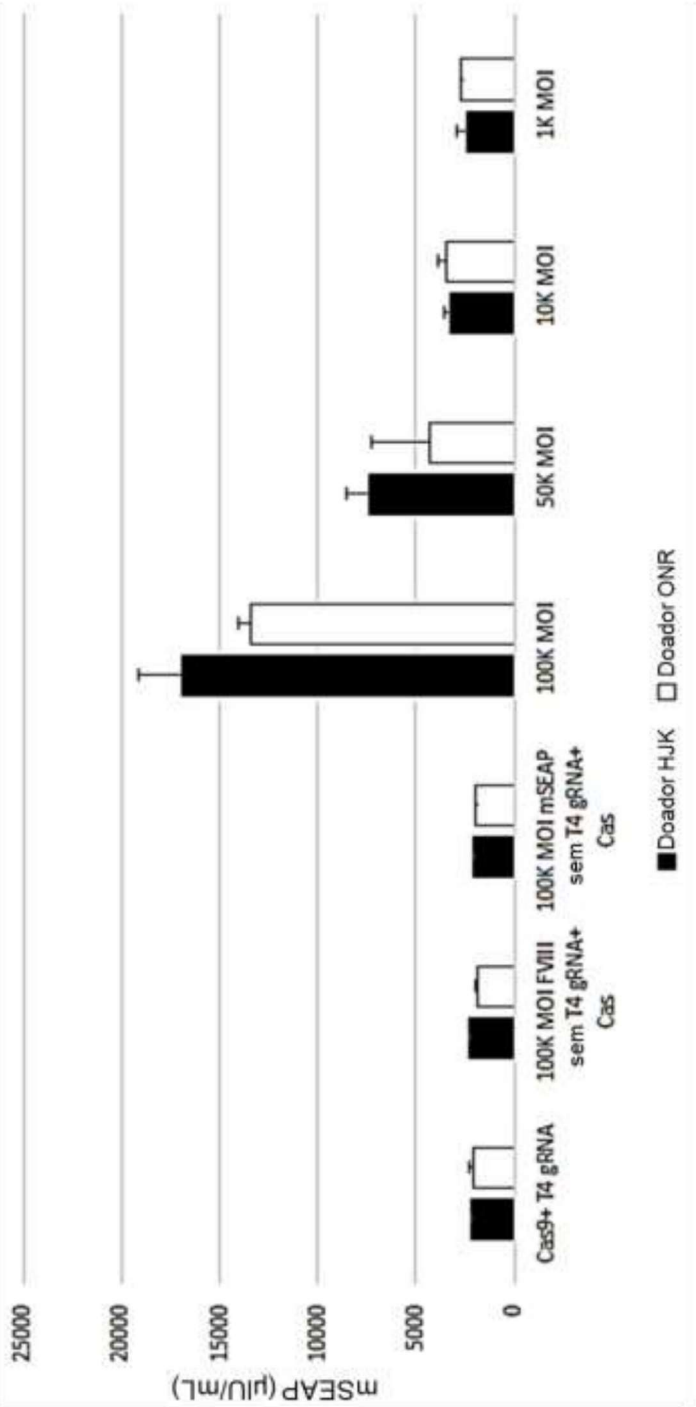
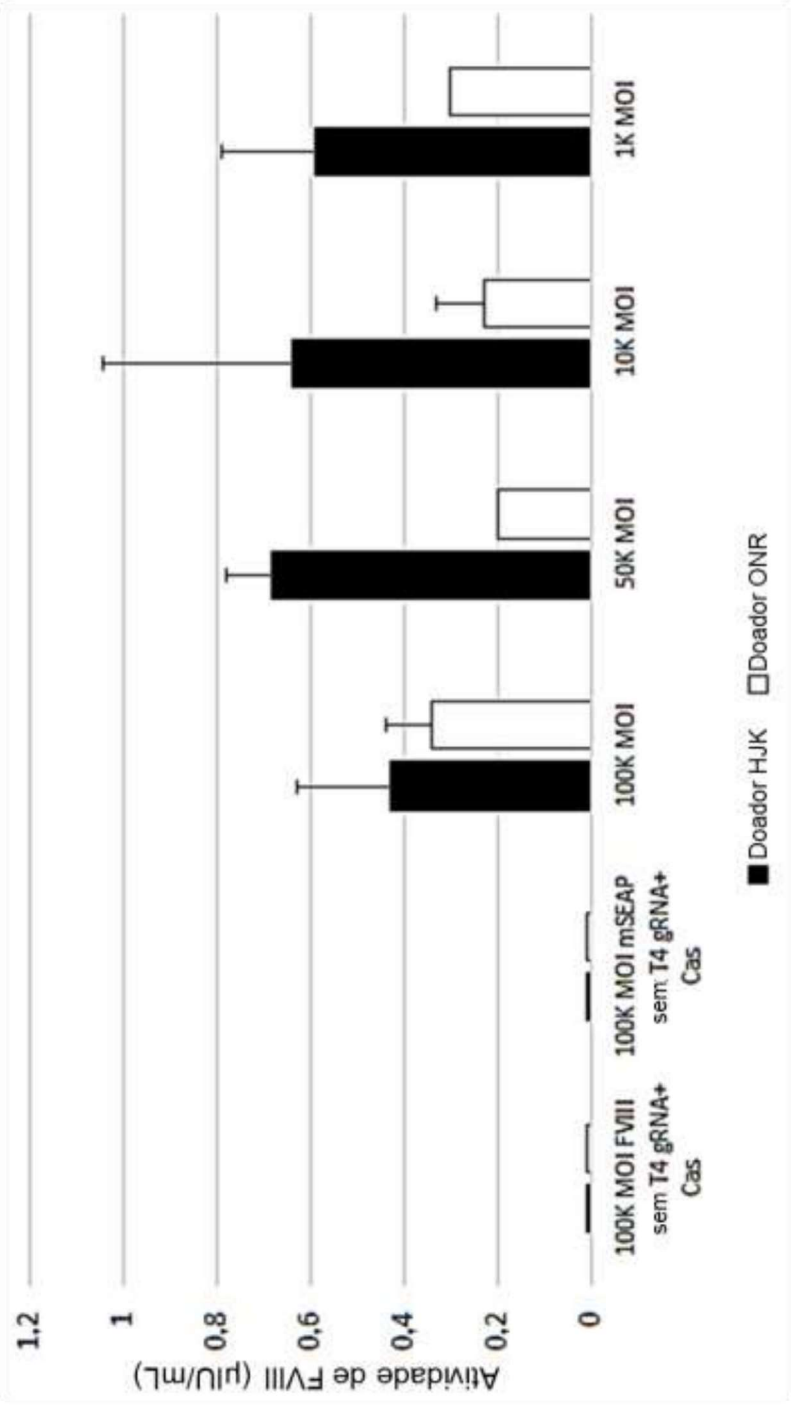

mSEAP (μIU/mL)
25000
20000
15000
10000
5000
0
Cas9+ T4 gRNA
100K MOI FVIII sem T4 gRNA+ Cas
100K MOI mSEAP sem T4 gRNA+ Cas
100K MOI
50K MOI
10K MOI
1K MOI
Doador HJK
Doador ONR

FIG. 16


Atividade de FVIII (µIU/mL)
1,2
1
0,8
0,6
0,4
0,2
0
100K MOI FVIII sem T4 gRNA+ Cas
100K MOI mSEAP sem T4 gRNA+ Cas
100K MOI
50K MOI
10K MOI
1K MOI
Doador HJK
Doador ONR

RESUMO

Patente de Invenção: **"COMPOSIÇÕES E MÉTODOS PARA EDIÇÃO GÊNICA PARA HEMOFILIA A".**

Fornecidos estão incluídos materiais e métodos para o tratamento da Hemofilia A em um sujeito *ex vivo* ou *in vivo*. Também fornecidos estão incluídos materiais e métodos para *knocking* em um gene que codifica FVIII em um genoma, em particular o lócus do gene da albumina.

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

Código de Controle

Campo 1



Campo 2



Outras Informações:

- Nome do Arquivo: LISTAGEM DE SEQUÊNCIA EMENDAS - 05-2020 -
- Data de Geração do Código: 19/05/2020
- Hora de Geração do Código: 15:20:54
- Código de Controle:
  - Campo 1: A88F4B2687297E34
  - Campo 2: 359A92A9E4664A14