

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2013.12.12	(73) Titular(es): MASSACHUSETTS INSTITUTE OF TECHNOLOGY 77 MASSACHUSETTS AVENUE CAMBRIDGE, MA 02142-1324 US PRESIDENT AND FELLOWS OF HARVARD COLLEGE US THE BROAD INSTITUTE, INC. US
(30) Prioridade(s): 2012.12.12 US 201261736527 P 2013.01.02 US 201361748427 P 2013.01.30 US 201361758468 P 2013.02.25 US 201361769046 P 2013.03.15 US 201361791409 P	
(43) Data de publicação do pedido: 2015.07.22	(72) Inventor(es): FENG ZHANG US PATRICK HSU US LE CONG US FEI RAN US
(45) Data e BPI da concessão: 2015.09.02 255/2015	(74) Mandatário: NUNO MIGUEL OLIVEIRA LOURENÇO RUA CASTILHO, Nº 50 - 9º 1269-163 LISBOA PT

(54) Epígrafe: **ENGENHARIA DE SISTEMAS, MÉTODOS E COMPOSIÇÕES GUIA OTIMIZADAS PARA A MANIPULAÇÃO DE SEQUÊNCIAS**

(57) Resumo:

A INVENÇÃO PREVÊ SISTEMAS, MÉTODOS E COMPOSIÇÕES PARA A MANIPULAÇÃO DE SEQUÊNCIAS E/OU ATIVIDADE DE SEQUÊNCIAS ALVO. SÃO PREVISTOS VETORES E SISTEMAS DE VETOR, ALGUNS DOS QUAIS CODIFICAM UM OU MAIS COMPONENTES DE UM COMPLEXO CRISPR, ASSIM COMO MÉTODOS PARA A CRIAÇÃO E UTILIZAÇÃO DE TAIS VETORES. SÃO TAMBÉM PREVISTOS MÉTODOS PARA O CONTROLO DA FORMAÇÃO DO COMPLEXO CRISPR EM CÉLULAS EUCARIÓTICAS E MÉTODOS PARA A SELEÇÃO DE CÉLULAS ESPECÍFICAS ATRAVÉS DA INTRODUÇÃO DE MUTAÇÕES PRECISAS UTILIZANDO O SISTEMA CRISPR-CAS.

RESUMO

"ENGENHARIA DE SISTEMAS, MÉTODOS E COMPOSIÇÕES GUIA OTIMIZADAS PARA A MANIPULAÇÃO DE SEQUÊNCIAS"

A invenção prevê sistemas, métodos e composições para a manipulação de sequências e/ou atividade de sequências alvo. São previstos vetores e sistemas de vetor, alguns dos quais codificam um ou mais componentes de um complexo CRISPR, assim como métodos para a criação e utilização de tais vetores. São também previstos métodos para o controle da formação do complexo CRISPR em células eucarióticas e métodos para a seleção de células específicas através da introdução de mutações precisas utilizando o sistema CRISPR-Cas.

DESCRIÇÃO

"ENGENHARIA DE SISTEMAS, MÉTODOS E COMPOSIÇÕES GUIA OTIMIZADAS PARA A MANIPULAÇÃO DE SEQUÊNCIAS"

Campo da Invenção

A presente invenção refere-se em geral a sistemas, métodos e composições utilizados para o controlo da expressão de genes envolvendo o direccionamento de sequências, tais como perturbação do genoma ou edição de genes, que podem utilizar sistemas de vetores relacionados com Repetições Palindrómicas Curtas Agrupadas e Regularmente Espaçadas (CRISPR, do inglês *Clustered Regularly Interspaced Short Palindromic Repeats*) e os seus componentes.

Declaração sobre Investigação Patrocinada pelo Governo Federal

Esta invenção foi realizada com o apoio concedido pelo governo através do National Institute of Health, NIH Pioneer Award DP1MH100706. O governo tem certos direitos sobre a invenção.

Antecedentes da Divulgação

Recentes avanços nas técnicas de sequenciação do genoma e métodos de análise aceleraram significativamente a capacidade de catalogar e mapear os fatores genéticos associados a uma gama variada de funções biológicas e doenças. São necessárias tecnologias de direccionamento genómico precisas para permitir a engenharia reversa sistemática das variações genéticas causais, permitindo a

alteração seletiva de elementos genéticos individuais, bem como para o avanço da biologia sintética, biotecnologia, e aplicações médicas. Embora as técnicas de edição do genoma, tais como desenho de dedos de zinco, efetores de tipo ativador de transcrição (TALEs, do inglês *transcription activator-like effectors*), ou *homing* meganucleases, estejam disponíveis para a produção de alterações direcionadas no genoma, permanece a necessidade de novas tecnologias de engenharia do genoma que sejam acessíveis, fáceis de instalar, escaláveis e passíveis de direcionar a várias posições dentro do genoma eucariota.

Sumário da Invenção

Existe uma necessidade premente de sistemas e técnicas alternativas e robustas para o direcionamento de sequências com um vasto leque de aplicações. A presente invenção aborda esta necessidade e prevê vantagens relacionadas. CRISPR/Cas ou o sistema CRISPR-Cas (os dois termos são utilizados intermutavelmente ao longo deste pedido) não requer a geração de proteínas personalizadas para as sequências alvo específicas, mas uma única enzima Cas que pode ser programada por uma molécula de RNA curto para reconhecer um específico DNA alvo, ou por outras palavras, a enzima Cas pode ser recrutada para um alvo específico do DNA utilizando a referida molécula de RNA curto. Adicionando o sistema CRISPR-Cas ao repertório de técnicas de sequenciação do genoma e métodos de análise pode simplificar significativamente a metodologia e acelerar a capacidade de catalogar e mapear os fatores genéticos associados a uma variada gama de funções biológicas e doenças. Para utilizar o sistema CRISPR-Cas de forma eficaz para a edição do genoma sem efeitos deletérios, é

fundamental compreender os aspetos de engenharia e otimização destas ferramentas de engenharia do genoma, que são aspetos da presente divulgação.

De acordo com a presente invenção, é previsto um sistema de vetor CRISPR-Cas de origem não natural e obtido por engenharia de acordo com a reivindicação 1. Outros aspetos da invenção são divulgados nas reivindicações dependentes. A invenção também prevê a utilização do sistema de acordo com as reivindicações 19 e 20.

Num aspeto, a presente divulgação prevê um sistema de vetor compreendendo um ou mais vetores. Em algumas formas de realização, o sistema compreende: (a) um primeiro elemento regulador ligado operacionalmente a uma sequência de emparelhamento tracr e um ou mais locais de inserção para a inserção de uma ou mais sequências guia a montante da sequência de emparelhamento tracr, em que, quando expressa, a sequência guia direciona a ligação específica da sequência de um complexo CRISPR a uma sequência alvo numa célula, por exemplo, célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR num complexo com (1) a sequência guia que hibrida com a sequência alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr; e (b) um segundo elemento regulador, ligado operacionalmente a uma sequência que codifica enzimas que codificam a referida enzima CRISPR compreendendo uma sequência de localização nuclear; em que os componentes (a) e (b) estão localizados no mesmo ou em diferentes vetores do sistema. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência de emparelhamento tracr sob o controlo do primeiro elemento regulador. Em algumas formas

de realização, o componente (a) compreende adicionalmente duas ou mais sequências guia ligadas de forma operativa ao primeiro elemento regulador, em que, quando expressas, cada uma das duas ou mais sequências guia orienta a ligação específica de um complexo CRISPR de uma sequência alvo diferente numa célula eucariótica. Em algumas formas de realização, o sistema compreende a sequência tracr sob o controlo de um terceiro elemento regulador, tal como um promotor de polimerase III. Em algumas formas de realização, a sequência tracr apresenta, pelo menos, 50%, 60%, 70%, 80%, 90%, 95%, ou 99% de complementaridade de sequência ao longo do comprimento da sequência de emparelhamento tracr quando otimamente alinhadas. Em algumas formas de realização, o complexo CRISPR compreende uma ou mais sequências de localização nuclear de resistência suficiente para dirigir a acumulação do referido complexo CRISPR numa quantidade detetável no núcleo de uma célula eucariótica. Sem se pretender ficar limitado pela teoria, acredita-se que um sinal de localização nuclear não é necessário para a atividade do complexo CRISPR em eucariotas, mas que incluindo tais sequências aumenta a atividade do sistema, especialmente para direccionar as moléculas de ácido nucleico no núcleo. Em algumas formas de realização, a enzima CRISPR é uma enzima do sistema CRISPR de tipo II. Em algumas formas de realização, a enzima CRISPR é uma enzima Cas9. Em algumas formas de realização, a enzima Cas9 é uma Cas9 de *S. pneumoniae*, *S. pyogenes*, ou *S. thermophilus*, e pode incluir uma Cas9 mutada derivada desses organismos. A enzima pode ser um homólogo ou ortólogo da Cas9. Em algumas formas de realização, a enzima CRISPR tem codões otimizados para expressão numa célula eucariótica. Em algumas formas de realização, a enzima CRISPR orienta a clivagem de uma ou

duas cadeias na posição da sequência alvo. Em algumas formas de realização, a enzima CRISPR carece de atividade de clivagem da cadeia de DNA. Em algumas formas de realização, o primeiro elemento regulador é um promotor de polimerase III. Em algumas formas de realização, o segundo elemento regulador é um promotor de polimerase II. Em algumas formas de realização, a sequência guia tem de comprimento, pelo menos, 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10-30, ou entre 15-25, ou entre 15-20 nucleótidos. Em geral, e ao longo desta especificação, o termo "vetor" refere-se a uma molécula de ácido nucleico capaz de transportar outro ácido nucleico ao qual foi ligada. Os vetores incluem, mas não estão limitados a, moléculas de ácido nucleico que são de cadeia simples, cadeia dupla, ou parcialmente de cadeia dupla; moléculas de ácidos nucleicos que compreendem uma ou mais extremidades livres, nenhuma extremidades livres (por exemplo circular); moléculas de ácido nucleico que compreendem DNA, RNA, ou ambas; e outras variedades de polinucleótidos conhecidos na técnica. Um tipo de vetor é um "plasmídeo", que se refere a uma alça de DNA de cadeia dupla circular ao qual podem ser inseridos segmentos de DNA adicionais, por exemplo, por técnicas de clonagem molecular padrão. Um outro tipo de vetor é um vetor viral, em que as sequências derivadas de DNA ou RNA viral estão presentes no vetor para empacotamento num vírus (por exemplo, retrovírus, retrovírus de replicação defeituosa, adenovírus, adenovírus de replicação defeituosa, e vírus adeno-associados). Os vetores virais também incluem polinucleótidos transportados por um vírus para a transfecção numa célula hospedeira. Determinados vetores são capazes de replicação autónoma numa célula hospedeira na qual são introduzidos (por exemplo, vetores bacterianos que têm uma origem de

replicação bacteriana e vetores epissomais de mamíferos). Outros vetores (por exemplo, vetores não epissomais de mamíferos) são integrados no genoma de uma célula hospedeira quando da introdução na célula hospedeira, e assim são replicados juntamente com o genoma do hospedeiro. Além disso, determinados vetores são capazes de dirigir a expressão de genes aos quais estão operacionalmente ligados. Estes vetores são aqui referidos como "vetores de expressão". Os vetores de expressão mais comuns com utilidade em técnicas de DNA recombinante estão muitas vezes na forma de plasmídeos.

Os vetores de expressão recombinante podem compreender um ácido nucleico da divulgação numa forma adequada para expressão do ácido nucleico numa célula hospedeira, o que significa que os vetores de expressão recombinantes incluem um ou mais elementos de regulação, que podem ser selecionados com base nas células hospedeiras a ser utilizado para a expressão, o que está operacionalmente ligado à sequência de ácido nucleico a ser expressa. Dentro de um vetor de expressão recombinante, "operativamente ligado" pretende significar que a sequência de nucleótidos de interesse é ligada ao elemento ou aos elementos reguladores de um modo que permite a expressão da sequência de nucleótidos (por exemplo, num sistema *in vitro* de transcrição/tradução ou numa célula hospedeira quando o vetor é introduzido na célula hospedeira).

O termo "elemento regulador" destina-se a incluir promotores, potenciadores (*enhancers*), locais internos de entrada do ribossoma (IRES), e outros elementos de controlo da expressão (por exemplo, sinais de terminação da transcrição, tais como sinais de poliadenilação e

sequências de poli-U). Tais elementos reguladores encontram-se descritos, por exemplo, em Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Califórnia (1990). Os elementos reguladores incluem aqueles que controlam a expressão constitutiva de uma sequência de nucleótidos em muitos tipos de células hospedeiras e os que controlam a expressão da sequência nucleotídica apenas em determinadas células hospedeiras (por exemplo, sequências reguladoras específicas de tecido). Um promotor específico de tecido pode controlar a expressão principalmente num tecido de interesse desejado, tal como músculo, neurónios, osso, pele, sangue, órgãos específicos (por exemplo, fígado, pâncreas), ou tipos de células em particular (por exemplo linfócitos). Os elementos reguladores podem também controlar a expressão de uma forma dependente de tempo, tal como de um modo dependente do ciclo celular ou da fase de desenvolvimento, o que também pode ou pode não ser específico do tecido ou tipo de célula. Em algumas formas de realização, um vetor compreende um ou mais promotor de pol III (por exemplo, 1, 2, 3, 4, 5, ou mais promotores de pol III), um ou mais promotores de pol II (por exemplo, 1, 2, 3, 4, 5, ou mais promotores de pol II), um ou mais promotores de pol I (por exemplo, 1, 2, 3, 4, 5, ou mais promotores de pol I), ou combinações dos mesmos. Exemplos de promotores de pol III incluem, mas não estão limitados a, promotores U6 e H1. Exemplos de promotores de pol II incluem, mas não estão limitados a, promotor LTR retroviral do vírus do sarcoma de Rous (RSV) (opcionalmente com o potenciador RSV), o citomegalovírus (CMV) (opcionalmente com o potenciador CMV) [ver, por exemplo, Boshart *et al*, Cell, 41:521-530 (1985)], o promotor SV40, o promotor dihidrofolato redutase, o promotor β -actina, o promotor

fosfoglicerol cinase (PGK) e o promotor EF1 α . Também englobados pelo termo "elemento regulador" são elementos potenciadores, tais como WPRE; potenciadores CMV; o segmento R-U5' em LTR de HTLV-I (Mol. Cell. Biol., Vol.8(1), p.466-472, 1988); potenciador SV40; e a sequência do intrão entre os exões 2 e 3 da β -globina de coelho (Proc. Natl. Acad. Sci. USA., Vol.78(3), p.1527-31, 1981). Será apreciado pelos peritos na técnica que o desenho do vetor de expressão pode depender de fatores tais como a escolha da célula hospedeira a ser transformada, o nível de expressão desejado, etc. Um vetor pode ser introduzido nas células hospedeiras para assim produzir transcritos, proteínas ou péptidos, incluindo proteínas ou péptidos de fusão, codificados por ácidos nucleicos como aqui descrito (por exemplo, transcritos de Repetições Palindrômicas Curtas Agrupadas e Regularmente Espaçadas (CRISPR), proteínas, enzimas, formas mutantes dos mesmos, proteínas de fusão dos mesmos, etc.).

Os vetores vantajosos incluem os lentivírus e os vírus adeno-associados, e também podem ser selecionados tipos de tais vetores para direcionar tipos de células em particular.

Num aspeto, a presente divulgação prevê um vetor compreendendo um elemento regulador ligado operativamente a uma sequência que codifica enzimas que codificam a enzima CRISPR compreendendo uma ou mais sequências de localização nuclear. Em algumas formas de realização, tais elementos reguladores conduzem à transcrição da enzima CRISPR numa célula eucariótica de tal forma que a referida enzima CRISPR acumula numa quantidade detetável no núcleo da célula eucariótica. Em algumas formas de realização, o

Metabolismo da Frutose e Manose	HK2; GCK; HK I
Metabolismo da Galactose	HK2; GCK; HK1
Biossíntese do Estilbeno, Cumarina e Lignina	PRDX6; PRDX1; TYR
Via de Apresentação do Antigene	CALR; B2M
Biossíntese de Esteroides	NQO1; DHCR7
Metabolismo do Butirato	ALDH1A1; NLGN1
Ciclo do Citrato	IDH2; IDH1
Metabolismo de Ácidos Gordos	ALDH1A1; CYP1B1
Metabolismo de Glicerofosfolípidos	PRDX6; CHKA
Metabolismo da Histidina	PRMT5; ALDH1A1
Metabolismo do Inositol	ERO1L; APEX1
Metabolismo de Xenobióticos pelo Citocromo p450	GSTP1; CYP1B1
Metabolismo do Metano	PRDX6; PRDX1
Metabolismo da Fenilalanina	PRDX6; PRDX1
Metabolismo do Propanoato	ALDH1A1; LDHA
Metabolismo de Aminoácidos com Selénio	PRMT5; AHCY
Metabolismo de Esfingolípidos	SPHK1; SPHK2
Metabolismo de Aminofosfonato	PRMT5
Metabolismo de Androgénio e Estrogénio	PRMT5
Metabolismo do Ascorbato e do Ácido Aldárico	ALDH1A1
Biossíntese de Ácidos Biliares	ALDH1A1
Metabolismo da Cisteína	LDHA
Biossíntese dos Ácidos Gordos	FASN
Via de sinalização do Recetor do Glutamato	GNB2L1
Resposta ao Stress Oxidativo mediada pelo	PRDX1

NRF2	
Via das Pentose Fosfato	GPI
Interconversão de Pentose e Glucuronato	UCHL1
Metabolismo do Retinol	ALDH1A1
Metabolismo da Riboflavina	TYR
Metabolismo da Tirosina	PRMT5, TYR
Biossíntese da Ubiquinona	PRMT5
Degradação da Valina, Leucina e Isoleucina	ALDH1A1
Metabolismo da Glicina, Serina e Treonina	CHKA
Degradação da Lisina	ALDH1A1
Dor/Sabor	TRPM5; TRPA1
Dor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkarla; Prkar2a
Função Mitochondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurologia de Desenvolvimento	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenina; Dkk-1; proteínas relacionadas com Frizzled; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4f1 ou Brn3a); Numb; Reln

As formas de realização da presente divulgação também se relacionam com métodos e composições relacionadas com inativação (*knocking out*) de genes, ampliação de genes e reparação de mutações específicas associadas com a instabilidade da repetição no DNA e distúrbios neurológicos (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Segunda Edição, Academic Press, 13 de outubro de 2011 - Medical). Encontrou-se aspectos específicos das sequências de repetição em tandem que são

responsáveis por mais de vinte doenças humanas (New insights into repeat instability: role of RNA·DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. Setembro-Outubro 2010; 7(5):551-8). O sistema CRISPR-Cas pode ser aproveitado para corrigir esses defeitos de instabilidade genómica.

Um outro aspeto da presente divulgação refere-se à utilização do sistema CRISPR-Cas para a edição de defeitos nos genes EMP2A e EMP2B que foram identificados como estando associados com a doença de Lafora. A doença de Lafora é uma condição autossómica recessiva que se caracteriza por epilepsia mioclónica progressiva que pode começar como convulsões epiléticas na adolescência. Em poucos casos da doença a causa pode estar em mutações em genes ainda não identificados. A doença provoca convulsões, espasmos musculares, dificuldade para caminhar, demência e, eventualmente, a morte. Não existe atualmente nenhuma terapia que se tenha mostrado eficaz contra a progressão da doença. Outras anormalidades genéticas associadas à epilepsia também pode ser alvo do sistema CRISPR-Cas e a genética subjacente é descrita adicionalmente em *Genetics of Epilepsy and Genetic Epilepsies*, editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology: 20; 2009).

Num outro aspeto da presente divulgação, o sistema CRISPR-Cas pode ser usado para corrigir defeitos oculares que surgem devido a diversas mutações genéticas que se descrevem adicionalmente em *Genetic Diseases of the Eye*, Segunda Edição, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Vários aspetos adicionais da presente divulgação relacionam-se com a edição de defeitos associados com uma vasta gama de doenças genéticas que são descritas adicionalmente no sítio Web dos Institutos Nacionais de Saúde na subseção sob o tópico Doenças Genéticas. As doenças cerebrais genéticas podem incluir, mas não estão limitadas a Adrenoleucodistrofia, Agenesia do Corpo Caloso, Síndrome de Aicardi, Doença de Alpers, Doença de Alzheimer, Síndrome de Barth, Doença de Batten, CADASIL, Degeneração Cerebelar, Doença de Fabry, Doença de Gerstmann-Straussler-Scheinker, Doença de Huntington e de outras Desordens de Repetição de Tripletos, Doença de Leigh, Síndrome de Lesch-Nyhan, Doença de Menkes, Miopatias Mitocondriais e Colpocefalia NINDS. Estas doenças são descritas com detalhes adicionais no site Web dos Institutos Nacionais de Saúde na subseção sob o tópico Doenças Genéticas do Cérebro.

Em algumas formas de realização, a condição pode ser a neoplasia. Em algumas formas de realização, em que a condição é neoplasia, os genes alvo são qualquer um dos listados na Tabela A (neste caso PTEN, e assim por diante). Em algumas formas de realização, a condição pode ser Degeneração Macular Relacionada com a Idade. Em algumas formas de realização, a condição pode ser a Doença Esquizofrenia. Em algumas formas de realização, a condição pode ser um Distúrbio de Repetições de Trinucleótidos. Em algumas formas de realização, a condição pode ser o Síndrome de X frágil. Em algumas formas de realização, a condição pode ser um Distúrbio Relacionado com a Secretase. Em algumas formas de realização, a condição pode ser distúrbio relacionado com Priões. Em algumas formas de realização, a condição pode ser ALS. Em algumas formas de

realização, a condição pode ser uma dependência de drogas. Em algumas formas de realização, a condição pode ser autismo. Em algumas formas de realização, a condição pode ser a Doença de Alzheimer. Em algumas formas de realização, a condição pode ser a inflamação. Em algumas formas de realização, a condição pode ser a Doença de Parkinson.

Exemplos de proteínas associadas com a doença de Parkinson incluem, mas não estão limitados a α -sinucleína, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Sinfilina-1, e NURR1.

Os exemplos de proteínas relacionadas com dependência podem incluir ABAT por exemplo.

Os exemplos de proteínas relacionadas com a inflamação podem incluir a proteína quimioatrativa de monócitos 1 (MCP1) codificada pelo gene CCR2, o recetor C-C de quimiocinas de tipo 5 (CCR5) codificada pelo gene CCR5, o recetor IIB de IgG (FCGR2b, também denominado CD32) codificado pelo gene Fcgr2b, ou a proteína Fc épsilon R1g (FCER1g) codificada pelo gene Fcer1g, por exemplo.

Exemplos de proteínas associadas a doenças cardiovasculares podem incluir IL1B (interleucina-1, beta), XDH (xantina desidrogenase), TP53 (proteína tumoral p53), PTGIS (sintase da prostaglandina I2 (prostaciclina)), MB (mioglobina), IL4 (interleucina 4), ANGPT1 (angiopoietina 1), ABCG8 (cassete de ligação a ATP, subfamília G (WHITE), membro 8), ou CTSK (catepsina K), por exemplo.

Exemplos de proteínas associadas à doença de Alzheimer podem incluir a proteína recetora de lipoproteínas de muito baixa densidade (VLDLR) codificada pelo gene de VLDLR, a

enzima 1 do modificador de tipo ubiquitina (UBA1) codificada pelo gene UBA1, ou a proteína de subunidade catalítica da enzima E1 ativadora de NEDD8 (UBE1C) codificada pelo gene UBA3, por exemplo.

Exemplos de proteínas associadas com Perturbações do Espectro Autista podem incluir a proteína 1 associada ao recetor de benzodiazepina (periférica) (BZRAP1) codificada pelo gene BZRAP1, a proteína membro 2 da família AF4/FMR2 (AFF2) codificada pelo gene AFF2 (também denominado MFR2), a proteína do homólogo 1 autossómico de atraso mental pelo cromossoma X frágil (FXR1) codificada pelo gene FXR1, ou a proteína do homólogo 2 autossómico de atraso mental pelo cromossoma X frágil (FXR2) codificada pelo gene FXR2, por exemplo.

Exemplos de proteínas associadas com a Degeneração Macular podem incluir a proteína 4 (ABCA4), membro da cassete de ligação a ATP, subfamília A (ABC1), codificada pelo gene ABCR, a proteína apolipoproteína E (APOE), codificada pelo gene da APOE, ou a proteína do Ligando 2 (CCL2) da quimiocina (motivo C-C) codificada pelo gene CCL2, por exemplo.

Exemplos de proteínas associadas com a Esquizofrenia podem incluir NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, Disc1, GSK3B, e suas combinações.

Exemplos de proteínas envolvidas na supressão de tumores podem incluir ATM (Ataxia telangiectasia mutada), a ATR (Ataxia Telangiectasia e relacionada com Rad3), EGFR (recetor do fator de crescimento epidérmico), ERBB2 (homólogo 2 do oncogene viral de leucemia eritroblástica v-

erb-b2), ErbB3 (homólogo 3 do oncogene viral de leucemia eritroblástica v-erb-b2), ErbB4 (homólogo 4 do oncogene viral de leucemia eritroblástica v-erb-b2), Notch1, Notch2, Notch3, ou Notch4, por exemplo.

Exemplos de proteínas associadas a uma desordem da secretase podem incluir PSENEN (homólogo do potenciador 2 da presenilina (*C. elegans*)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (proteína precursora (A4) de beta amiloide), APOE4 (homólogo B da faringe defeituosa anterior 1 (*C. elegans*)), PSEN2 (presenilina 2 (doença de Alzheimer 4)), ou de BACE1 (enzima 1 de clivagem de APP no sítio beta), por exemplo.

Exemplos de proteínas associadas com a Esclerose Lateral Amiotrófica podem incluir SOD1 (superóxido dismutase 1), ALS2 (esclerose lateral amiotrófica 2), proteína FUS (proteína fundida de sarcoma), TARDBP (proteína TAR de ligação ao DNA), VEGFA (fator de crescimento endotelial vascular A), VEGFB (fator de crescimento endotelial vascular B), e VEGFC (fator de crescimento endotelial vascular C), e qualquer combinação das mesmas.

Exemplos de proteínas associadas a doenças de priões podem incluir SOD1 (superóxido dismutase 1), ALS2 (esclerose lateral amiotrófica 2), proteína FUS (proteína fundida de sarcoma), TARDBP (proteína TAR de ligação ao DNA), VEGFA (fator de crescimento endotelial vascular A), VEGFB (fator de crescimento endotelial vascular B), e VEGFC (fator de crescimento endotelial vascular C), e qualquer combinação das mesmas.

Os exemplos de proteínas relacionadas com condições neurodegenerativas em desordens de priões podem incluir A2M (Alfa-2-Macroglobulina), AATF (Fator de transcrição antagonista da Apoptose), ACP (Fosfatase ácida prostática), ACTA2 (Alfa Actina 2 do músculo liso da aorta), ADAM22 (domínio de metalopeptidase ADAM), ADORA3 (Recetor de Adenosina A3), ou ADRA1D (recetor adrenérgico Alfa-1D para recetor alfa-adrenérgicos 1D), por exemplo.

Exemplos de proteínas associadas à imunodeficiência podem incluir A2M [alfa-2-macroglobulina]; AANAT [arilalquilamina N-acetiltransferase]; ABCA1 [cassete de ligação a ATP, subfamília A (ABC1), membro 1]; ABCA2 [cassete de ligação a ATP, subfamília A (ABC1), membro 2]; ou ABCA3 [cassete de ligação a ATP, subfamília A (ABC1), membro 3]; por exemplo.

Exemplos de proteínas associadas com Distúrbios de Repetições de Trinucleótidos incluem AR (recetor dos androgénios), FMR1 (atraso mental ligado ao X frágil tipo 1), HTT (huntingtina), ou DMPK (proteína cinase da distrofia miotónica), FXN (frataxina), ATXN2 (ataxina 2), por exemplo.

Exemplos de proteínas associadas a Desordens dos neurotransmissores incluem SST (somatostatina), NOS1 (sintase de óxido nítrico 1 (neuronal)), ADRA2A (recetor adrenérgico, alfa-2A-), ADRA2C (recetor adrenérgico, alfa-2C-), TACR1 (recetor da taquicinina 1), ou HTR2C (recetor 2C de 5-hidroxitriptamina (serotonina)), por exemplo.

Exemplos de sequências associadas ao desenvolvimento neurológico incluem A2BP1 [proteína 1 de ligação a ataxina 2], AADAT [aminoácido aminotransferase], AANAT [arilalquilamina N-acetiltransferase], ABAT [4-

aminobutirato aminotransferase], ABCA1 [cassete de ligação a ATP, subfamília A (ABC1), membro 1], ou ABCA13 [cassete de ligação a ATP, subfamília A (ABC1), membro 13], por exemplo.

Exemplos adicionais de condições preferidas que podem ser tratadas com o presente sistema incluem as que podem ser selecionadas a partir de: Síndrome de Aicardi-Goutieres; Doença de Alexander; Síndrome de Allan-Herndon-Dudley; Síndromes relacionados com POLG; Alfa-Manosidose (Tipo II e III); Síndrome de Alström; Síndrome de Angelman; Ataxia Telangiectasia; Lipofuscinose Ceroide Neuronal; Beta-talassemia; Atrofia Bilateral Ótica e (Infantil) Atrofia Ótica tipo 1; Retinoblastoma (bilateral); Doença de Canavan; Síndrome 1 Cerebro-óculo-facio-esquelético [COFS1]; Xantomatose Cerebrotendinosa; Síndrome de Cornelia de Lange; Transtornos relacionados com MAPT; Doenças genéticas de Priões; Síndrome de Dravet; Doença de Alzheimer de Início Precoce-Familiar; Ataxia de Friedreich [FRDA]; Síndrome de Fryns; Fucosidose; Distrofia Muscular Congénita de Fukuyama; Galactosialidose; Doença de Gaucher; Acidémias orgânicas; Linfocitose Hemofagocítica; Síndrome de Progeria de Hutchinson-Gilford; Mucopolisacaridose II; Doença infantil do Armazenamento de Ácido siálico; Neurodegeneração associada com PLA2G6; Síndrome de Jervell e Lange-Nielsen; Epidermólise bolhosa nas articulações; Doença de Huntington; Doença de Krabbe (Infantil); Síndrome de Leigh associado ao DNA mitocondrial e NARP; Síndrome de Lesch-Nyhan; Lisencefalia associada LIS1; Síndrome de Lowe; Doença da urina com odor a xarope de bordo; Síndrome de duplicação MECP2; Desordem do transporte de cobre relacionada com ATP7A; Distrofia Muscular Relacionado com LAMA2; Deficiência na Arilsulfatase A; Mucopolissacaridose

elemento regulador é um promotor de polimerase II. Em algumas formas de realização, a enzima CRISPR é uma enzima do sistema CRISPR de tipo II. Em algumas formas de realização, a enzima CRISPR é uma enzima Cas9. Em algumas formas de realização, a enzima Cas9 é uma Cas9 de *S. pneumonias*, *S. pyogenes*, ou *S. thermophilus*, e pode incluir uma Cas9 mutada derivada desses organismos. Em algumas formas de realização, a enzima CRISPR tem codões otimizados para a expressão numa célula eucariótica. Em algumas formas de realização, a enzima CRISPR orienta a clivagem de uma ou duas cadeias na posição da sequência alvo. Em algumas formas de realização, a enzima CRISPR carece de atividade de clivagem da cadeia de DNA.

Num aspeto, a presente divulgação prevê uma enzima CRISP compreendendo uma ou mais sequências de localização nuclear de resistência suficiente para dirigir a acumulação da referida enzima CRISPR numa quantidade detetável no núcleo de uma célula eucariótica. Em algumas formas de realização, a enzima CRISPR é uma enzima do sistema CRISPR de tipo II. Em algumas formas de realização, a enzima CRISPR é uma enzima Cas9. Em algumas formas de realização, a enzima Cas9 é uma Cas9 de *S. pneumonias*, *S. pyogenes*, ou *S. thermophilus*, e pode incluir uma Cas9 mutada derivada desses organismos. A enzima pode ser um homólogo ou ortólogo da Cas9. Em algumas formas de realização, a enzima CRISPR não tem a capacidade para clivar uma ou mais cadeias de uma sequência alvo à qual se liga.

Num aspeto, a presente divulgação prevê uma célula hospedeira eucariótica, compreendendo (a) um primeiro elemento regulador ligado operacionalmente a uma sequência de emparelhamento tracr e um ou mais locais de inserção

tipo I, II ou III; Desordem na biogénese dos Peroxissomas, Espectro do Síndrome de Zellweger; Neurodegeneração com distúrbios na Acumulação de Ferro no Cérebro; Deficiência de Esfingomielinase Ácida; Doença de Niemann-Pick de tipo C; Encefalopatia de Glicina; Distúrbios relacionados com ARX; Distúrbios no Ciclo da Ureia; Osteogénese Imperfeita Relacionada com COL1A1/2; Síndromes de deleção do DNA mitocondrial; Distúrbios relacionados com PLP1; Síndrome de Perry; Síndrome de Phelan-McDermid; Doença do Armazenamento de Glicogénio de Tipo II (Doença de Pompe) (Infantil); Distúrbios relacionados com MAPT; Distúrbios Relacionados com MECP2; Condrodisplasia Punctata Rizomélica tipo 1; Síndrome de Roberts; Doença de Sandhoff; Doença de Schindler - Tipo 1; Deficiência de Adenosina Desaminase; Síndrome de Smith-Lemli-Opitz; Atrofia Muscular Espinhal; Ataxia Espinocerebelar de início Infantil; Deficiência da Hexosaminidase A; Displasia Tanatofórica de Tipo 1; Distúrbios relacionados com o colagénio de tipo VI; Síndrome de Usher de Tipo I; Distrofia Muscular Congénita; Síndrome de Wolf-Hirschhorn; Deficiência da Lípase Ácida Lisossomal; e Xerodermia Pigmentosa.

A administração crónica de terapêuticas com proteínas pode desencadear respostas imunes inaceitáveis para a proteína específica. A imunogenicidade dos fármacos de proteína pode ser atribuída a alguns epítomos imunodominantes de linfócitos T auxiliares (HTL). A redução da afinidade da ligação de MHC a estes epítomos HTL contidos nestas proteínas pode gerar fármacos com menor imunogenicidade (Tangri S, *et al.* ("Rationally engineered therapeutic proteins with reduced immunogenicity" J Immunol. 15 de Março 2005; 174(6):3187-96.) Na presente divulgação, a imunogenicidade da enzima CRISPR, em particular, pode ser

reduzida seguindo a primeira abordagem descrita em Tangri *et al* no que diz respeito à eritropoietina e posteriormente desenvolvida. Por conseguinte, podem ser utilizado a evolução dirigida ou desenho racional para reduzir a imunogenicidade da enzima CRISPR (por exemplo uma Cas9) nas espécies hospedeiras (ser humano ou outras espécies).

Nas plantas, frequentemente os patógenos são específicos do hospedeiro. Por exemplo, *Fusarium oxysporum* f. sp. *lycopersici* faz com que o tomate muque mas ataca só o tomate, e *F. oxysporum* f. *dianthii* *Puccinia graminis* f sp. *tritici* ataca apenas o trigo. As plantas têm defesas existentes e induzidas para resistir a maioria dos patógenos. As mutações e eventos de recombinação entre as gerações de plantas conduzem a variabilidade genética que dá origem a suscetibilidade, especialmente na medida que os patógenos se reproduzem com mais frequência do que as plantas. Em plantas pode não haver resistência do hospedeiro, por exemplo, o hospedeiro e patógeno são incompatíveis. Também pode haver Resistência Horizontal, por exemplo, resistência parcial contra todas as raças de um patógeno, normalmente controladas por muitos genes e Resistência Vertical, por exemplo, resistência completa a algumas raças de um patógeno mas não para outras raças, geralmente controlada por poucos genes. Num nível Gene-para-Gene, plantas e patógenos evoluem em conjunto, e as alterações genéticas em equilíbrio alteram-se em outras. Assim, usando Variabilidade Natural, os criadores combinam a maioria dos genes úteis para Produção, Qualidade, Uniformidade, Rigidez, e Resistência. As fontes de genes de resistência incluem Variedades nativas ou estrangeiras, Variedades de Herança, Parentes de Plantas Selvagens, e Mutações Induzidas, por exemplo, tratando o material

vegetal com agentes mutagénicos. Usando a presente invenção, os criadores de plantas são fornecidos com uma nova ferramenta para induzir mutações. Por conseguinte, um perito na técnica pode analisar o genoma de fontes de genes de resistência, e em Variedades que têm as características ou traços desejados empregam a presente invenção para induzir o aumento de genes de resistência, com mais precisão do que os agentes mutagénicos anteriores e, portanto, podem acelerar e melhoram os programas de reprodução de plantas.

Como será evidente, prevê-se que o presente sistema possa ser dirigido qualquer sequência polinucleótida de interesse. Alguns exemplos de distúrbios ou doenças que podem ser tratados de forma útil utilizando o presente sistema estão incluídos nas Tabelas mencionadas anteriormente e exemplos de genes atualmente associados com estes distúrbios ou doenças também são fornecidos nessas mesmas tabelas. No entanto, os genes exemplificados não o são de modo exaustivo.

Exemplos

Os exemplos seguintes são apresentados com a finalidade de ilustrar várias formas de realização da invenção e não se destinam a limitar de qualquer forma a presente invenção. Os presentes exemplos, juntamente com os métodos descritos na presente divulgação são representativos de realizações preferidas na atualidade, são exemplificativos, e não pretendem servir de limitações ao âmbito da invenção.

Exemplo 1: Atividade do Complexo CRISPR no núcleo de uma célula eucariótica

Um exemplo do sistema CRISPR de tipo II é o locus CRISPR de tipo II SF370 de *Streptococcus pyogenes*, que contém um conjunto (cluster) de quatro genes Cas9, Cas1, Cas2, e Csn1, bem como dois elementos de RNA não codificantes, tracrRNA e uma matriz característica de sequências repetitivas (repetições diretas) intercaladas por fragmentos curtos de sequências não repetitivas (espaçadores, cada um com aproximadamente 30 pb). Neste sistema, é gerado o corte da cadeia dupla de DNA alvo (DSB) em quatro passos sequenciais (Figura 2A). Em primeiro lugar, os dois RNAs não codificantes, o pré-crRNA matriz e o tracrRNA, são transcritos a partir do locus CRISPR. Em segundo lugar, o tracrRNA hibrida com as repetições diretas de pré-crRNA, o qual é em seguida processado em crRNA maduro contendo sequências separadoras individuais. Em terceiro lugar, o complexo crRNA:tracrRNA maduro orienta a Cas9 para o alvo de DNA que consiste no proto-espaçador e PAM correspondente através da formação de um hétero-duplex entre a região espaçadora do crRNA e o DNA proto-espaçador. Finalmente, a Cas9 medeia a clivagem do DNA alvo a montante de PAM para criar um DSB dentro do proto-espaçador (Figura 2A). Este exemplo descreve um exemplo de processo para adaptar este sistema nucleáse programável de RNA para dirigir a atividade do complexo CRISPR nos núcleos das células eucarióticas.

Para melhorar a expressão de componentes CRISPR em células de mamíferos foram otimizados os códons dos dois genes do locus 1 de SF370 de *Streptococcus pyogenes* (*S. pyogenes*), Cas9 (*SpCas9*) e RNase III (*SpRNase III*). Para facilitar a localização nuclear, um sinal de localização nuclear (NLS), foi incluído no terminal amino (N) ou carboxílico (C) tanto em *SpCas9* como *SpRNase III* (Figura 2B). Para facilitar a

visualização da expressão da proteína, também foi incluída um marcador de proteína fluorescente, nos terminais N ou C de ambas as proteínas (Figura 2B). Uma versão de SpCas9 com um NLS ligado a ambos os terminais N e C também foi gerada (2xNLS-SpCas9). Os vetores contendo SpCas9 e SpRNase III fundidos com NLS foram transfetados para em células de rim embrionário humano (HEK) 293FT, e observou-se que o posicionamento relativo de NLS para SpCas9 e SpRNase III afeta a sua eficiência de localização nuclear. Considerando que o C-terminal de NLS era suficiente para orientar a SpRNase III para o núcleo, a ligação de uma única cópia destas NLS em particular a qualquer dos terminais N ou C de SpCas9 foi incapaz de atingir a localização nuclear adequada neste sistema. Neste exemplo, o C-terminal de NLS foi o de nucleoplasmina (KRPAATKKAGQAKKKK), e NLS C-terminal foi de antígeno T grande de SV40 (PKKKRKV). Das versões de SpCas9 testadas, apenas 2xNLS-SpCas9 apresentou uma localização nuclear (Figura 2B).

O tracrRNA do locus CRISPR SF370 de *S. pyogenes* tem dois sítios de início da transcrição, dando origem a dois transcritos de 89 nucleótidos (nt), e 171 nt que são posteriormente processados em tracrRNAs maduros idênticos de 75 nt. O tracrRNA de 89 nt mais curto foi selecionado para a expressão em células de mamífero (vetores de expressão ilustrados na Figura 6, com a funcionalidade tendo sido determinada por resultados de ensaio Surveyor mostrado na Figura 6B). Os locais de iniciação da transcrição são marcados como +1, e também são indicados o terminador da transcrição e a sequência sonda da transferência de Northern. A expressão de tracrRNA processado também foi confirmada mediante a transferência de Northern. A Figura 7C mostra os resultados de uma

análise de transferência de Northern de um RNA total extraído de células 293FT transfectadas com vetores de expressão U6 que suportam tracrRNA longo ou curto, bem como SpCas9 e DR-EMX1(1)-DR. Os painéis esquerdo e direito são de células 293FT transfectadas com ou sem SpRNase III, respetivamente. U6 indica o controlo de carga transferido com um snRNA de U6 humano dirigido pela sonda. A transfecção do vetor de expressão de tracrRNA curto levou a níveis abundantes da forma processada de tracrRNA (~75 pb). Quantidades muito baixas de tracrRNA longo são detetadas na transferência de Northern.

Para promover a iniciação precisa da transcrição, foi selecionado o promotor U6 baseado na RNA polimerase III para orientar a expressão de tracrRNA (Figura 2C). Do mesmo modo, foi desenvolvido um vetor baseado em promotor U6 para expressar uma matriz de pré-crRNA que consiste num único espaçador flanqueado por duas repetições diretas (DRS, também englobados pelo termo "sequências de emparelhamento tracr"; Figura 2C). O espaçador inicial foi desenhado para se dirigir a um local alvo de 33 pares de bases (pb) (proto-espaçador de 30 pb e mais uma sequência motivo CRISPR (PAM) de 3 pb que satisfaz o motivo de reconhecimento de NGG de Cas9) no locus *EMX1* humano (Figura 2C), um gene essencial no desenvolvimento do córtex cerebral.

Para testar se a expressão heteróloga do sistema CRISPR (SpCas9, SpRNase III, tracrRNA, e pré-crRNA) em células de mamíferos pode ser conseguida com a clivagem direcionada dos cromossomas de mamíferos, foram transfectadas células HEK 293FT com combinações de componentes CRISPR. Dado que as DSB em núcleos de mamífero são parcialmente reparadas

pela via de ligação de extremidade não-homóloga (NHEJ), o que leva à formação de indels, o ensaio Surveyor foi utilizado para detetar a atividade de clivagem potencial no locus *EMX1* alvo (ver, por exemplo Guschin *et al.*, 2010, Mol Biol Methods 649:247). A co-transfecção de todos os quatro componentes CRISPR foi capaz de induzir até 5,0% a clivagem de uma cadeia no proto-espaçador (ver Figura 2D). A co-transfecção de todos os componentes CRISPR menos a SpRNase III também foi capaz de induzir uma indel até 4,7% no proto-espaçador, o que sugere que pode haver RNases de mamíferos endógenos que são capazes de ajudar na maturação crRNA, tais como por exemplo as enzimas Dicer e Drosha relacionadas. A remoção de qualquer uma das restantes três componentes aboliu a atividade de clivagem do genoma do sistema CRISPR (Figura 2D). A sequenciação Sanger de amplicons contendo o locus alvo verificou que a atividade de clivagem: se encontrava em 43 clones sequenciados, 5 alelos mutados (11,6%). Experiências semelhantes usando uma variedade de sequências guia produziu percentagens de indel tão elevadas quanto 29% (ver Figuras 4-8, 10 e 11). Estes resultados definem um sistema de três componentes para a modificação eficiente do genoma mediada por CRISPR em células de mamífero.

A fim de otimizar a eficiência de clivagem, os Requerentes também testaram se diferentes isoformas de tracrRNA influenciavam a eficiência de clivagem e verificaram que, neste sistema de exemplo, apenas a forma de transcrito (89 pb) curta foi capaz de mediar a clivagem do locus genómico *EMX1* humano. A Figura 9 fornece uma análise de transferência de Northern de processamento adicional de crRNA em células de mamífero. A Figura 9A ilustra uma representação esquemática que mostra o vetor de expressão

para um único espaçador flanqueado por duas repetições diretas (DR-EMX1(1)-DR). O espaçador de 30 pb que dirige o proto-espaçador 1 do locus *EMX1* humano e as sequências de repetição direta são mostrados na sequência abaixo da Figura 9A. A linha indica a região cuja sequência de complemento inverso foi utilizada para gerar sondas para a transferência de Northern para a detecção de crRNA de EMX1(1). A Figura 9B mostra uma análise de transferência Northern de RNA total extraído de células 293FT transfectadas com vetores de expressão U6 que carregam DR-EMX1 (1)-DR. Os painéis esquerdo e direito são de células 293FT transfectadas com ou sem SpRNase III respectivamente. DR-EMX1 (1)-DR foi processado em crRNA maduros apenas na presença de SpCas9 e tracrRNA curto e não era dependente da presença de SpRNase III. O crRNA maduro detectado a partir RNA total de 293FT transfectadas tem ~33 pb e é mais curto do que o crRNA maduro de 39-42 pb de *S. pyogenes*. Estes resultados demonstram que um sistema CRISPR se pode transplantar e reprogramar em células eucarióticas para facilitar a clivagem de polinucleótidos alvo endógenos de mamífero.

A Figura 2 ilustra o sistema CRISPR bacteriano descrito neste exemplo. A Figura 2A ilustra uma *representação esquemática* que mostra o sítio CRISPR 1 de SF370 de *Streptococcus pyogenes* e um mecanismo proposto de clivagem de DNA mediada por CRISPR com este sistema. O crRNA maduro processado a partir da matriz de repetição direta-espaçador dirige a Cas9 a alvos genômicos que consistem em proto-espaçadores complementares e um motivo adjacente proto-espaçador (PAM). Após o emparelhamento do alvo-espaçador, a Cas9 medeia uma quebra da cadeia dupla no DNA alvo. Figura 2B ilustra a modificação por engenharia da Cas9 de *S.*

pyogenes (SpCas9) e RNase III (SpRNase III) com sinais de localização nuclear (NLS) para permitir a importação para o núcleo dos mamíferos. A Figura 2C ilustra a expressão em mamífero de SpCas9 e de SpRNase III conduzida pelo promotor constitutivo EF1a e tracrRNA e matriz de pré-crRNA (DR-espaçador-DR) dirigida pelo promotor U6 de RNA de Pol3 para promover a iniciação e terminação precisas da transcrição. Um proto-espaçador do locus de *EMX1* humano com uma sequência de PAM satisfatória é usado como o espaçador na matriz de pré-crRNA. A Figura 2D ilustra o ensaio de nuclease Surveyor para inserções e deleções menores mediadas por SpCas9. SpCas9 foi expressa com e sem SpRNase III, tracrRNA, e uma matriz de pré-crRNA carregando o espaçador *EMX1* alvo. Figura 2E ilustra uma representação esquemática de emparelhamento de bases entre locus alvo e *EMX1*-alvo crRNA, bem como um exemplo de um cromatograma que mostra a eliminação de micro adjacente ao local de clivagem da SpCas9. A Figura 2F ilustra alelos mutados identificados a partir da análise de sequenciação de 43 amplicons clonais mostrando uma variedade de micro inserções e deleções. Os traços indicam bases excluídas e as bases não-alinhadas ou mal emparelhadas indicam inserções ou mutações. Escala da barra = 10 μ m.

Para simplificar ainda mais o sistema de três componentes, foi adaptado um desenho híbrido crRNA-tracrRNA quimérico, onde um crRNA maduro (que compreende uma sequência guia) está condensado com um tracrRNA parcial através de uma alça de haste para imitar o duplex de crRNA:tracrRNA natural (Figura 3A).

As sequências guia podem ser inseridas entre os locais BbsI utilizando oligonucleótidos emparelhados. Os proto-

espaçadores nas cadeias sense e anti-sense são indicados acima e abaixo das sequências de DNA, respectivamente. Uma taxa de modificação de 6,3% e de 0,75% foi atingida para os loci *PVALB* humano e *Th* de ratinho, respectivamente, o que demonstra a ampla aplicabilidade do sistema CRISPR na modificação de diferentes loci em vários organismos. Enquanto a clivagem foi detetada apenas com um dos três espaçadores para cada locus utilizando as construções quiméricas, todas as sequências alvo foram clivados com uma eficiência de produção indel que alcança 27% quando se utiliza a disposição de co-expressão de pré-crRNA (Figuras 4 e 5).

A Figura 5 fornece uma ilustração adicional de que a SpCas9 pode ser reprogramada para dirigir múltiplos loci genômicos alvo em células de mamífero. A figura 5A fornece uma representação esquemática do locus *EMX1* humano mostrando a localização de cinco proto-espaçadores, indicado pelas sequências sublinhadas. A figura 5B fornece uma representação esquemática do complexo de pré-crRNA/tracrRNA mostrando hibridação entre a região de repetição direta do pré-crRNA e tracrRNA (topo), e um desenho esquemático de um design de RNA quimérico que compreende uma sequência guia de 20 pb, e sequências de emparelhamento tracr e sequências tracr que consistem em sequências de repetição direta parcial e tracrRNA hibridadas numa estrutura em forquilha (parte inferior). Os resultados de um ensaio de Surveyor que comparam a eficácia da clivagem mediada por Cas9 em cinco proto-espaçadores no locus *EMX1* humano são ilustrados na Figura 5C. Cada proto-espaçador é direcionado através do complexo processado pré-crRNA/ tracrRNA (crRNA) ou RNA quimérico (chiRNA).

para a inserção de uma ou mais sequências guia a montante da sequência de emparelhamento tracr, em que, quando expressa, a sequência guia direciona a ligação específica da sequência de um complexo CRISPR a uma sequência alvo numa célula, por exemplo, célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR num complexo com (1) a sequência guia que hibrida com a sequência alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr; e/ou (b) um segundo elemento regulador, ligado operacionalmente a uma sequência que codifica enzimas que codificam a referida enzima CRISPR compreendendo uma sequência de localização nuclear. Em algumas formas de realização, a célula hospedeira compreende os componentes (a) e (b). Em algumas formas de realização, o componente (a), componente (b), ou os componentes (a) e (b) são integrados de forma estável no genoma da célula eucariótica hospedeira. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência de emparelhamento tracr sob o controlo do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende adicionalmente duas ou mais sequências guia operacionalmente ligadas ao primeiro elemento regulador, em que, quando expressas, cada uma das duas ou mais sequências guia direcionam a ligação específica de um complexo CRISPR para uma sequência alvo diferente numa célula eucariótica. Em algumas formas de realização, a célula hospedeira eucariótica compreende ainda um terceiro elemento regulador, tal como um promotor de polimerase III, operacionalmente ligado à referida sequência tracr. Em algumas formas de realização, a sequência tracr exhibe, pelo menos, 50%, 60%, 70%, 80%, 90%, 95%, ou 99% de complementaridade de sequência ao longo do comprimento da

Uma vez que a estrutura secundária do RNA pode ser crucial para interações intermoleculares, foi utilizado um algoritmo de predição da estrutura com base na energia livre mínima e conjunto de estrutura ponderada de Boltzmann para comparar a estrutura secundária putativa de todas as sequências guia utilizadas na nossa experiência dirigida ao genoma (Figura 3B) (ver por exemplo, Gruber *et al.*, 2008, *Nucleic Acids Research*, 36:W70). A análise revelou que, na maioria dos casos, as sequências guia eficientes no contexto do crRNA quimérico eram substancialmente livres de motivos de estrutura secundária, enquanto que as sequências guia ineficientes tinham maior probabilidade de formar estruturas secundárias internas que poderiam impedir o emparelhamento de bases com o DNA proto-espaçador alvo. Assim, é possível que a variabilidade na estrutura secundária do espaçador tenha um impacto na eficiência da interferência mediada por CRISPR quando se utiliza um crRNA quimérico.

A Figura 3 ilustra vetores de expressão a modo de exemplo. A Figura 3A fornece uma representação esquemática de um vetor bicistrônico para dirigir a expressão de uma quimera sintética crRNA-tracrRNA (RNA quimérico), assim como SpCas9. O RNA guia quimérico contém uma sequência guia de 20 pb correspondendo ao proto-espaçador no local genômico alvo. A figura 3B fornece um diagrama esquemático que mostra as sequências guia a serem dirigidas aos loci *EMX1* e *PVALB* humanos, e *Th* de ratinho, assim como as suas estruturas secundárias previstas. A eficiência da modificação em cada local alvo é indicada abaixo da figura da estrutura secundária do RNA (*EMX1*, $n = 216$ leituras de sequenciação dos amplicons; *PVALB*, $n = 224$ leituras; *Th*, $n = 265$ leituras). O algoritmo de dobragem origina uma saída

com cada base colorida de acordo com a sua probabilidade de assumir a estrutura secundária prevista, tal como indicado por uma escala de arco-íris que está representada na figura 3B em escala de cinzentos. Outros modelos de vetores para SpCas9 são mostrados na Figura 3A, incluindo vetores de expressão simples que incorpora um promotor U6 ligado a um sítio de inserção para um oligo guia, e um promotor Cbh ligado à sequência de codificação da SpCas9.

Para testar se os espaçadores contendo estruturas secundárias são capazes de funcionar em células procarióticas onde CRISPR opera naturalmente, foi usado um ensaio de interferência de transformação de plasmídeos portadores de proto-espaçador de uma estirpe de *E. coli* que expressa de forma heteróloga o locus 1 de CRISPR SF370 de *S. pyogenes* (Figura 3C). O locus CRISPR foi clonado num vetor de expressão de *E. coli* de baixo número de cópias e a matriz crRNA foi substituída com um único espaçador flanqueado por um par de DR (pCRISPR). As estirpes de *E. coli* contendo plasmídeos diferentes pCRISPR foram transformadas com plasmídeos contendo as sequências em teste com o correspondente proto-espaçador e PAM (Figura 3C). No ensaio bacteriano, todos os espaçadores facilitavam a interferência eficiente de CRISPR (Figura 3C). Estes resultados sugerem que podem existir fatores adicionais que afetam a eficiência da atividade CRISPR em células de mamífero.

Para investigar a especificidade da clivagem mediada por CRISPR, o efeito de mutações de um único nucleótido na sequência guia na clivagem do proto-espaçador no genoma de mamífero foi analisada utilizando uma série de crRNAs quiméricos que têm como alvo EMX1 com mutações pontuais

individuais (Figura 4A). A Figura 4B ilustra os resultados de um ensaio de nuclease de Surveyor que compara a eficácia de clivagem da Cas9 quando emparelhada com diferentes RNAs quiméricos mutantes. Os emparelhamentos incorretos de uma única base até 12 pb na posição 5' do PAM abole substancialmente a clivagem genômica por SpCas9, enquanto que os espaçadores com mutações nas posições mais a montante mantiveram a atividade frente ao alvo proto-espaçador original (Figura 4B). Além do PAM, SpCas9 tem especificidade única de bases nos últimos 12 pb do espaçador. Além disso, CRISPR é capaz de mediar a clivagem genômica tão eficientemente quanto um par de nucleases TALE (TALEN) que se dirigem ao mesmo proto-espaçador de *EMX1*. A Figura 4C fornece uma representação esquemática que mostra o desenho de TALEN com alvo em *EMX1*, e a Figura 4D mostra um gel de Surveyor que compara a eficiência de TALEN e Cas9 (n = 3).

Depois de ter sido estabelecido um conjunto de componentes para conseguir a edição de genes mediada por CRISPR em células de mamífero através do mecanismo de NHEJ propenso a erros, foi testada a capacidade de CRISPR para estimular a recombinação homóloga (HR), uma via de reparação de genes de alta-fidelidade para realizar correções precisas no genoma, foi testado. A SpCas9 de tipo selvagem é capaz de mediar DSB específicas de local, que se podem reparar, tanto através de NHEJ como de HR. Além disso, foi manipulada por engenharia a substituição de um aspartato por alanina (D10A) no domínio catalítico de RuvC I de SpCas9 para converter a nuclease numa *nickase* (SpCas9n; ilustrado na Figura 5A) (ver por exemplo Sapranasaks *et al.*, 2011, *Nucleic Acids Research*, 39:9275; Gasiunas *et al.*, 2012, *Proc. Natl. Acad. Sci. USA*, 109:E2579), de tal

modo que o DNA genómico nicked sofre a reparação dirigida por homologia de alta fidelidade (HDR). O ensaio Surveyor confirmou que a SpCas9n não gera indels no proto-espçador alvo de *EMX1*. Tal como ilustrado na Figura 5B, a co-expressão de crRNA quimérico para ser dirigido a *EMX1* com SpCas9 produz indels no local alvo, enquanto que a co-expressão com SpCas9n não produz ($n = 3$). Além disso, a sequenciação de 327 amplicons não detetou quaisquer indels induzidos por SpCas9n. O mesmo locus foi escolhido para testar HR mediada por CRISPR mediante a co-transfecção de células HEK 293FT com o RNA alvo quimérico a ser dirigido a *EMX1*, hSpCas9 ou hSpCas9n, bem como um modelo de HR para introduzir um par de locais de restrição (*HindIII* e *NheI*) perto do proto-espçador. A Figura 5C fornece uma ilustração esquemática da estratégia de HR, com as localizações relativas dos pontos de recombinação e as sequências de hibridação dos primers (setas). De fato, SpCas9 e SpCas9n catalisam a integração do molde HR no locus *EMX1*. A amplificação por PCR da região alvo seguida de digestão com enzima de restrição *HindIII* revelou produtos de clivagem correspondentes aos tamanhos dos fragmentos esperados (na Figura 5D as setas que se mostram no gel da análise dos polimorfismo de comprimento de fragmento de restrição), e com SpCas9 SpCas9n a mediar níveis semelhantes de eficiência de HR. Os Requerentes verificaram ainda HR utilizando sequenciação de Sanger de amplicons genómicos (Figura 5E). Estes resultados demonstram a utilidade de CRISPR para facilitar a inserção do gene de forma dirigida no genoma de mamífero. Dada a especificidade do alvo de 14 pb (12 pb do espçador e 2 pb do PAM) da SpCas9 do tipo selvagem, a disponibilidade de uma *nickase* pode reduzir de forma significativa a probabilidade de alterações para fora do alvo, uma vez que

quebras das cadeias individuais não são substratos para a via NHEJ mais propensa a erros.

A construção de vetores de expressão que imitam a arquitetura natural dos loci CRISPR com espaçadores agrupados (Figura 2A) de modo a testar a possibilidade de direcionamento de múltiplas sequências. Usando uma única matriz CRISPR que codifica um par de direcionamento *EMX1* e *PVALB*, foi detetada a clivagem eficiente em ambos os loci (Figura 4F, que mostra tanto um desenho esquemático da matriz de crRNA como uma transferência de Surveyor que mostra a mediação eficiente da clivagem). Também foi testada a eliminação direcionada de regiões genômicas maiores através DSB simultâneos usando espaçadores contra dois alvos dentro de *EMX1* espaçados por 119 pb, e foi detetada uma eficácia de deleção de 1,6% (3 de 182 amplicons; Figura 5G) foi detetado. Isto demonstra que o sistema CRISPR pode mediar a edição multiplex dentro de um único genoma.

Exemplo 2: Modificações e alternativas do sistema CRISPR

A capacidade para usar o RNA para programar a clivagem específica da sequência de DNA define uma nova classe de ferramentas de modificação mediante engenharia do genoma para uma variedade de aplicações de investigação e industriais. Vários aspetos do sistema CRISPR podem ser ainda melhorados para aumentar a eficiência e versatilidade do direcionamento de CRISPR. A atividade Cas9 ótima pode depender da disponibilidade de Mg^{2+} livre em níveis mais elevados do que o que se encontra presente no núcleo de mamíferos (ver, por exemplo Jinek et al., 2012, Science, 337:816), e a preferência para um motivo NGG imediatamente

a jusante do proto-espaçador restringe a capacidade de se dirigir em média a cada 12 pb no genoma humano. Algumas destas limitações podem ser superadas explorando a diversidade de loci de CRISPR em todo o metagenoma microbiano (ver, por exemplo Makarova *et al.*, 2011, *Nat Rev Microbiol*, 9:467). Outros loci de CRISPR podem ser transplantados para o meio celular de mamífero mediante um processo semelhante ao descrito no Exemplo 1. A eficiência da modificação em cada local alvo é indicada a seguir às estruturas secundárias de RNA. O algoritmo para gerar as estruturas de cores para cada de base de acordo com a sua probabilidade de assumir a estrutura secundária prevista. Os espaçadores de RNA guia 1 e 2 induziram 14% e 6,4%, respetivamente. A análise estatística da atividade de clivagem através de réplicas biológicas nestes dois locais de proto-espaçadores também é fornecida na Figura 7.

Exemplo 3: Algoritmo de seleção de sequência alvo na amostra

Foi projetado um programa de software para identificar sequências alvo CRISPR candidatas em ambas as cadeias de uma sequência de DNA de entrada com base no comprimento da sequência guia desejada e uma sequência motivo CRISPR (PAM) para uma enzima CRISPR especificada. Por exemplo, podem-se identificar os locais alvo para a Cas9 de *S. pyogenes*, com sequências NGG de PAM, mediante a procura de 5'-N_x-NGG-3', tanto na sequência de entrada como no complemento reverso da entrada. Da mesma forma, se podem identificar os locais alvo para Cas9 de CRISPR1 de *S. thermophilus*, com sequência PAM NNAGAAW, mediante a procura de 5'-N_x-NNAGAAW-3', tanto na sequência de entrada e como no complemento reverso da entrada. Da mesma forma, podem ser identificados os locais

alvo para Cas9 de CRISPR3 de *S. thermophilus*, com sequência PAM NGGNG, através da procura de 5'-N_x-NGGNG-3', tanto na sequência de entrada como no complemento reverso da entrada. O valor "x" em N_x pode ser fixado pelo programa ou especificado pelo usuário, tal como 20.

Uma vez que várias ocorrências no genoma do local alvo de DNA podem conduzir à edição não específica do genoma, após a identificação de todos os locais potenciais, o programa filtra as sequências com base no número de vezes que aparecem no genoma de referência relevante. Para essas enzimas CRISPR para as quais a especificidade de sequência é determinada por uma sequência "semente", tal como de 11-12 pb na região 5' da sequência de PAM, incluindo a própria sequência PAM, o passo de filtragem pode basear-se na sequência semente. Assim, para evitar a edição em loci genômicos adicionais, os resultados são filtrados com base no número de ocorrências da sequência semente:PAM no genoma relevante. O utilizador pode ser autorizado a escolher o comprimento da sequência semente. O usuário também pode ser permitido especificar o número de ocorrências da sequência semente:PAM num genoma para fins de passar o filtro. O valor padrão é identificar sequências únicas. O nível do filtro é alterado mudando tanto o comprimento da sequência semente como o número de ocorrências da sequência no genoma. O programa pode em adição ou em alternativa fornecer a sequência de uma sequência guia complementar da sequência ou sequências alvo relatadas através do fornecimento do complemento reverso da sequência ou sequências alvo identificadas.

Exemplo 4: Avaliação de múltiplos híbridos de crRNA-tracrRNA quiméricos

Este exemplo descreve os resultados obtidos para os RNA quiméricos (chiRNAs; que compreende uma sequência guia, uma sequência de emparelhamento tracr, e uma sequência tracr num único transcrito) que têm sequências tracr que incorporam diferentes comprimentos de sequência tracrRNA de tipo selvagem. A Figura 18a ilustra uma representação esquemática de um vetor de expressão bicistrônico para RNA quimérico e Cas9. Cas9 é acionada pelo promotor CBh e o RNA quimérico é acionado por um promotor U6. O RNA guia quimérico consiste numa sequência guia de 20 pb (Ns) ligada à sequência tracr (que acontece desde o primeiro "U" da cadeia inferior até ao fim do transcrito), o qual é truncado em várias posições, tal como indicado. As sequências guia e tracr são separadas pela sequência de emparelhamento tracr GUUUUAGAGCUA seguida pela sequência de alça GAAA. Os resultados dos ensaios SURVEYOR para indels mediados por Cas9 nos loci *EMX1* e *PVALB* humanos estão ilustrados na Figura 18b e 18c, respetivamente. As setas indicam os fragmentos SURVEYOR esperados. Os chiRNA são indicados pela designação "+n", e crRNA refere-se a um RNA híbrido no qual as sequências guia e tracr são expressas como transcritos separados. A quantificação destes resultados, realizados em triplicado, são ilustrados em histograma nas Figuras 11a e 11b, que correspondem às Figuras 10b e 10c, respetivamente ("N.D." indica que não há indels detetados). Os ID dos proto-espaçadores e seus alvos genómicos, sequência proto-espaçador, sequência de PAM, e localização da cadeia são fornecidos na Tabela D. As sequências guia foram desenhadas para serem complementares à sequência proto-espaçador completa no caso de transcritos

separados no sistema híbrido, ou apenas para a porção sublinhada no caso de RNA quiméricos.

Tabela D:

ID Proto-espacedor	Alvo genómico	Sequência Proto-espacedor (5' to 3')	PA M	Cadeia
1	<i>EMX1</i>	GGACATCGAT <u>GTCACCTCCAATGACTAG</u> <u>GG</u>	TG G	+
2	<i>EMX1</i>	CATTGGAGGT <u>GACATCGATGTCCTCCCC</u> <u>AT</u>	TG G	-
3	<i>EMX1</i>	<u>GGAAGGGCCTGAGTCCGAGCAGAAGAA</u> <u>GAA</u>	GG G	+
4	<i>PVALB</i>	GGTGGCGAGAG <u>GGGGCCGAGATTGGGTGT</u> <u>TC</u>	AG G	+
5	<i>PVALB</i>	ATGCAGGAGCGTGGCGAGAG <u>GGGGCCGA</u> <u>GAT</u>	TG G	+

Cultura de células e transfecção

A linha celular 293FT de rim embrionário humano (HEK) (Life Technologies) foi mantida em Dulbecco's Modified Eagle's Medium (DMEM) suplementado com 10% de soro fetal bovino (HyClone), 2 mM GlutaMAX (Life Technologies), 100U/mL de penicilina, e 100 µg/mL de estreptomicina a 37°C com incubação com 5% de CO₂. As células 293FT foram inoculadas em placas de 24 poços (Corning) 24 horas antes da transfecção a uma densidade de 150.000 células por poço. As células foram transfectadas utilizando Lipofectamina 2000 (Life Technologies) seguindo o protocolo recomendado pelo fabricante. Para cada poço de uma placa de 24 poços, foi usado um total de 500 ng de plasmídeo.

Ensaio SURVEYOR para a modificação do genoma

As células 293FT foram transfectadas com DNA plasmídico, como descrito anteriormente. As células foram incubadas a 37°C durante 72 horas após a transfecção antes da extração do DNA genómico. O DNA genómico foi extraído utilizando o QuickExtract DNA Extraction Solution (Epicentre) seguindo o protocolo do fabricante. Resumidamente, as células sedimentadas foram ressuspensas em solução QuickExtract e incubou-se a 65°C durante 15 minutos e 98°C durante 10 minutos. A região genómica que flanqueia o local alvo CRISPR para cada gene foi amplificada por PCR (primers listados na Tabela E), e os produtos foram purificados utilizando uma coluna de centrifugação QiaQuick (Qiagen) seguindo o protocolo do fabricante. Um total de 400 ng dos produtos de PCR purificados foram misturados com 2 µL de tampão 10X Taq DNA polimerase PCR (Enzymatics) e água ultrapura até um volume final de 20 µL, e submetido a um processo de re-emparelhamento (re-annealing) para permitir a formação de heteroduplexes: 95°C durante 10 min, 95°C a 85°C aumentando - 2°C/s, de 85°C a 25°C a - 0,25°C/s, e 25°C mantido durante 1 minuto. Depois de re-emparelhamento, os produtos foram tratados com nuclease SURVEYOR e potenciador S SURVEYOR (Transgenomics) seguindo o protocolo recomendado pelo fabricante, e analisados em géis Novex de 4-20% de poliacrilamida em TBE (Life Technologies). Os géis foram corados com o corante SYBR Gold DNA (Life Technologies) durante 30 minutos e fotografados com um sistema de imagem para géis Gel Doc (Bio-Rad). A quantificação foi baseada em intensidades relativas de bandas.

sequência de emparelhamento tracr quando otimamente alinhadas. Em algumas formas de realização, a enzima CRISPR compreende uma ou mais sequências de localização nuclear de resistência suficiente para dirigir a acumulação da referida enzima CRISPR no núcleo de uma célula eucariótica numa quantidade detetável. Em algumas formas de realização, a enzima é uma enzima CRISPR do sistema CRISPR de tipo II. Em algumas formas de realização, a enzima CRISPR é uma enzima Cas9. Em algumas formas de realização, a enzima Cas9 é uma Cas9 de *S. pneumonias*, *S. pyogenes*, ou *S. thermophilus*, e pode incluir uma Cas9 mutada derivada desses organismos. A enzima pode ser um homólogo ou ortólogo da Cas9. Em algumas formas de realização, a enzima CRISPR tem codões otimizados para a expressão numa célula eucariótica. Em algumas formas de realização, a enzima CRISPR orienta a clivagem de uma ou duas cadeias na posição da sequência alvo. Em algumas formas de realização, a enzima CRISPR carece de atividade de clivagem da cadeia de DNA. Em algumas formas de realização, o primeiro elemento regulador é um promotor de polimerase III. Em algumas formas de realização, o segundo elemento regulador é um promotor de polimerase II. Em algumas formas de realização, a sequência guia tem de comprimento pelo menos 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10-30, ou entre 15-25, ou entre 15-20 nucleótidos. Num aspeto, a presente divulgação prevê um organismo eucariota não humano; de preferência um organismo eucariota multicelular, que compreende uma célula eucariota hospedeira de acordo com qualquer uma das formas de realização descritas. Em outros aspetos, a invenção fornece um organismo eucariota; de preferência um organismo eucariota multicelular, que compreende uma célula hospedeira eucariota de acordo com qualquer uma das realizações descritas. O organismo em

Tabela E:

Nome do Primer	Alvo genómico	Sequência do primer (5' para 3')
Sp-EMX1-F	<i>EMX1</i>	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	<i>EMX1</i>	GGAGATTGGAGACACGGAGA G
Sp-PVALB-F	<i>PVALB</i>	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	<i>PVALB</i>	GGCAGCAAACCTCCTTGTCCT

Identificação computacional de locais alvo exclusivos de CRISPR

Para identificar locais alvo exclusivos para a enzima Cas9 SF370 de *S. pyogenes* (SpCas9) no genoma humano, de ratinho, rato, peixe zebra, mosca da fruta, e *C. elegans*, foi desenvolvido pelos inventores um pacote de software para digitalizar ambas as cadeias de uma sequência de DNA e identificar todos os possíveis locais alvo SpCas9. Para este exemplo, cada local alvo SpCas9 foi operacionalmente definido como uma sequência de 20 pb seguido por uma sequência de motivo adjacente proto-espaçador NGG (PAM), e os inventores identificaram todas as sequências que satisfazem essa definição 5'-N₂₀-NGG-3' em todos os cromossomas. Para impedir a edição de genoma não-específico, após a identificação de todos os locais potenciais, todos os locais alvo foram filtradas com base no número de vezes em que aparecem no genoma de referência relevante. Para tirar vantagem da especificidade da sequência da atividade Cas9 conferida por uma sequência de "semente", que pode ser, por exemplo, uma sequência na posição 5' de aproximadamente 11-12 pb da sequência de PAM, as sequências 5'-NNNNNNNNNN-NGG-3' foram selecionadas para ser única no genoma relevante. Todas as sequências genómicas foram retiradas a partir do UCSC Genome Browser (genoma Humano hg19, genoma Ratinho MM9, genoma Rato m5,

genoma do Peixe Zebra danRer7, genoma de *D. melanogaster* dm4 e genoma de *C. elegans* cel10). Os resultados da pesquisa completa estão disponíveis para visualização usando a informação do UCSC Genome Browser. Um exemplo visualização de alguns locais alvo no genoma humano é apresentado na Figura 22.

Inicialmente, três locais dentro do locus *EMX1* em células HEK 293FT humanas foram alvo. Foi avaliada a eficiência da modificação do genoma de cada chiRNA utilizando o ensaio de nuclease SURVEYOR, que deteta mutações que resultam de quebras da cadeia dupla de DNA (DSB) e sua posterior reparação pela via de reparação dos danos no DNA que une o extremo não homólogo (NHEJ). As construções designadas chiRNA (+n) indicam que até ao nucleótido +n do tracrRNA de tipo selvagem é incluído na construção de RNA quimérico, com valores de 48, 54, 67, e 85 usado para n. Os RNA quiméricos que contêm fragmentos mais longos de tipo selvagem tracrRNA (chiRNA (+67) e chiRNA (+85)) mediam a clivagem de DNA em todos os três locais alvo *EMX1*, com chiRNA (+85), em especial, demonstrando níveis significativamente mais elevados de clivagem de DNA do que os híbridos correspondente de crRNA/tracrRNA que expressam sequências guia e tracr em transcritos separados (Figuras 10b e 10a). Dois locais no locus *PVALB* que não originaram nenhuma clivagem detetável usando o sistema híbrido (sequência guia e sequência tracr expressa como transcritos separados) também foram alvo usando chiRNAs. Os chiRNA (+67) e chiRNA (+85) foram capazes de mediar clivagem significativa nos dois proto-espaçadores *PVALB* (Figuras 10c e 10b).

Para todos os cinco alvos nos loci *EMX1* e *PVALB*, foi observado um aumento consistente na eficiência da modificação do genoma com o aumento do comprimento da sequência tracr. Sem se pretender ficar limitado por qualquer teoria, a estrutura secundária formada pela extremidade 3' do tracrRNA pode desempenhar um papel no aumento da taxa de formação do complexo CRISPR. Uma ilustração de estruturas secundárias previstas para cada um dos RNA quiméricos utilizados no presente exemplo é fornecido na Figura 21. A estrutura secundária foi prevista utilizando RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) usando o algoritmo de energia livre mínima e de função de emparelhamento. A pseudo-cor para cada base (reproduzida em escala de cinza) indica a probabilidade de emparelhamento. Dado que os chiRNAs com sequências tracr mais longas foram capazes de clivar alvos que não foram clivados por híbridos de crRNA/tracrRNA CRISPR nativos, é possível que o RNA quimérico pode ser carregado numa Cas9 mais eficientemente do que o seu homólogo híbrido nativo. Para facilitar a aplicação de Cas9 para edição de genoma específico do local em células e organismos eucarióticas, todos os locais alvo únicos previstos para Cas9 de *S. pyogenes* foram identificados computacionalmente nos genomas de humano, ratinho, rato, peixe zebra, *C. elegans*, e *D. melanogaster*. Os RNA quiméricos podem ser desenhados para as enzimas Cas9 de outros micróbios para expandir o espaço alvo de nucleases de RNA programáveis CRISPR.

As Figuras 11 e 21 ilustram exemplos de vetores de expressão bicistrônica para a expressão de RNA quimérico incluindo até +85 nucleótidos da sequência de RNA tracr de tipo selvagem, e SpCas9 com sequências de localização nuclear. A SpCas9 é expressa a partir de um promotor CBh e

termina com o sinal bGH poliA (bGH pA). A sequência ampliada apresentada imediatamente abaixo corresponde à região em torno do local de inserção de sequência guia, e inclui, de 5' para 3', a porção 3' do promotor U6 (primeira região sombreada), locais de clivagem BbsI (setas), repetição direta parcial (sequência de emparelhamento tracr GTTTTAGAGCTA, sublinhada), sequência de alça GAAA, e sequência tracr +85 (sequência sublinhada depois da sequência alça). Uma inserção da sequência guia a modo de exemplo é ilustrado abaixo do local de inserção da sequência guia, com a sequência de nucleótidos guia para um alvo selecionado representado por um "N".

As sequências descritas nos exemplos anteriores são as seguintes (as sequências de polinucleótidos são apresentadas de 5' para 3'):

tracrRNA curto U6 (SF370 de *Streptococcus pyogenes*):

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAAT
AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(Negrito= sequência tracrRNA; sublinhado= sequência de finalização)

tracrRNA longo U6 (SF370 de *Streptococcus pyogenes*):

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTT
 CTTTGAATTTCTCCTTGATTATTTGTTATAAAAAGTTATAAAATAATCTTGTTGGAACC
 ATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG
 TGGCACCGAGTCGGTGCTTTTTTTT

Cadeia principal-DR U6-DR-BbsI (SF370 de *Streptococcus pyogenes*) :

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAA
 ACCGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

Cadeia principal de U6-RNA quimérico-BbsI (SF370 de *Streptococcus pyogenes*)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAAT
 AGCAAGTTAAAATAAGGCTAGTCCG

NLS-SpCas9-EGFP:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGIHGVPAAADKKYSIGLDIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDL
AEDAKLQLSKDQYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNR
EKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTN
FDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPAIKKG
ILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID
NKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDF
QFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILV
ELDGDVNGHKFSVSGEGEGDATYGKLTLLKFICTTGKLPVPWPPTLVTTLTYGVCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
GNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYK

SpCas9-EGFP-NLS :

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIF
 GNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
 NLIASLSGLTPNFKS NFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
 HAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
 FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL
 KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG

DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKN
 SRERMKRIIEGIGELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
 YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
 QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYG
 DYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
 VWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
 YGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKE
 VKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
 PEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVL SAYNKH RDKPIREQAENI
 IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDAA
 AVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATY GKLTLKFICTTGKLPVPWP
 TLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG
 DTLVNRIELKGIDFKEDGNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDGSV
 QLADHYQQNTPIGDGPVLLPDNH YLSTQSALSKDPNEKRDH MVLLEFVTAAGITLGM
 ELYKKRPAATKKAGQAKKKK

NLS-SpCas9-EGFP-NLS :

MDYKDHDGDYKDHIDYKDDDDKMAPKKRKVGIHGVPAADKKYSIGLDIGTNSVG
 WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
 NRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
 LRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF
 ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
 AEDAKLQLSKD TYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLS
 ASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
 PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFLKDNR
 EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTN
 FDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
 KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
 ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEELGKELGSGI

LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID
 NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
 LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
 IGTKATKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
 AKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIILPKYSLFELE
 NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKY
 FDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILV
 ELGDGVNGHKFSVS GEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFSRYP
 DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
 GNILGHKLEYNNSHN VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
 VLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDLEYKKRPAATKKAGQA
 KKKK

NLS-SpCas9-NLS:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGIHGVPAADKKYSIGLDIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTTYH
LRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNR
EKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTN
FDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEELGELGSKI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID
NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKVR EINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
IGKATAKYFFYSNMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIKLPKYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK

NLS-mCherry-SpRNase3:

MFLFLSLTSFLSSRTLVSKEEDNMAIIEFMRFKVHMEGSVNGHEFEIEGEGEGRPYE
GTQTAKLKVTGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLLKLSFPEGFKWERVM
NFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPMQKKTMGWEASSERMYPED
GALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVE
QYERAEGRHSTGGMDELYKGSQLEELLSTSFDIQFNDLTLLTAFTHTSYANEHRLN
VSHNERLEFLGDAVLQLIISEYLF AKYPKKTEGDMSKLRSMIVREESLAGFSRFSFDAYI
KLKGEEKSGGRRRDTILGDLFEAFLGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVK
DYKTCLQEFLQTKGDVAIDYQVISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQD
AAKNALAQLESEV

SpRNase3-mCherry-NLS:

MKQLEELLSTSFDIQFNDLTLLTAFTHTSYANEHRLLNVSHNERLEFLGDAVLQLIISEY
LFAKYPKKTEGDMSKLRSMIVREESLAGFSRFCSDAYIKLGKGEEKSGGRRRDTILGDL
FEAFLGALLLDKGIDA VRRFLKQVMIPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQ
VISEKGPAHAKQFEVSIVNGAVLSKGLGKSKKLAEQDAAKNALAQLSEVGSVSKGEE
DNMAIIEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDIL
SPQFMYGSKAYVKHPADIPDYKLKSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFI
YKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAE
VKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKKRP
AATKKAQAKKKK

NLS-SpCas9n-NLS (a mutação *nickase* D10A está
em minúsculas):

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKRKVGIHGVPAADKKYSIGLaIGTNSVGW
AVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRKN
RICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDQYDDDLNLLAQIGDQYADFLAAKNLSDAILSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI
LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREK
IEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIIEGKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSID
NKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLKPYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGDKRPAATKKAQAKKKK

algumas formas de realização destes aspetos pode ser um animal; por exemplo, um mamífero. Além disso, o organismo pode ser um artrópode, tal como um inseto. O organismo também pode ser uma planta. Além disso, o organismo pode ser um fungo.

Num aspeto, a presente divulgação prevê um kit que compreende um ou mais dos componentes aqui descritos. Em algumas formas de realização, o kit compreende um sistema de vetor e instruções para utilizar o kit. Em algumas formas de realização, o sistema de vetor compreende (a) um primeiro elemento regulador ligado operacionalmente a uma sequência de emparelhamento tracr e um ou mais locais de inserção para a inserção de uma ou mais sequências guia a montante da sequência de emparelhamento tracr, em que, quando expressa, a sequência guia direciona a ligação específica da sequência de um complexo CRISPR a uma sequência alvo na célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR num complexo com (1) a sequência guia que hibrida com a sequência alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr; e/ou (b) um segundo elemento regulador, ligado operacionalmente a uma sequência que codifica enzimas que codificam a referida enzima CRISPR compreendendo uma sequência de localização nuclear. Em algumas formas de realização, o kit compreende os componentes (a) e (b) localizados no mesmo ou em diferentes vetores do sistema. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência de emparelhamento tracr sob o controlo do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência de emparelhamento tracr sob o

hEMX1-HR Molde-HindII-NheI:

GAATGCTGCCCTCAGACCCGCTTCCTCCCTGTCCTTGTCTGTCCAAGGAGAATGAGG
TCTCACTGGTGGATTTCGGACTACCCCTGAGGAGCTGGCACCTGAGGGACAAGGCCC
CCCACCTGCCCAGCTCCAGCCTCTGATGAGGGGTGGGAGAGAGCTACATGAGGTTG
CTAAGAAAGCCTCCCCTGAAGGAGACCACACAGTGTGTGAGGTTGGAGTCTCTAGC
AGCGGGTTCTGTGCCCCCAGGGATAGTCTGGCTGTCCAGGCACTGCTCTTGATATAA
ACACCACCTCCTAGTTATGAAACCATGCCCCATTCTGCCTCTCTGTATGGAAAAGAGC
ATGGGGCTGGCCCCGTGGGGTGGTGTCCACTTTAGGCCCTGTGGGAGATCATGGGAA
CCCACGCACTGGGTCATAGGCTCTCTCATTTACTACTCACATCCACTCTGTGAAGAA
GCGATTATGATCTCTCCTCTAGAACTCGTAGAGTCCCATGTCTGCCGGCTTCCAGA
GCCTGCACTCCTCCACCTTGGCTTGGCTTTGCTGGGGCTAGAGGAGCTAGGATGCAC
AGCAGCTCTGTGACCCTTTGTTTGAGAGGAACAGGAAAACCAACCTTCTCTCTGGCC
CACTGTGTCTCTTCTGCCCCTGCCATCCCCCTTCTGTGAATGTTAGACCCATGGGAGC
AGCTGGTCAGAGGGGACCCCGCCCTGGGGCCCCCTAACCCCTATGTAGCCTCAGTCTTC
CCATCAGGCTCTCAGCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTGCTCTGGGGGCC
TCCTGAGTTTCTCATCTGTGCCCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAG
AACCGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCG
AGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAG
GCCAATGGGGAGGACATCGATGTCACCTCCAATGACaagcttgetageGGTGGGCAACCAC
AAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGCCCCCTGCGTGGGGCCAA
GCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGGCCCCA
CAGGGCTTGAAGCCCCGGGGCCGCCATTGACAGAGGGACAAGCAATGGGCTGGCTGA
GGCCTGGGACCACTTGGCCTTCTCCTCGGAGAGCCTGCCTGCCTGGGCGGGCCCCGCC
CGCCACCGCAGCCTCCCAGCTGCTCTCCGTGTCTCCAATCTCCCTTTTGTTTTGATGC
ATTTCTGTTTTAATTTATTTTCCAGGCACCACTGTAGTTTAGTGATCCCCAGTGTCCC
CCTTCCCTATGGGAATAATAAAAGTCTCTCTCTTAATGACACGGGCATCCAGCTCCA
GCCCCAGAGCCTGGGGTGGTAGATTCCGGCTCTGAGGGCCAGTGGGGGCTGGTAGA
GCAAACGCGTTCAGGGCCTGGGAGCCTGGGGTGGGGTACTGGTGGAGGGGGTCAAG
GGTAATTCATTAACCTCTCTTTTGTGGGGACCCTGGTCTCTACCTCCAGCTCCA
CAGCAGGAGAAACAGGCTAGACATAGGGAAGGGCCATCCTGTATCTTGAGGGAGGA
CAGGCCCAGGTCTTTCTTAACGTATTGAGAGGTGGGAATCAGGCCCAGGTAGTTCAA
TGGGAGAGGGAGAGTGCTTCCCTCTGCCTAGAGACTCTGGTGGCTTCTCCAGTTGAG
GAGAAACCAGAGGAAAGGGGAGGATTGGGGTCTGGGGGAGGGAACACCATTCACA
AAGGCTGACGGTTCAGTCCGAAGTCGTGGGCCCACCAGGATGCTCACCTGTCTTG
GAGAACCGCTGGGCAGGTTGAGACTGCAGAGACAGGGCTTAAGGCTGAGCCTGCAA
CCAGTCCCCAGTGA CT CAGGGCCTCCTCAGCCCAAGAAAGAGCAACGTGCCAGGGC
CCGCTGAGCTCTTGTGTTACCTG

RNA quimérico contendo +48 tracrRNA (SF370 de *S. pyogenes*)
gagggcctatttcccatgattccctcatatttgcataatagatacagggctgttagagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatagtgacgtagaaagtaataattcttgggtagtttgcagttttaaataattgttttaaatggactatcatatgcttaccgtaactt
gaaagtatttgcatttcttggctttatatacttgggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNNggttttaga
gctagaaatagcaagttaaataaggctagtcgTTTTTTT

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

RNA quimérico contendo +54 tracrRNA (SF370 de *S. pyogenes*)
gagggcctatttcccatgattccctcatatttgcataatagatacagggctgttagagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatagtgacgtagaaagtaataattcttgggtagtttgcagttttaaataattgttttaaatggactatcatatgcttaccgtaactt
gaaagtatttgcatttcttggctttatatacttgggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNNggttttaga
gctagaaatagcaagttaaataaggctagtcggtttcaTTTTTTTT

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

RNA quimérico contendo +67 tracrRNA (SF370 de *S. pyogenes*)
gagggcctatttcccatgattccctcatatttgcataatagatacagggctgttagagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatagtgacgtagaaagtaataattcttgggtagtttgcagttttaaataattgttttaaatggactatcatatgcttaccgtaactt
gaaagtatttgcatttcttggctttatatacttgggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNNggttttaga
gctagaaatagcaagttaaataaggctagtcggtttcaacttgaaaaagtgTTTTTTT

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

RNA quimérico que contém +85 tracrRNA (SF370 de *S. pyogenes*)
gagggcctatttcccatgattccctcatatttgcataatagatacagggctgttagagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatagtgacgtagaaagtaataattcttgggtagtttgcagttttaaataattgttttaaatggactatcatatgcttaccgtaactt
gaaagtatttgcatttcttggctttatatacttgggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNNggttttaga
gctagaaatagcaagttaaataaggctagtcggtttcaacttgaaaaagtgccaccgagtcggtgcTTTTTTT

(N = Sequência guia; primeira parte sublinhada = sequência

de emparelhamento tracr; segunda parte sublinhada =
sequência tracr; negrito = terminador)

CBh-NLS-SpCas9-NLS

CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCC
ATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG
ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA
TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCA
TTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTA
GTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCT
CCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTAAATTATTTGTGCAGCG
ATGGGGGCGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAG
GGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGGCGGCAGCCAATCAGAGCGGCGCGC
TCCGAAAGTTTCCTTTTATGGCGAGGCGGGCGGCGGGCGGCCCTATAAAAAGCGA
AGCGCGCGGGCGGGCGGGAGTCGCTGCGACGCTGCCTTCGCCCCGTGCCCCGCTCCG
CCGCCGCTCGCGCCGCCGCCCGGCTCTGACTGACCGCGTTACTCCACAGGTGA
GCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGGG
TTTAAGGGATGGTTGGTTGGTGGGTATTAATGTTTAATTACCTGGAGCACCTGCCT
GAAATCACTTTTTTTCAGGTTGGaccggtgccaccATGGACTATAAGGACCACGACGGAGA
CTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCTGATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATC
GGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCGGTGATCACCGACGAGTACAA
GGTGGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGA
AGAACCTGATCGGAGCCCTGCTGTTTCGACAGCGGCGAAACAGCCGAGGCCACCCGG
CTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTC
GGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCT
GAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGG
CCCTGGCCCATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGGCGACCTGAACC
CCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAG
CTGTTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTC
TGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCCGGCG
AGAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCC
AACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGA
CACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAAGTACGCCG
ACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATACCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGATAC

GACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCC
TGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACA
TTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAA
AAGATGGACGGCACCCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCG
GAAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGAGAGC
TGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGG
AAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAAACAGCAGATTTCGCTGGATGACCAGAAAAGAGCGAGGAAACCATCACCCC
CTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGC
GGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGC
CTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACC
GAGGGAATGAGAAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGG
ACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTAC
TTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC
AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTT
CCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACAC
TGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTC
GACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC
TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
GATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGAC
GACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGA
TAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCA
TCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAG
CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGAC
AGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
CAGCCAGATCCTGAAAGAACACCCCGTGAAAACACCCAGCTGCAGAACGAGAAG
CTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGA
CATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAA
GGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAG
AGCGACAACGTGCCCTCCGAAGAGGTCTGTGAAGAAGATGAAGAATACTGGCGGCA
GCTGCTGAACGCCAAGCTGATTACCCAGAGAAAAGTTCGACAATCTGACCAAGGCCG

NNNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaagataaggcttcatgccgaaatca
acaccctgtcattttatggcaggggtgttttcgttatttaa**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaagataaggcttcatgccgaaatca
acaccctgtcattttatggcaggggtgt**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttcagaagctacaagataaggctt
catgccgaaatcaacaccctgtcattttatggcaggggtgttttcgttatttaa**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaagataaggcttcatgccgaaatc
aacaccctgtcattttatggcaggggtgttttcgttatttaa**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacacccgtcattttatggcagggtg**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaagatttGAAAtaaatcttcgagaagctacaatgataaggcttcatgccgaaatcaacacccgtcattttatggcagggtgttttcgttatttaa**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacacccgtcattttatggcagggtgttttcgttatttaa**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR1 Cas9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacacccgtcattttatggcagggtg**TTTTTT**

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Exemplo de RNA quimérico LMD-9 CRISPR3 Cas9 de *S. thermophilus* (com PAM de NGGNG)

NNNNNNNNNNNNNNNNNNNNNNNNgttttagagctgtgGAAAcacagcgagtaaaaataaggcttagtccgtactcaact
gaaaagggtggcaccgattcgggtgTTTTT

(N = Sequência guia; primeira parte sublinhada = sequência de emparelhamento tracr; segunda parte sublinhada = sequência tracr; negrito = terminador)

Versão otimizada para codão de Cas9 do locus LMD-9 CRISPR3 de *S. thermophilus* (com NLS em ambas as terminações 5' e 3')

ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGACCA
AGCCCTACAGCATCGGCCTGGACATCGGCACCAATAGCGTGGGCTGGGCCGTGACC
ACCGACAACCTACAAGGTGCCCAGCAAGAAAATGAAGGTGCTGGGCAACACCTCCAA
GAAGTACATCAAGAAAAACCTGCTGGGCGTGCTGCTGTTTCGACAGCGGCATTACAG
CCGAGGGCAGACGGCTGAAGAGAACCGCCAGACGGCGGTACACCCGGCGGAGAAA
CAGAATCCTGTATCTGCAAGAGATCTTCAGCACCGAGATGGCTACCCTGGACGACG
CCTTC'TCCAGCGGCTGGACGACAGCTTCCTGGTGCCCGACGACAAGCGGGACAGC
AAGTACCCCATCTTCGGCAACCTGGTGGAAGAGAAGGCCTACCACGACGAGTTCCC
CACCATCTACCACCTGAGAAAGTACCTGGCCGACAGCACCAAGAAGGCCGACCTGA
GACTGGTGTATCTGGCCCTGGCCACATGATCAAGTACCGGGGCCACTTCCTGATCG
AGGGCGAGTTCAACAGCAAGAACAACGACATCCAGAAGAACTTCCAGGACTTCCTG

GACACCTACAACGCCATCTTCGAGAGCGACCTGTCCCTGGAAAAACAGCAAGCAGCT
GGAAGAGATCGTGAAGGACAAGATCAGCAAGCTGGAAAAGAAGGACCGCATCCTG
AAGCTGTTCCCCGGCGAGAAGAACAGCGGAATCTTCAGCGAGTTTCTGAAGCTGAT
CGTGGGCAACCAGGCCGACTTCAGAAAGTGCTTCAACCTGGACGAGAAAGCCAGCC
TGCACTTCAGCAAAGAGAGCTACGACGAGGACCTGGAAACCCTGCTGGGATATATC
GGCGACGACTACAGCGACGTGTTCTGAAGGCCAAGAAGCTGTACGACGCTATCCT
GCTGAGCGGCTTCCTGACCGTGACCGACAACGAGACAGAGGCCCCACTGAGCAGCG
CCATGATTAAGCGGTACAACGAGCACAAAGAGGATCTGGCTCTGCTGAAAGAGTAC
ATCCGGAACATCAGCCTGAAAACCTACAATGAGGTGTTCAAGGACGACACCAAGAA
CGGCTACGCCGGCTACATCGACGGCAAGACCAACCAGGAAGATTTCTATGTGTACC
TGAAGAAGCTGCTGGCCGAGTTTCGAGGGGGCCGACTACTTTCTGGAAAAAATCGAC
CGCGAGGATTTCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCTACCA
GATCCATCTGCAGGAAATGCGGGCCATCCTGGACAAGCAGGCCAAGTTCTACCCAT
TCCTGGCCAAGAACAAGAGCGGATCGAGAAGATCCTGACCTTCCGCATCCCTTACT
ACGTGGGCCCCCTGGCCAGAGGCAACAGCGATTTTGCCTGGTCCATCCGGAAGCGC
AATGAGAAGATCACCCCCTGGAACCTTCGAGGACGTGATCGACAAAGAGTCCAGCGC
CGAGGCCTTCATCAACCGGATGACCAGCTTCGACCTGTACCTGCCCAGGAAAAGG
TGCTGCCCAAGCACAGCCTGCTGTACGAGACATTCAATGTGTATAACGAGCTGACCA
AAGTGCGGTTTATCGCCGAGTCTATGCGGGACTACCAGTTCCTGGACTCCAAGCAGA
AAAAGGACATCGTGCGGCTGTACTTCAAGGACAAGCGGAAAGTGACCGATAAGGAC
ATCATCGAGTACCTGCACGCCATCTACGGCTACGATGGCATCGAGCTGAAGGGCAT
CGAGAAGCAGTTCAACTCCAGCCTGAGCACATACCACGACCTGCTGAACATTATCA
ACGACAAAGAATTTCTGGACGACTCCAGCAACGAGGCCATCATCGAAGAGATCATC
CACACCCTGACCATCTTTGAGGACCGCGAGATGATCAAGCAGCGGCTGAGCAAGTT
CGAGAACATCTTCGACAAGAGCGTGCTGAAAAAGCTGAGCAGACGGCACTACACCG
GCTGGGGCAAGCTGAGCGCCAAGCTGATCAACGGCATCCGGGACGAGAAGTCCGGC
AACACAATCCTGGACTACCTGATCGACGACGGCATCAGCAACCGGAACCTTCATGCA
GCTGATCCACGACGACGCCCTGAGCTTCAAGAAGAAGATCCAGAAGGCCCAGATCA
TCGGGGACGAGGACAAGGGCAACATCAAAGAAGTCGTGAAGTCCCTGCCCGGCAGC
CCCGCCATCAAGAAGGGAATCCTGCAGAGCATCAAGATCGTGGACGAGCTCGTGAA
AGTGATGGGCGGCAGAAAGCCCGAGAGCATCGTGGTGGAAATGGCTAGAGAGAAC

controle do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende adicionalmente duas ou mais sequências guia ligadas de forma operativa ao primeiro elemento regulador, em que, quando expressas, cada uma das duas ou mais sequências guia orienta a ligação específica de um complexo CRISPR de uma sequência alvo diferente numa célula eucariótica. Em algumas formas de realização, o sistema compreende ainda um terceiro elemento regulador, tal como um promotor de polimerase III, operacionalmente ligado à referida sequência tracr. Em algumas formas de realização, a sequência tracr apresenta, pelo menos, 50%, 60%, 70%, 80%, 90%, 95%, ou 99% de complementaridade de sequência ao longo do comprimento da sequência de emparelhamento tracr quando otimamente alinhadas. Em algumas formas de realização, o complexo CRISPR compreende uma ou mais sequências de localização nuclear de resistência suficiente para dirigir a acumulação do referido complexo CRISPR numa quantidade detetável no núcleo de uma célula eucariótica. Em algumas formas de realização, a enzima CRISPR é uma enzima do sistema CRISPR de tipo II. Em algumas formas de realização, a enzima CRISPR é uma enzima Cas9. Em algumas formas de realização, a enzima Cas9 é uma Cas9 de *S. pneumoniae*, *S. pyogenes*, ou *S. thermophilus*, e pode incluir uma Cas9 mutada derivado desses organismos. A enzima pode ser um homólogo ou ortólogo da Cas9. Em algumas formas de realização, a enzima CRISPR tem codões otimizados para expressão numa célula eucariótica. Em algumas formas de realização, a enzima CRISPR orienta a clivagem de uma ou duas cadeias na posição da sequência alvo. Em algumas formas de realização, a enzima CRISPR carece de atividade de clivagem da cadeia de DNA. Em algumas formas de realização, o primeiro elemento regulador é um promotor de polimerase III. Em algumas

CAGTACACCAATCAGGGCAAGAGCAACAGCCAGCAGAGACTGAAGAGACTGGAAA
 AGTCCCTGAAAGAGCTGGGCAGCAAGATTCTGAAAGAGAATATCCCTGCCAAGCTG
 TCCAAGATCGACAACAACGCCCTGCAGAACGACCGGCTGTACCTGTACTACCTGCA
 GAATGGCAAGGACATGTATACAGGCGACGACCTGGATATCGACCGCCTGAGCAACT
 ACGACATCGACCATATTATCCCCCAGGCCTTCCTGAAAGACAACAGCATTGACAAC
 AAAGTGCTGGTGTCTCCGCCAGCAACCGCGGCAAGTCCGATGATGTGCCCAGCCT
 GGAAGTCGTGAAAAAGAGAAAGACCTTCTGGTATCAGCTGCTGAAAAGCAAGCTGA
 TTAGCCAGAGGAAGTTCGACAACCTGACCAAGGCCGAGAGAGGCGGCCTGAGCCCT
 GAAGATAAGGCCGGCTTCATCCAGAGACAGCTGGTGGAAACCCGGCAGATCACCAA
 GCACGTGGCCAGACTGCTGGATGAGAAGTTTAACAACAAGAAGGACGAGAACAACC
 GGGCCGTGCGGACCGTGAAGATCATCACCTGAAGTCCACCCTGGTGTCCCAGTTCC
 GGAAGGACTTCGAGCTGTATAAAGTGCGCGAGATCAATGACTTTCACCACGCCCAC
 GACGCCTACCTGAATGCCGTGGTGGCTTCCGCCCTGCTGAAGAAGTACCCTAAGCTG
 GAACCCGAGTTCGTGTACGGCGACTACCCCAAGTACAACCTCCTTCAGAGAGCGGAA
 GTCCGCCACCGAGAAGGTGTACTTCTACTCCAACATCATGAATATCTTTAAGAAGTC
 CATCTCCCTGGCCGATGGCAGAGTGATCGAGCGGCCCTGATCGAAGTGAACGAAG
 AGACAGGCGAGAGCGTGTGGAACAAAGAAAGCGACCTGGCCACCGTGCGGCGGGT
 GCTGAGTTATCCTCAAGTGAATGTCTGTAAGAAGGTGGAAGAACAGAACCGGCC
 TGGATCGGGGCAAGCCCAAGGGCCTGTTCAACGCCAACCTGTCCAGCAAGCCTAAG
 CCCAACTCCAACGAGAATCTCGTGGGGGCCAAAGAGTACCTGGACCCTAAGAAGTA
 CGGCGGATACGCCGGCATCTCCAATAGCTTCACCGTGCTCGTGAAGGGCACAATCG
 AGAAGGGCGCTAAGAAAAAGATCACAACCGTGCTGGAATTTTCAGGGGATCTCTATC
 CTGGACCGGATCAACTACCGGAAGGATAAGCTGAACCTTTCTGCTGGAAAAAGGCTA
 CAAGGACATTGAGCTGATTATCGAGCTGCCTAAGTACTCCCTGTTTCAACTGAGCGA
 CGGCTCCAGACGGATGCTGGCCTCCATCCTGTCCACCAACAACAAGCGGGGCGAGA
 TCCACAAGGGAAACCAGATCTTCCTGAGCCAGAAATTTGTGAAACTGCTGTACCACG
 CCAAGCGGATCTCCAACACCATCAATGAGAACCACCGGAAATACGTGGAAAACAC
 AAGAAAGAGTTTGAGGAACTGTTCTACTACATCCTGGAGTTCAACGAGAACTATGTG
 GGAGCCAAGAAGAACGGCAAACTGCTGAACTCCGCCCTCCAGAGCTGGCAGAACCA
 CAGCATCGACGAGCTGTGCAGCTCCTTCATCGGCCCTACCGGCAGCGAGCGGAAGG
 GACTGTTTGAGCTGACCTCCAGAGGCTCTGCCGCCGACTTTGAGTTCCTGGGAGTGA
 AGATCCCCCGGTACAGAGACTACACCCCTCTAGTCTGCTGAAGGACGCCACCCTGA
 TCCACCAGAGCGTGACCGGCCTGTACGAAACCCGGATCGACCTGGCTAAGCTGGGC
 GAGGGAAAGCGTCCTGCTGCTACTAAGAAAGCTGGTCAAGCTAAGAAAAAGAAATA
 A

Exemplo 5: Otimização de RNA guia para Cas9
de *Streptococcus pyogenes* (referido como SpCas9).

Os Requerentes mutaram as sequências tracrRNA e de repetição direta, ou mutaram o RNA guia quimérico para aumentar os RNA nas células.

A otimização é baseada na observação de que havia trechos de timinas (T) no tracrRNA e no RNA guia, que poderiam levar à terminação precoce da transcrição pelo promotor pol 3. Portanto, os Requerentes geraram as seguintes sequências otimizadas que são apresentadas em seguida. O tracrRNA otimizado e a correspondente repetição direta otimizada são apresentados em pares.

tracrRNA 1 Otimizado (mutação a sublinhado):

**GGAACCATTCA1AACAGCATAGCAAGTTA1AATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTT**

Repetição direta 1 otimizada (mutação a sublinhado):

GTT2TAGAGCTATGCTGTT2TGAATGGTCCCAAAC

tracrRNA 2 Otimizado (mutação a sublinhado):

**GGAACCATTCAA1ACAGCATAGCAAGTTAA1ATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTT**

Repetição direta 2 otimizada (mutação a sublinhado):

GT2TTAGAGCTATGCTGT2TTGAATGGTCCCAAAC

Os Requerentes também otimizaram o RNA guia quimérico para uma atividade ótima em células eucarióticas.

RNA guia original:

**NNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT**

Sequência de RNA guia quimérico 1 otimizada:

NNNNNNNNNNNNNNNNNNNNNGTATTAGAGCTAGAAATAGCAAGTTAATAAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

Sequência de RNA guia quimérico 2 otimizada:

NNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCA
TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
GTGCTTTTTTT

Sequência de RNA guia quimérico 3 otimizada:

NNNNNNNNNNNNNNNNNNNNNGTATTAGAGCTATGCTGTATTGGAAACAATACAGC
ATAGCAAGTTAAATAAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTT

Os requerentes mostraram que o RNA guia quimérico otimizado funciona melhor tal como indicado na Figura 9. A experiência foi conduzida mediante a co-transfecção de células 293FT com Cas9 e uma cassete de DNA de U6- RNA guia para expressar uma das quatro formas de RNA mostradas anteriormente. O alvo do RNA guia é o mesmo sítio alvo no locus EMX1 humano: "GTCACCTCCAATGACTAGGG"

Exemplo 6: Otimização de LMD-9 CRISPR1 Cas9 de *Streptococcus thermophilus* (denominado St1Cas9).

Os requerentes desenharam RNA guias quiméricos tal como se mostra na Figura 12.

O RNA guia St1Cas9 pode ser submetido ao mesmo tipo de otimização como o efetuado para RNA guia SpCas9, quebrando os fragmentos de poli timinas (T).

Exemplo 7: Melhoria do sistema Cas9 para a aplicação *in vivo*

Os Requerentes conduziram uma busca Meta genômica para uma Cas9 com baixo peso molecular. A maioria dos homólogos Cas9 são bastante grandes. Por exemplo, a SpCas9 tem aproximadamente 1368aa de comprimento, sendo muito grande para ser facilmente empacotada em vetores virais para administração. Algumas das sequências podem ter sido anotadas de forma errônea e, portanto, a frequência exata para cada comprimento pode não ser necessariamente precisa. No entanto, fornece um vislumbre da distribuição de proteínas Cas9 e sugerem que existem homólogos Cas9 mais curtos.

Através da análise computacional, os Requerentes verificaram que na estirpe bacteriana de *Campylobacter* existem duas proteínas Cas9 com menos do que 1000 aminoácidos. A sequência para um Cas9 de *Campylobacter jejuni* é apresentada em seguida. Com este comprimento, a CjCas9 pode ser facilmente empacotada em AAV, lentivírus, Adenovírus, e outros vetores virais para a administração robusta em células primárias e em modelos animais *in vivo*.

> Cas9 de *Campylobacter jejuni* (CjCas9)

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARR
 KARLNHLKHLIANEFKLNIEDYQSFDES LAKAYKGS LISP YELRFRALNELL SKQDFAR
 VILHI AKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSK
 EFTNVRNKKESYERCIAQSFLKDELKLFKKQREFGFSFSKKFEEVLSVAFYKRALKDFS
 HLVGNC SFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDL NALLNEVLK
 NGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDIT
 LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNE
 LNLKVAINEDKKDFLPAFNETYKDEVTPVVLRAIKEYRKVLNALLKKYGVHKNIE
 LAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGKINSKNILKRLFKKEQKEFCA
 YSGEKIKISDLQDEKMLEIDHIYPYRSFDDSYMNVLVFTKQNEKLNQTPFEAFGND
 AKWQKIEVLAKNLP TKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVNLYTKDYL
 DFLPLSDDENTKLN DTQKGSKVHVEAKSGMLTSALRHTWGFS AKDRNNHLHHAIDAVI
 IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIF
 VSKPERKKPSGALHEETFRKEEEFYQSYGGKEGV LKALELGKIRKVNGKIVKNGDMFR
 VDFKHKKTNKFYAVPIYTMDFA LKVLPNKAVARSKKGEIKDWILMDENYEFCSLYK
 DSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKS
 IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK.

O elemento tracrRNA putativo para esta CjCas9 é:

**TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTG
 CGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT**

A sequência da Repetição Direta é:

ATTTTACCATAAGAAATTTAAAAAGGGACTAAAAC

A estrutura de co-dobra do tracrRNA e a repetição direta são fornecidos na Figura 6.

Um exemplo de um RNA guia quimérico para CjCas9 é:

**NNNNNNNNNNNNNNNNNNNNNGUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUU
 UGCGGGACUCUGCGGGGUACAAUCCCCUAAAACCGCUUUU**

Os Requerentes também têm otimizado RNA guia Cas9 utilizando métodos *in vitro*. A Figura 18 mostra os dados da otimização de RNA guia quimérico StlCas9 *in vitro*.

Embora as formas de realização preferidas da presente invenção tenham sido aqui apresentadas e descritas, será evidente para os peritos na técnica que tais formas de realização são fornecidas apenas a modo de um exemplo. Os

peritos na técnica encontrarão numerosas variações, alterações e substituições sem se afastar da invenção. Deve ser entendido que na prática da invenção podem ser empregues várias alternativas para as formas de realização da invenção aqui descrita. Pretende-se que as reivindicações seguintes definam o âmbito da invenção e os métodos e estruturas dentro do âmbito destas reivindicações e seus equivalentes sejam assim cobertos.

Exemplo 8: Otimização de sgRNA de Sa

Os Requerentes desenharam cinco variantes sgRNA para SaCas9 para uma arquitetura truncada ideal com maior eficiência de clivagem. Além disso, foi testado o sistema duplex nativo de repetição direta:tracr junto com os sgRNAs. As sequências guia com comprimentos indicados foram co-transfectadas com SaCas9 e testadas em células HEK 293FT para a atividade. Um total de 100 ng de amplicon U6-PCR sgRNA (ou 50 ng de repetição direta e 50 ng de tracrRNA) e 400 ng de plasmídeo SaCas9 foram co-transfectados em 200.000 de hepatócitos Hepal-6 de ratinho, e o DNA foi colhido 72 horas após a transfecção para análise de SURVEYOR. Os resultados são mostrados na Fig. 23.

Referências:

1. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636-646 (2010).

2. Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843-1846 (2011).
3. Stoddard, B.L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38**, 49-95 (2005).
4. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* **55**, 58-63 (2006).
5. Sung, C.K., Li, H., Claverys, J.P. & Morrison, D.A. An *rpsL* cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **67**, 5190-5196 (2001).
6. Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. & Court, D.L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206-223 (2009).
7. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
8. Deveau, H., Garneau, J.E. & Moineau, S. CRISPR-Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475-493 (2010).
9. Horvath, P. & Barrangou, R. CRISPR-Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170 (2010).
10. Terns, M.P. & Terns, R.M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321-327 (2011).
11. van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M. & Brouns, S.J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends. Biochem. Sci.* **34**, 401-407 (2009).

12. Brouns, S.J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
13. Carte, J., Wang, R., Li, H., Terns, R.M. & Terns, M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489-3496 (2008).
14. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).
15. Hatoum-Aslan, A., Maniv, I. & Marraffini, L.A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 21218-21222 (2011).
16. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355-1358 (2010).
17. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390-1400 (2008).
18. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* (2012).
19. Makarova, K.S., Aravind, L., Wolf, Y.I. & Koonin, E.V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct.* **6**, 38 (2011).
20. Barrangou, R. RNA-mediated programmable DNA cleavage. *Nat. Biotechnol.* **30**, 836-838 (2012).

21. Brouns, S.J. Molecular biology. A Swiss army knife of immunity. *Science* **337**, 808-809 (2012).
22. Carroll, D. A CRISPR Approach to Gene Targeting. *Mol. Ther.* **20**, 1658-1660 (2012).
23. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L.A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12**, 177-186 (2012).
24. Sapranaukas, R. et al. The *Streptococcus thermophilus* CRISPR-Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* (2011).
25. Semenova, E. et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* (2011).
26. Wiedenheft, B. et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* (2011).
27. Zahner, D. & Hakenbeck, R. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J. Bacteriol.* **182**, 5919-5921 (2000).
28. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192-221 (2006).
29. Motamedi, M.R., Szigety, S.K. & Rosenberg, S.M. Double-strand-break repair recombination in *Escherichia coli*: physical evidence for a DNA replication mechanism in vivo. *Genes Dev.* **13**, 2889-2903 (1999).
30. Hosaka, T. et al. The novel mutation K87E in ribosomal protein S 12 enhances protein synthesis

activity during the late growth phase in *Escherichia coli*. *Mol. Genet. Genomics* **271**, 317-324 (2004).

31. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15748-15753 (2003).

32. Edgar, R. & Qimron, U. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J. Bacteriol.* **192**, 6291-6294 (2010).

33. Marraffini, L.A. & Sontheimer, E.J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568-571 (2010).

34. Fischer, S. *et al.* An archaeal immune system can detect multiple Protospacer Adjacent Motifs (PAMs) to target invader DNA. *J. Biol. Chem.* **287**, 33351-33363 (2012).

35. Gudbergisdottir, S. *et al.* Dynamic properties of the *Sulfolobus* CRISPR-Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35-49 (2011).

36. Wang, H.H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nat Methods* **9**, 591-593 (2012).

37. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR-Cas Systems. *Science* In press (2013).

38. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* In press (2013).

39. Hoskins, J. *et al.* Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183**, 5709-5717 (2001).

40. Havarstein, L.S., Coomaraswamy, G. & Morrison, D.A. An unmodified heptadecapeptide pheromone induces

formas de realização, o segundo elemento regulador é um promotor de polimerase II. Em algumas formas de realização, a sequência guia tem de comprimento pelo menos 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10-30, ou entre 15-25, ou entre 15-20 nucleótidos.

Num aspeto, a presente divulgação prevê um método para modificar um polinucleótido alvo numa célula eucariótica. Em algumas formas de realização, o método compreende que seja permitido que um complexo CRISPR se ligue ao polinucleótido alvo para efetuar a clivagem do referido polinucleótido alvo modificando assim o polinucleótido alvo, em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com uma sequência guia hibridada com uma sequência alvo no interior do referido polinucleótido alvo, em que a referida sequência guia é ligada a uma sequência de emparelhamento tracr que por sua vez hibrida com uma sequência tracr. Em algumas formas de realização, a referida clivagem compreende a clivagem de uma ou duas cadeias na posição da sequência alvo pela referida enzima CRISPR. Em algumas formas de realização, a referida clivagem resulta na diminuição da transcrição de um gene alvo. Em algumas formas de realização, o método compreende ainda a reparação do dito polinucleótido alvo clivado por recombinação homóloga com um polinucleótido molde exógeno, em que a referida reparação resulta numa mutação compreendendo uma inserção, deleção ou substituição de um ou mais nucleótidos do referido polinucleótido alvo. Em algumas formas de realização, a referida mutação resulta na alteração de um ou mais de aminoácidos numa proteína expressa a partir de um gene que compreende a sequência alvo. Em algumas formas de realização, o método compreende ainda a entrega de um ou mais vetores para a referida

competence for genetic transformation in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11140-11144 (1995).

41. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol.* **150**, 815-825 (1982).

42. Horton, R.M. In Vitro Recombination and Mutagenesis of DNA : SOEing Together Tailor-Made Genes. *Methods Mol. Biol.* **15**, 251-261 (1993).

43. Podbielski, A., Spellerberg, B., Woischnik, M., Pohl, B. & Luttmann, R. Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS). *Gene* **177**, 137-147 (1996).

44. Husmann, L.K., Scott, J.R., Lindahl, G. & Stenberg, L. Expression of the Arp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*. *Infection and immunity* **63**, 345-348 (1995).

45. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).

46. Tangri S, et al. ("Rationally engineered therapeutic proteins with reduced immunogenicity" *J Immunol.* 2005 Mar 15; 174(6):3187-96.

Embora as realizações preferidas da presente invenção tenham sido aqui apresentadas e descritas, será óbvio para os peritos na técnica que tais formas de realização são fornecidas por meio de apenas exemplo. Numerosas variações, alterações e substituições ocorrerão aos peritos na técnica sem se afastar da invenção. Deve ser entendido que várias

alternativas para as formas de realização da invenção aqui descritas podem ser empregues na prática da invenção.

Outros aspetos da divulgação são definidos nas seguintes cláusulas numeradas.

1. Uma composição de origem não natural ou obtida por engenharia compreendendo:

A) uma sequência de polinucleótidos de RNA quimérico (chiRNA) do sistema CRISPR-Cas, em que a sequência de polinucleótidos compreende

- (a) uma sequência guia capaz de hibridar com uma sequência alvo numa célula eucariótica,
- (b) uma sequência de emparelhamento tracr, e
- (c) uma sequência tracr

em que (a), (b) e (c) se encontrarem dispostos na orientação de 5' para 3',

em que, quando transcrita, a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia dirige a ligação específica da sequência de um complexo CRISPR para a sequência alvo,

em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com (1) a sequência guia que é hibridada com a sequência alvo, e (2) a sequência de emparelhamento tracr que é hibridada com a sequência tracr,

ou

B) um sistema de enzima CRISPR, em que o sistema é codificado por um sistema de vetor compreendendo um ou mais vetores compreendendo

I. um primeiro elemento regulador ligado operacionalmente a uma sequência de polinucleótido de RNA quimérico (chiRNA) do sistema CRISPR-Cas, em que a sequência de polinucleótido compreende

- (a) uma ou mais sequências guia capazes de hibridar com uma ou mais sequências alvo numa célula eucariota,
- (b) uma sequência de emparelhamento tracr, e
- (c) uma ou mais sequências tracr, e

II. um segundo elemento regulador ligado operacionalmente a uma sequência codificante de enzima que codifica uma enzima CRISPR compreendendo, pelo menos, uma ou mais sequências de localização nuclear,

em que (a), (b) e (c) se encontram dispostas na orientação de 5' para 3',

em que os componentes I e II estão localizados no mesmo ou em diferentes vetores do sistema,

em que, quando transcrita a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia dirige a ligação específica da sequência de um complexo CRISPR a uma sequência alvo, em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que é hibridada com a sequência alvo, e (2) a sequência de

emparelhamento tracr que é hibridada com a sequência tracr,

ou

C) um sistema de enzima CRISPR multiplexo, em que o sistema é codificado por um sistema de vetor que compreende um ou mais vetores compreendendo

I. um primeiro elemento regulador ligado operacionalmente a uma sequência de polinucleótido de RNA quimérico (chiRNA) do sistema CRISPR-Cas, em que a sequência de polinucleótido compreende

(a) uma ou mais sequências guia capazes de hibridar com uma ou mais sequências alvo numa célula eucariota,

(b) uma sequência de emparelhamento tracr, e

(c) um ou mais sequências tracr, e

II. um segundo elemento regulador ligado operacionalmente a uma sequência codificante de enzima que codifica uma enzima CRISPR compreendendo, pelo menos, uma ou mais sequências de localização nuclear,

em que (a), (b) e (c) se encontram dispostas na orientação de 5' para 3',

em que os componentes I e II estão localizados no mesmo ou em diferentes vetores do sistema, em que, quando transcrita a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia dirige a ligação específica da sequência de um complexo CRISPR a uma sequência alvo,

em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que é

hibridada com a sequência alvo, e (2) a sequência de emparelhamento tracr que é hibridada com a sequência tracr

em que no sistema multiplexo múltiplas sequências de polinucleótidos de chiRNA são utilizadas

ou

D) um sistema de enzima CRISPR multiplexo, em que o sistema é codificado por um sistema de vetor que compreende um ou mais vetores compreendendo

I. um primeiro elemento regulador ligado operacionalmente a

(a) uma ou mais sequências guia capazes de hibridar com uma sequência alvo numa célula, e

(b) pelo menos uma ou mais sequências de emparelhamento tracr,

II. um segundo elemento regulador ligado operacionalmente a uma sequência codificante de enzima que codifica uma enzima CRISPR, e

III. um terceiro elemento regulador ligado operacionalmente a uma sequência tracr,

em que os componentes I, II e III, estão localizados no mesmo ou em diferentes vetores do sistema,

em que, quando transcrita a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia dirige a ligação específica da sequência de um complexo CRISPR a uma sequência alvo, em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que é hibridada com a sequência alvo, e (2) a sequência de

emparelhamento tracr que é hibridada com a sequência tracr, e em que no sistema multiplexo múltiplas sequências guia e uma única sequência tracr são utilizadas;

e

em que, na sequência de polinucleótido de A), ou no sistema de B), C) ou D), uma ou mais das sequências guia, tracr e de emparelhamento tracr são modificadas para melhorar a estabilidade.

2. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 1 em que a modificação compreende uma estrutura secundária obtida por engenharia.

3. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 1 ou cláusula 2, em que a modificação compreende uma redução numa região de hibridação entre a sequência de emparelhamento tracr e a sequência tracr.

4. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende a fusão da sequência de emparelhamento tracr e a sequência tracr através de uma alça artificial.

5. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende que a sequência de tracr apresenta um comprimento entre 40 e 120 pb.

6. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência tracr tem entre 40 pb e o comprimento total da sequência tracr.

7. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência tracr inclui, pelo menos, os nucleótidos 1-67 do correspondente tracrRNA do tipo selvagem.

8. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência tracr inclui, pelo menos, os nucleótidos 1-85 do correspondente tracrRNA do tipo selvagem.

9. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência tracr compreende os nucleótidos correspondentes aos nucleótidos 1-67 de tracrRNA de Cas9 de *S. pyogenes* do tipo selvagem.

10. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência tracr compreende os nucleótidos correspondentes aos nucleótidos 1-85 de tracrRNA de Cas9 de *S. pyogenes* do tipo selvagem.

11. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 9 em que a sequência tracr consiste essencialmente em nucleótidos correspondentes aos nucleótidos 1-67 de tracrRNA de Cas9 de *S. pyogenes* do tipo selvagem.

12. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 10 em que a sequência tracr consiste essencialmente em nucleótidos correspondentes aos nucleótidos 1-85 de tracrRNA de Cas9 de *S. pyogenes* do tipo selvagem.

13. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende a otimização de sequência.

14. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 13 em que a modificação compreende a redução das sequências poli-T nas sequências tracr e/ou de emparelhamento tracr.

15. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 14, em que um ou mais T presentes numa sequência poli-T da sequência do tipo selvagem relevante foi substituído por um nucleótido que não seja T.

16. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 13, 14, ou 15, em que a sequência modificada não compreende qualquer sequência poli-T com mais de 4 T contíguos.

17. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende a adição de uma sequência de finalização poli-T.

18. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 17, em que a modificação

compreende a adição de uma sequência de finalização poli-T em sequências tracr e/ou de emparelhamento tracr.

19. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 17 ou 18, em que a modificação compreende a adição de uma sequência de finalização poli-T na sequência guia.

20. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende a alteração de alças e/ou forquilhas.

21. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 20 em que a modificação compreende que sejam fornecidas um mínimo de duas forquilhas na sequência guia.

22. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 20 ou 21, em que a modificação compreende que seja fornecida uma forquilha formada por complementaridade entre a sequência tracr e de emparelhamento tracr.

23. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 20 a 22 em que a modificação compreende que seja fornecida uma ou mais forquilhas na extremidade 3' da sequência tracrRNA.

24. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 20 a 23 em que a modificação compreende que seja fornecida uma ou mais forquilhas

adicionais adicionadas à extremidade 3' da sequência guia.

25. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende uma extensão da extremidade 5' da sequência guia.

26. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 25 em que a modificação compreende que sejam fornecidas uma ou mais forquilhas na extremidade 5' da sequência guia.

27. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 25 ou 26, em que a modificação compreende a anexação da sequência (5'-AGGACGAAGTCCTAA) à extremidade 5' da sequência guia.

28. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende que seja fornecida reticulação ou que seja fornecido um ou mais nucleótidos modificados na sequência de polinucleótidos.

29. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 28, em que os nucleótidos modificados são fornecidos em qualquer ou em todas as sequências tracr, de emparelhamento tracr, e/ou guia, e/ou na sequência codificante da enzima, e/ou nas sequências do vetor.

30. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 28 ou 29, em que fornecer

célula eucariótica, em que um ou mais vetores dirigirem a expressão de um ou mais dos seguintes: a enzima CRISPR, a sequência guia ligada à sequência de emparelhamento tracr, e a sequência tracr. Em algumas formas de realização, os referidos vetores são fornecidos para a célula eucariótica num sujeito. Em algumas formas de realização, referida modificação tem lugar na referida célula eucariótica de uma cultura de células. Em algumas formas de realização, o método compreende ainda o isolamento da referida célula eucariótica de um sujeito antes da referida modificação. Em algumas formas de realização, o método compreende ainda o retorno da referida célula e/ou células eucarióticas delas derivadas para o referido sujeito.

Num aspeto, a presente divulgação prevê um método para a modificação da expressão de um polinucleótido numa célula eucariótica. Em algumas formas de realização, o método compreende que seja permitido que um complexo CRISPR se ligue ao polinucleótido de modo que a referida ligação resulta num aumento ou diminuição da expressão do referido polinucleótido; em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com uma sequência guia hibridada com uma sequência alvo dentro do referido polinucleótido, em que a referida sequência guia é ligada a uma sequência de emparelhamento tracr que por sua vez hibrida com uma sequência tracr. Em algumas formas de realização, o método compreende ainda a entrega de um, ou mais, vetores para as referidas células eucarióticas, em que um ou mais vetores dirigem a expressão de uma ou mais: enzima CRISPR, sequência guia ligada à sequência de emparelhamento tracr e sequência tracr.

nucleótidos modificados compreende a inclusão de pelo menos um nucleótido que não ocorre de modo natural, ou um nucleótido modificado, ou seus análogos.

31. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 30, em que os nucleótidos modificados são modificados na região da ribose, do fosfato, e/ou da base.

32. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 30, em que o nucleótido modificado é selecionado a partir do grupo que consiste em análogos 2'-O-metilo, análogos 2'-desoxi, ou análogos 2'-fluoro.

33. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR da cláusula 30, em que o nucleótido modificado é selecionado a partir do grupo que consiste de 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina.

34. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende duas forquilhas.

35. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende três forquilhas.

36. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende um máximo de cinco forquilhas.

37. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a enzima CRISPR é uma enzima do sistema CRISPR de tipo II.

38. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a enzima CRISPR é uma enzima Cas9.

39. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a enzima CRISPR compreende menos de mil aminoácidos.

40. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer uma das cláusulas 1 a 38, em que a enzima CRISPR compreende menos de quatro mil aminoácidos.

41. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a enzima Cas9 é StCas9 ou StlCas9.

42. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a enzima Cas9 é uma enzima Cas9 de um organismo selecionado a partir do grupo que compreende o género *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* ou *Corynebacter*.

43. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a

enzima CRISPR é uma nuclease que dirige a clivagem de ambas as cadeias no local da sequência alvo.

44. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer número anterior, em que o primeiro elemento regulador é um promotor de polimerase III.

45. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que o segundo elemento regulador é um promotor de polimerase II.

46. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a sequência guia compreende pelo menos quinze nucleótidos.

47. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer cláusula anterior, em que a modificação compreende uma sequência tracr otimizada e/ou uma sequência guia de RNA otimizada, e/ou uma estrutura dobra da sequência tracr e/ou sequência ou sequências de emparelhamento tracr e/ou estruturas secundárias para estabilização da sequência tracr e/ou uma sequência tracr com uma região reduzida de emparelhamento de bases e/ou elementos de RNA fundidos de sequência tracr; e/ou, em sistema multiplex existem dois RNA compreendendo um marcador e compreendendo uma pluralidade de guias ou um RNA que compreende uma pluralidade de quiméricos.

48. O chiRNA do sistema CRISPR-Cas ou o sistema de enzima CRISPR de qualquer uma das cláusulas 1-47, em que a enzima CRISPR tem codões otimizados para a expressão numa célula eucariótica.

49. A composição de qualquer cláusula precedente, em que a composição compreende uma sequência de polinucleótido de RNA quimérico (chiRNA) do sistema CRISPR-Cas.

50. A composição da cláusula 49, em que a composição compreende ainda uma sequência de polinucleótidos que codifica uma enzima CRISPR compreendendo pelo menos uma ou mais sequências de localização nuclear.

51. O sistema de enzima CRISPR de qualquer cláusula anterior.

52. O sistema de enzima CRISPR multiplexo de qualquer cláusula anterior.

53. O produto de transcrição ou tradução da composição de cláusula 49 ou 50, o sistema de enzima CRISPR da cláusula 51, ou o sistema de enzima CRISPR multiplexo da cláusula 52.

Listagem de Sequências

<110> THE BROAD INSTITUTE, INC. MASSACHUSETTS INSTITUTE OF TECHNOLOGY

<120> ENGENHARIA DE SISTEMAS, MÉTODOS E COMPOSIÇÕES GUIA OTIMIZADAS PARA A MANIPULAÇÃO DE SEQUÊNCIAS

<130> PC927381EPA

<140> PCT/US2013/074819

<141> 2013-12-12

<150> 61/836,127

<151> 2013-06-17

<150> 61/835,931

<151> 2013-06-17

<150> 61/828,130

<151> 2013-05-28

<150> 61/819,803

<151> 2013-05-06

<150> 61/814,263

<151> 2013-04-20

<150> 61/806,375

<151> 2013-03-28

<150> 61/802,174

<151> 2013-03-15

<150> 61/791,409

<151> 2013-03-15

<150> 61/769,046

<151> 2013-02-25

<150> 61/758,468

<151> 2013-01-30

<150> 61/748,427

<151> 2013-01-02

<150> 61/736,527

<151> 2012-12-12

<160> 264

<170> PatentIn version 3.5

<210> 1

<211> 15

<212> DNA

<213> Sequência Artificial

<220>

<221>

<223>/nota="Descrição de Sequência Artificial:
Oligonucleótido sintético"


```

<400> 1
aggacgaagt cctaa                                     15

<210> 2
<211> 7
<212> PRT
<213> Virus Simio 40

<400> 2
Pro Lys Lys Lys Arg Lys Val
1                               5

<210> 3
<211> 16
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição de Organismo Desconhecido:
        Sequência NLS da nucleoplasmina bipartida"

<400> 3
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1           5           10           15

<210> 4
<211> 9
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição de Organismo
        Desconhecido: Sequência NLS C-myc"

<400> 4
Pro Ala Ala Lys Arg Val Lys Leu Asp
1           5

<210> 5
<211> 11
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição de Organismo
        Desconhecido: Sequência NLS C-myc"

<400> 5
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1           5           10

<210> 6
<211> 38
<212> PRT
<213> Homo sapiens

```

```

<400> 6
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1          5          10          15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
          20          25          30

Arg Asn Gln Gly Gly Tyr
          35

<210> 7
<211> 42
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição de Organismo Desconhecido: Sequência
do domínio IBB da importina-alpha"

<400> 7
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1          5          10          15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
          20          25          30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
          35          40

<210> 8
<211> 8
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição do Organismo Desconhecido:
Sequência da proteína Mioma T"

<400> 8
Val Ser Arg Lys Arg Pro Arg Pro
1          5

<210> 9
<211> 8
<212> PRT
<213> Organismo Desconhecido

<220>
<221> source
<223> /nota="Descrição do Organismo Desconhecido:
Sequência da proteína Mioma T"

<400> 9

```

Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 10
<211> 8
<212> PRT
<213> Homo sapiens

<400> 10
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 11
<211> 12
<212> PRT
<213> Mus musculus

<400> 11
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 12
<211> 5
<212> PRT
<213> Virus Influenza

<400> 12
Asp Arg Leu Arg Arg
1 5

<210> 13
<211> 7
<212> PRT
<213> Virus Influenza

<400> 13
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 14
<211> 10
<212> PRT
<213> Virus da Hepatitis delta

<400> 14
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 15
<211> 10
<212> PRT
<213> Mus musculus

<400> 15
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 16
 <211> 20
 <212> PRT
 <213> Homo sapiens

<400> 16
 Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
 1 5 10 15

Lys Ser Lys Lys
 20

<210> 17
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 17
 Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
 1 5 10 15

Lys

<210> 18
 <211> 27
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t or g

<220>
 <221> base_modificada
 <222> (21)..(22)
 <223> a, c, t, g, outro ou desconhecido

<400> 18
 nnnnnnnnnn nnnnnnnnnn nnagaaw

27

<210> 19
 <211> 19
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Oligonucleótido sintético"

Num aspeto, a presente divulgação prevê um método para gerar um modelo de célula eucariótica compreendendo um gene mutado da doença. Em algumas formas de realização, um gene da doença é qualquer gene associado com um aumento no risco de ter ou desenvolver uma doença. Em algumas formas de realização, o método compreende (a) introdução de um ou mais vetores numa célula eucariótica, em que o um ou mais vetores de controlam a expressão de um ou mais dos seguintes: uma enzima CRISPR, uma sequência guia ligada a uma sequência de emparelhamento tracr, e uma sequência tracr; e (b) permitir que um complexo CRISPR se ligue a um polinucleótido alvo para efetuar a clivagem do polinucleótido alvo no interior do referido gene da doença, em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que é hibridada com a sequência alvo dentro do polinucleótido alvo, e (2) a sequência de emparelhamento tracr que é hibridada com a sequência tracr, gerando assim um modelo de célula eucariótica compreendendo um gene mutante da doença. Em algumas formas de realização, a referida clivagem compreende a clivagem de uma ou duas cadeias na posição da sequência alvo pela referida enzima CRISPR. Em algumas formas de realização, a referida clivagem resulta na diminuição da transcrição de um gene alvo. Em algumas formas de realização, o método compreende ainda reparar o referido polinucleótido alvo clivado por recombinação homóloga com um polinucleótido molde exógeno, em que a referida reparação resulta numa mutação compreendendo uma inserção, deleção ou substituição de um ou mais nucleótidos do referido polinucleótido alvo. Em algumas formas de realização, a referida mutação resulta numa ou mais alterações de aminoácidos na expressão de uma proteína a partir de um gene que compreende a sequência alvo.

<220>
 <221> base_modificada
 <222> (1)..(12)
 <223> a, c, t ou g

<220>
 <221> base_modificada
 <222> (13)..(14)
 <223> a, c, t, g, outro ou desconhecido

<400> 19
 nnnnnnnnnn nnnnagaaw

19

<210> 20
 <211> 27
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t ou g

<220>
 <221> base_modificada
 <222> (21)..(22)
 <223> a,c,t,g, outro ou desconhecido

<400> 20
 nnnnnnnnnn nnnnnnnnnn nnagaaw

27

<210> 21
 <211> 18
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(11)
 <223> a, c, t ou g

<220>
 <221> base_modificada
 <222> (12)..(13)
 <223> a, c, t, g, outro ou desconhecido

<400> 21

nnnnnnnnnn nnnagaaw

18

<210> 22
 <211> 137
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 22
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60
 gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg caggggtgttt 120
 tcgttattta attttttt 137

<210> 23
 <211> 123
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 23
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60
 gcttcatgcc gaaatcaaca cctgttcatt ttatggcagg gtgttttcgt tatttaattt 120
 ttt 123

<210> 24
 <211> 110
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial:
 Polinucleótido sintético"

<220>

```

<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 24
nnnnnnnnnn nnnnnnnnnn gtttttgtac totcagaaat gcagaagcta caaagataag      60
ggttcattgcc gaaatcaaca cctgtcatt ttatggcagg gtgttttttt      110

<210> 25
<211> 102
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial:
        Polinucleótido sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 25
nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc      60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtcgtttt tt      102

<210> 26
<211> 88
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial:
        Oligonucleótido sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 26
nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc      60
cgttatcaac ttgaaaaagt gttttttt      88

<210> 27
<211> 76
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial:
        Oligonucleótido sintético"

```



```

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 27
nnnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc      60
cgttatcatt tttttt                                                                76

<210> 28
<211> 12
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição da Sequência Artificial:
      Oligonucleótido sintético"

<400> 28
guuuuagagc ua                                                                12

<210> 29
<211> 33
<212> DNA
<213> Homo sapiens

<400> 29
ggacatcgat gtcacctcca atgactaggg tgg                                          33

<210> 30
<211> 33
<212> DNA
<213> Homo sapiens

<400> 30
cattggaggt gacatcgatg tcctcccccatt tgg                                        33

<210> 31
<211> 33
<212> DNA
<213> Homo sapiens

<400> 31
ggaagggcct ggtccgagc agaagaagaa ggg                                          33

<210> 32
<211> 33
<212> DNA
<213> Homo sapiens

<400> 32
ggtgacgaga gggccgaga ttgggtgttc agg                                          33

```

```

<210> 33
<211> 33
<212> DNA
<213> Homo sapiens

<400> 33
atgcaggagg gtggcgagag gggccgagat tgg 33

<210> 34
<211> 21
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Primer
        sintético"

<400> 34
aaaaccaccc ttctctctgg c 21

<210> 35
<211> 21
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Primer
        sintético"

<400> 35
ggagattgga gacacggaga g 21

<210> 36
<211> 20
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Primer
        sintético"

<400> 36
ctggaaagcc aatgcctgac 20

<210> 37
<211> 20
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Primer
        sintético"

<400> 37
ggcagcaaac tccttgctct 20

```

<210> 38
 <211> 12
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 38
 gtttttagagc ta 12

<210> 39
 <211> 335
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 39
 gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag 60
 ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
 aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180
 atgcttaccg taacttgaaa gtatttcgat ttcttggcctt tatatatctt gtggaaagga 240
 cgaaacaccg gaaccattca aaacagcata gcaagttaaa ataaggctag tccgttatca 300
 acttgaaaaa gtggcaccga gtcggtgctt ttttt 335

<210> 40
 <211> 423
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 40
 gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag 60
 ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
 aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180
 atgcttaccg taacttgaaa gtatttcgat ttcttggcctt tatatatctt gtggaaagga 240
 cgaaacaccg gtagtattaa gtattgtttt atggctgata aatttctttg aatttctcct 300
 tgattatttg ttataaaaagt tataaaaataa tcttgttgga accattcaaa acagcatage 360

```

aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cgggtgctttt 420
ttt 423

```

```

<210> 41
<211> 339
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<400> 41
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg ggttttagag ctatgctggt ttgaatggtc ccaaaacggg tcttcgagaa 300
gacgttttag agctatgctg tttgaatgg tcccaaac 339

```

```

<210> 42
<211> 309
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<400> 42
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag ttaaaataag 300
gctagtcgg 309

```

```

<210> 43
<211> 1648
<212> PRT
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polipéptido
        sintético"

```

<400> 43

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
 1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
 20 25 30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
 35 40 45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
 50 55 60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
 65 70 75 80

Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
 85 90 95

Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
 100 105 110

Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
 115 120 125

Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
 130 135 140

Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
 145 150 155 160

Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
 165 170 175

Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
 180 185 190

Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
 195 200 205

Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
 210 215 220

Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
 225 230 235 240

Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
 245 250 255

Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
260 265 270

Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
275 280 285

Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
290 295 300

Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
325 330 335

Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
340 345 350

Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
355 360 365

Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
370 375 380

Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
385 390 395 400

Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
405 410 415

Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
420 425 430

Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
435 440 445

Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg
450 455 460

Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
465 470 475 480

Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
485 490 495

Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile

500					505					510					
Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln
		515					520					525			
Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu
		530				535					540				
Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr
545					550					555					560
Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro
				565					570					575	
Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe
			580					585						590	
Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe
		595					600					605			
Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp
	610					615					620				
Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile
625					630					635					640
Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu
				645					650					655	
Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu
			660					665					670		
Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys
		675				680						685			
Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys
	690					695					700				
Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp
705					710					715					720
Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile
				725					730					735	
His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val
			740					745					750		

Num aspeto, a presente divulgação prevê um método para o desenvolvimento de um agente biologicamente ativo que modula um evento de sinalização celular associado com um gene da doença. Em algumas formas de realização, um gene da doença é qualquer gene associado com um aumento no risco de ter ou desenvolver uma doença. Em algumas formas de realização, o método compreende (a) pôr em contacto um composto de teste com uma célula modelo de qualquer uma das formas de realização descritas; e (b) detetar uma alteração numa leitura que é indicativa de uma redução ou um aumento de um acontecimento de sinalização celular associado com a referida mutação no referido gene da doença, desenvolvendo assim o referido agente biologicamente ativo que modula o referido evento de sinalização celular associado com o referido gene da doença.

Num aspeto, a presente divulgação prevê um polinucleótido recombinante compreendendo uma sequência guia a montante de uma sequência de emparelhamento tracr, em que a sequência guia, quando expressa, orienta a ligação específica de sequências de um complexo CRISPR a uma sequência alvo correspondente presente numa célula eucariótica. Em algumas formas de realização, a sequência alvo é uma sequência viral presente numa célula eucariótica. Em algumas formas de realização, a sequência alvo é um proto-oncogene ou um oncogene.

Num aspeto, a presente divulgação prevê um método de seleção de uma ou mais células procarióticas através da introdução de uma ou mais mutações num gene numa ou mais células procarióticas, o método compreendendo: a introdução de um ou mais vetores na célula ou células procarióticas,

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
 755 760 765
 Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
 770 775 780
 Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
 785 790 795 800
 Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
 805 810 815
 Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
 820 825 830
 Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
 835 840 845
 Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
 850 855 860
 Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
 865 870 875 880
 Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
 885 890 895
 Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
 900 905 910
 Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
 915 920 925
 Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 930 935 940
 Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
 945 950 955 960
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
 965 970 975
 Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
 980 985 990
 Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
 995 1000 1005

Phe Gln	Phe Tyr Lys Val	Arg	Glu Ile Asn Asn Tyr	His His Ala
1010		1015		1020
His Asp	Ala Tyr Leu Asn Ala	Val Val Gly Thr Ala	Leu Ile Lys	
1025		1030		1035
Lys Tyr	Pro Lys Leu Glu Ser	Glu Phe Val Tyr Gly	Asp Tyr Lys	
1040		1045		1050
Val Tyr	Asp Val Arg Lys Met	Ile Ala Lys Ser Glu	Gln Glu Ile	
1055		1060		1065
Gly Lys	Ala Thr Ala Lys Tyr	Phe Phe Tyr Ser Asn	Ile Met Asn	
1070		1075		1080
Phe Phe	Lys Thr Glu Ile Thr	Leu Ala Asn Gly Glu	Ile Arg Lys	
1085		1090		1095
Arg Pro	Leu Ile Glu Thr Asn	Gly Glu Thr Gly Glu	Ile Val Trp	
1100		1105		1110
Asp Lys	Gly Arg Asp Phe Ala	Thr Val Arg Lys Val	Leu Ser Met	
1115		1120		1125
Pro Gln	Val Asn Ile Val Lys	Lys Thr Glu Val Gln	Thr Gly Gly	
1130		1135		1140
Phe Ser	Lys Glu Ser Ile Leu	Pro Lys Arg Asn Ser	Asp Lys Leu	
1145		1150		1155
Ile Ala	Arg Lys Lys Asp Trp	Asp Pro Lys Lys Tyr	Gly Gly Phe	
1160		1165		1170
Asp Ser	Pro Thr Val Ala Tyr	Ser Val Leu Val Val	Ala Lys Val	
1175		1180		1185
Glu Lys	Gly Lys Ser Lys Lys	Leu Lys Ser Val Lys	Glu Leu Leu	
1190		1195		1200
Gly Ile	Thr Ile Met Glu Arg	Ser Ser Phe Glu Lys	Asn Pro Ile	
1205		1210		1215
Asp Phe	Leu Glu Ala Lys Gly	Tyr Lys Glu Val Lys	Lys Asp Leu	
1220		1225		1230
Ile Ile	Lys Leu Pro Lys Tyr	Ser Leu Phe Glu Leu	Glu Asn Gly	
1235		1240		1245

Arg Lys	Arg Met	Leu Ala	Ser	Ala Gly	Glu Leu	Gln	Lys Gly	Asn	
1250			1255			1260			
Glu Leu	Ala Leu	Pro Ser	Lys	Tyr Val	Asn Phe	Leu	Tyr Leu	Ala	
1265			1270			1275			
Ser His	Tyr Glu	Lys Leu	Lys	Gly Ser	Pro Glu	Asp	Asn Glu	Gln	
1280			1285			1290			
Lys Gln	Leu Phe	Val Glu	Gln	His Lys	His Tyr	Leu	Asp Glu	Ile	
1295			1300			1305			
Ile Glu	Gln Ile	Ser Glu	Phe	Ser Lys	Arg Val	Ile	Leu Ala	Asp	
1310			1315			1320			
Ala Asn	Leu Asp	Lys Val	Leu	Ser Ala	Tyr Asn	Lys	His Arg	Asp	
1325			1330			1335			
Lys Pro	Ile Arg	Glu Gln	Ala	Glu Asn	Ile Ile	His	Leu Phe	Thr	
1340			1345			1350			
Leu Thr	Asn Leu	Gly Ala	Pro	Ala Ala	Phe Lys	Tyr	Phe Asp	Thr	
1355			1360			1365			
Thr Ile	Asp Arg	Lys Arg	Tyr	Thr Ser	Thr Lys	Glu	Val Leu	Asp	
1370			1375			1380			
Ala Thr	Leu Ile	His Gln	Ser	Ile Thr	Gly Leu	Tyr	Glu Thr	Arg	
1385			1390			1395			
Ile Asp	Leu Ser	Gln Leu	Gly	Gly Asp	Ala Ala	Ala	Val Ser	Lys	
1400			1405			1410			
Gly Glu	Glu Leu	Phe Thr	Gly	Val Val	Pro Ile	Leu	Val Glu	Leu	
1415			1420			1425			
Asp Gly	Asp Val	Asn Gly	His	Lys Phe	Ser Val	Ser	Gly Glu	Gly	
1430			1435			1440			
Glu Gly	Asp Ala	Thr Tyr	Gly	Lys Leu	Thr Leu	Lys	Phe Ile	Cys	
1445			1450			1455			
Thr Thr	Gly Lys	Leu Pro	Val	Pro Trp	Pro Thr	Leu	Val Thr	Thr	
1460			1465			1470			
Leu Thr	Tyr Gly	Val Gln	Cys	Phe Ser	Arg Tyr	Pro	Asp His	Met	

1475	1480	1485
Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val		
1490	1495	1500
Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr		
1505	1510	1515
Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile		
1520	1525	1530
Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly		
1535	1540	1545
His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met		
1550	1555	1560
Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg		
1565	1570	1575
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln		
1580	1585	1590
Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn		
1595	1600	1605
His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu		
1610	1615	1620
Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly		
1625	1630	1635
Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys		
1640	1645	

<210> 44
 <211> 1625
 <212> PRT
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polipeptido Sintético"

<400> 44
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800
 Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 820 825 830
 Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020
 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025		1030		1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala				
1040		1045		1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu				
1055		1060		1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val				
1070		1075		1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr				
1085		1090		1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys				
1100		1105		1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro				
1115		1120		1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val				
1130		1135		1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys				
1145		1150		1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser				
1160		1165		1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys				
1175		1180		1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu				
1190		1195		1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly				
1205		1210		1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val				
1220		1225		1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser				
1235		1240		1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys				
1250		1255		1260

His Tyr 1265	Leu Asp Glu Ile Ile 1270	Glu Gln Ile Ser Glu 1275	Phe Ser Lys
Arg Val 1280	Ile Leu Ala Asp Ala 1285	Asn Leu Asp Lys Val 1290	Leu Ser Ala
Tyr Asn 1295	Lys His Arg Asp Lys 1300	Pro Ile Arg Glu Gln 1305	Ala Glu Asn
Ile Ile 1310	His Leu Phe Thr Leu 1315	Thr Asn Leu Gly Ala 1320	Pro Ala Ala
Phe Lys 1325	Tyr Phe Asp Thr Thr 1330	Ile Asp Arg Lys Arg 1335	Tyr Thr Ser
Thr Lys 1340	Glu Val Leu Asp Ala 1345	Thr Leu Ile His Gln 1350	Ser Ile Thr
Gly Leu 1355	Tyr Glu Thr Arg Ile 1360	Asp Leu Ser Gln Leu 1365	Gly Gly Asp
Ala Ala 1370	Ala Val Ser Lys Gly 1375	Glu Glu Leu Phe Thr 1380	Gly Val Val
Pro Ile 1385	Leu Val Glu Leu Asp 1390	Gly Asp Val Asn Gly 1395	His Lys Phe
Ser Val 1400	Ser Gly Glu Gly Glu 1405	Gly Asp Ala Thr Tyr 1410	Gly Lys Leu
Thr Leu 1415	Lys Phe Ile Cys Thr 1420	Thr Gly Lys Leu Pro 1425	Val Pro Trp
Pro Thr 1430	Leu Val Thr Thr Leu 1435	Thr Tyr Gly Val Gln 1440	Cys Phe Ser
Arg Tyr 1445	Pro Asp His Met Lys 1450	Gln His Asp Phe Phe 1455	Lys Ser Ala
Met Pro 1460	Glu Gly Tyr Val Gln 1465	Glu Arg Thr Ile Phe 1470	Phe Lys Asp
Asp Gly 1475	Asn Tyr Lys Thr Arg 1480	Ala Glu Val Lys Phe 1485	Glu Gly Asp
Thr Leu 1490	Val Asn Arg Ile Glu 1495	Leu Lys Gly Ile Asp 1500	Phe Lys Glu

em que um ou mais vetores dirigem a expressão de um ou mais dos seguintes: uma enzima CRISPR, uma sequência guia ligada a uma sequência de emparelhamento tracr, uma sequência tracr, e um molde de edição; em que o molde de edição compreende uma ou mais mutações que suprimem a clivagem enzimática da CRISPR; permitindo a recombinação homóloga do molde de edição com o polinucleótido alvo na célula ou células a selecionar; permitindo um complexo CRISPR se ligue a um polinucleótido alvo para efetuar a clivagem do polinucleótido alvo no interior do referido gene, em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que hibrida com a sequência alvo dentro do polinucleótido alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr, em que a ligação do complexo CRISPR ao polinucleótido alvo induz a morte celular, permitindo assim a seleção de uma ou mais células procarióticas em que foram introduzidas uma ou mais mutações. Numa forma de realização preferida, a enzima CRISPR é a Cas9. Num outro aspeto da presente divulgação, a célula a ser selecionada pode ser uma célula eucariótica. Alguns aspetos da divulgação permitem a seleção de células específicas sem necessidade de um marcador de seleção ou um processo de dois passos que podem incluir um sistema de contrasseleção.

Em alguns aspetos, a presente divulgação prevê uma composição de origem não-natural ou manipulada que compreende uma sequência de polinucleótidos de RNA quimérico (chiRNA) do sistema CRISPR-Cas, em que a sequência de polinucleótidos compreende: (a) uma sequência guia capaz de hibridar com uma sequência alvo numa célula eucariótica, (b) uma sequência de emparelhamento tracr, e (c) uma sequência tracr em que (a), (b) e (c) se

```

Asp Gly  Asn Ile Leu Gly His  Lys Leu Glu Tyr Asn  Tyr Asn Ser
1505                                1510                    1515

His Asn  Val Tyr Ile Met Ala  Asp Lys Gln Lys Asn  Gly Ile Lys
1520                                1525                    1530

Val Asn  Phe Lys Ile Arg His  Asn Ile Glu Asp Gly  Ser Val Gln
1535                                1540                    1545

Leu Ala  Asp His Tyr Gln Gln  Asn Thr Pro Ile Gly  Asp Gly Pro
1550                                1555                    1560

Val Leu  Leu Pro Asp Asn His  Tyr Leu Ser Thr Gln  Ser Ala Leu
1565                                1570                    1575

Ser Lys  Asp Pro Asn Glu Lys  Arg Asp His Met Val  Leu Leu Glu
1580                                1585                    1590

Phe Val  Thr Ala Ala Gly Ile  Thr Leu Gly Met Asp  Glu Leu Tyr
1595                                1600                    1605

Lys Lys  Arg Pro Ala Ala Thr  Lys Lys Ala Gly Gln  Ala Lys Lys
1610                                1615                    1620

Lys Lys
1625

<210> 45
<211> 1664
<212> PRT
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição da Sequência Artificial: Polipéptido
        sintético"

<400> 45
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1          5          10          15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
          20          25          30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
          35          40          45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
          50          55          60

```

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His
 65 70 75 80
 Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu
 85 90 95
 Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr
 100 105 110
 Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu
 115 120 125
 Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe
 130 135 140
 Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn
 145 150 155 160
 Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His
 165 170 175
 Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu
 180 185 190
 Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu
 195 200 205
 Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe
 210 215 220
 Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile
 225 230 235 240
 Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser
 245 250 255
 Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys
 260 265 270
 Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr
 275 280 285
 Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln
 290 295 300
 Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln
 305 310 315 320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
 325 330 335
 Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
 340 345 350
 Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
 355 360 365
 Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
 370 375 380
 Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
 385 390 395 400
 Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
 405 410 415
 Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
 420 425 430
 Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
 435 440 445
 Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg
 450 455 460
 Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
 465 470 475 480
 Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
 485 490 495
 Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
 500 505 510
 Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
 515 520 525
 Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
 530 535 540
 Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
 545 550 555 560
 Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro

565					570					575					
Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe
		580						585					590		
Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe
		595					600					605			
Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp
	610					615					620				
Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile
625						630					635				640
Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu
				645					650					655	
Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu
			660					665					670		
Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys
		675					680					685			
Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys
	690					695					700				
Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp
705						710					715				720
Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile
				725					730					735	
His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val
			740					745					750		
Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly
	755						760					765			
Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp
	770					775					780				
Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile
785						790					795				800
Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser
				805					810					815	

Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
 820 825 830
 Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
 835 840 845
 Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
 850 855 860
 Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
 865 870 875 880
 Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
 885 890 895
 Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
 900 905 910
 Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
 915 920 925
 Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 930 935 940
 Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
 945 950 955 960
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
 965 970 975
 Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
 980 985 990
 Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
 995 1000 1005
 Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
 1010 1015 1020
 His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
 1025 1030 1035
 Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
 1040 1045 1050
 Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
 1055 1060 1065

Gly Lys	Ala Thr	Ala Lys	Tyr	Phe Phe	Tyr Ser	Asn	Ile Met	Asn	
1070			1075			1080			
Phe Phe	Lys Thr	Glu Ile	Thr	Leu Ala	Asn Gly	Glu	Ile Arg	Lys	
1085			1090			1095			
Arg Pro	Leu Ile	Glu Thr	Asn	Gly Glu	Thr Gly	Glu	Ile Val	Trp	
1100			1105			1110			
Asp Lys	Gly Arg	Asp Phe	Ala	Thr Val	Arg Lys	Val	Leu Ser	Met	
1115			1120			1125			
Pro Gln	Val Asn	Ile Val	Lys	Lys Thr	Glu Val	Gln	Thr Gly	Gly	
1130			1135			1140			
Phe Ser	Lys Glu	Ser Ile	Leu	Pro Lys	Arg Asn	Ser	Asp Lys	Leu	
1145			1150			1155			
Ile Ala	Arg Lys	Lys Asp	Trp	Asp Pro	Lys Lys	Tyr	Gly Gly	Phe	
1160			1165			1170			
Asp Ser	Pro Thr	Val Ala	Tyr	Ser Val	Leu Val	Val	Ala Lys	Val	
1175			1180			1185			
Glu Lys	Gly Lys	Ser Lys	Lys	Leu Lys	Ser Val	Lys	Glu Leu	Leu	
1190			1195			1200			
Gly Ile	Thr Ile	Met Glu	Arg	Ser Ser	Phe Glu	Lys	Asn Pro	Ile	
1205			1210			1215			
Asp Phe	Leu Glu	Ala Lys	Gly	Tyr Lys	Glu Val	Lys	Lys Asp	Leu	
1220			1225			1230			
Ile Ile	Lys Leu	Pro Lys	Tyr	Ser Leu	Phe Glu	Leu	Glu Asn	Gly	
1235			1240			1245			
Arg Lys	Arg Met	Leu Ala	Ser	Ala Gly	Glu Leu	Gln	Lys Gly	Asn	
1250			1255			1260			
Glu Leu	Ala Leu	Pro Ser	Lys	Tyr Val	Asn Phe	Leu	Tyr Leu	Ala	
1265			1270			1275			
Ser His	Tyr Glu	Lys Leu	Lys	Gly Ser	Pro Glu	Asp	Asn Glu	Gln	
1280			1285			1290			
Lys Gln	Leu Phe	Val Glu	Gln	His Lys	His Tyr	Leu	Asp Glu	Ile	
1295			1300			1305			


```

1535              1540              1545

His Lys  Leu Glu Tyr Asn Tyr  Asn Ser His Asn Val  Tyr Ile Met
1550              1555              1560

Ala Asp  Lys Gln Lys Asn Gly  Ile Lys Val Asn Phe  Lys Ile Arg
1565              1570              1575

His Asn  Ile Glu Asp Gly Ser  Val Gln Leu Ala Asp  His Tyr Gln
1580              1585              1590

Gln Asn  Thr Pro Ile Gly Asp  Gly Pro Val Leu Leu  Pro Asp Asn
1595              1600              1605

His Tyr  Leu Ser Thr Gln Ser  Ala Leu Ser Lys Asp  Pro Asn Glu
1610              1615              1620

Lys Arg  Asp His Met Val Leu  Leu Glu Phe Val Thr  Ala Ala Gly
1625              1630              1635

Ile Thr  Leu Gly Met Asp Glu  Leu Tyr Lys Lys Arg  Pro Ala Ala
1640              1645              1650

Thr Lys  Lys Ala Gly Gln Ala  Lys Lys Lys Lys
1655              1660

<210> 46
<211> 1423
<212> PRT
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polipeptido
        sintético"

<400> 46
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1          5          10          15

Tyr Lys Asp Asp Asp Asp Lys Met  Ala Pro Lys Lys Lys Arg Lys Val
          20          25          30

Gly Ile His Gly Val Pro Ala Ala Asp Lys Lys Tyr Ser Ile Gly Leu
          35          40          45

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr
50          55          60

Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His

```

65					70					75				80	
Ser	Ile	Lys	Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu
				85					90					95	
Thr	Ala	Glu	Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr
			100					105						110	
Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu
		115					120					125			
Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe
	130					135					140				
Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn
145					150					155					160
Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His
				165					170					175	
Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu
			180					185					190		
Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu
		195					200					205			
Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe
	210					215					220				
Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile
225					230					235					240
Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser
				245					250					255	
Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys
			260				265						270		
Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr
		275					280					285			
Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln
	290					295					300				
Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln
305					310				315						320

Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser
 325 330 335
 Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr
 340 345 350
 Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His
 355 360 365
 Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu
 370 375 380
 Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly
 385 390 395 400
 Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys
 405 410 415
 Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu
 420 425 430
 Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser
 435 440 445
 Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg
 450 455 460
 Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu
 465 470 475 480
 Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg
 485 490 495
 Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile
 500 505 510
 Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
 515 520 525
 Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
 530 535 540
 Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
 545 550 555 560
 Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
 565 570 575

encontrarem dispostos numa orientação 5' para 3', em que, quando transcrita, a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia orienta a ligação específica de sequências de um complexo CRISPR para a sequência alvo, em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com (1) a sequência guia que hibrida com a sequência alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr,

ou

um sistema enzimático CRISPR, em que o sistema é codificado por um sistema de vetor compreendendo um ou mais vetores compreendendo I. um primeiro elemento regulador ligado operacionalmente a uma sequência de polinucleótidos de RNA quimérico (chiRNA) do sistema CRISPR-Cas, em que a sequência de polinucleótidos compreende (a) uma ou mais sequências guia capazes de hibridar com uma ou mais sequências alvo numa célula eucariótica, (b) uma sequência de emparelhamento tracr, e (c) uma ou mais sequências tracr, e II. um segundo elemento regulador operacionalmente ligado a uma sequência que codifica enzimas que codificam uma enzima CRISPR que compreende, pelo menos, uma ou mais sequências de localização nuclear, em que (a), (b) e (c) se encontrarem dispostos numa orientação 5' para 3', em que os componentes I e II estão localizados no mesmo ou em diferentes vetores do sistema, em que quando transcrita, a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia orienta a ligação específica da sequência de um complexo CRISPR para a sequência alvo, em que o complexo CRISPR compreende a enzima CRISPR num complexo com (1) a sequência guia que é hibridada com a sequência alvo, e (2) a sequência de emparelhamento tracr que é hibridada com a sequência tracr, ou um sistema enzimático CRISPR multiplex, em que o sistema é codificado

Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
 580 585 590
 Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe
 595 600 605
 Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp
 610 615 620
 Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
 625 630 635 640
 Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
 645 650 655
 Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
 660 665 670
 Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
 675 680 685
 Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
 690 695 700
 Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
 705 710 715 720
 Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
 725 730 735
 His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
 740 745 750
 Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
 755 760 765
 Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
 770 775 780
 Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
 785 790 795 800
 Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
 805 810 815
 Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
 820 825 830

Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
 835 840 845
 Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
 850 855 860
 Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
 865 870 875 880
 Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
 885 890 895
 Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
 900 905 910
 Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
 915 920 925
 Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 930 935 940
 Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
 945 950 955 960
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
 965 970 975
 Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
 980 985 990
 Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
 995 1000 1005
 Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
 1010 1015 1020
 His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys
 1025 1030 1035
 Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys
 1040 1045 1050
 Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
 1055 1060 1065
 Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn

1070	1075	1080
Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys 1085 1090 1095		
Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp 1100 1105 1110		
Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met 1115 1120 1125		
Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly 1130 1135 1140		
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu 1145 1150 1155		
Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe 1160 1165 1170		
Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val 1175 1180 1185		
Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu 1190 1195 1200		
Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile 1205 1210 1215		
Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu 1220 1225 1230		
Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly 1235 1240 1245		
Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn 1250 1255 1260		
Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala 1265 1270 1275		
Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln 1280 1285 1290		
Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile 1295 1300 1305		

Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp
1310 1315 1320

Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
1325 1330 1335

Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr
1340 1345 1350

Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1355 1360 1365

Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1370 1375 1380

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg
1385 1390 1395

Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr
1400 1405 1410

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1415 1420

<210> 47

<211> 483

<212> PRT

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Polipéptido sintético"

<400> 47

Met Phe Leu Phe Leu Ser Leu Thr Ser Phe Leu Ser Ser Ser Arg Thr
1 5 10 15

Leu Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe
20 25 30

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
35 40 45

Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr
50 55 60

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
65 70 75 80

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
 85 90 95

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
 100 105 110

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
 115 120 125

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
 130 135 140

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
 145 150 155 160

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
 165 170 175

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
 180 185 190

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
 195 200 205

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
 210 215 220

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
 225 230 235 240

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Ser Lys Gln
 245 250 255

Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe Asn Asp Leu
 260 265 270

Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala Asn Glu His
 275 280 285

Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe Leu Gly Asp
 290 295 300

Ala Val Leu Gln Leu Ile Ile Ser Glu Tyr Leu Phe Ala Lys Tyr Pro
 305 310 315 320

Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met Ile Val Arg
 325 330 335

Glu Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe Asp Ala Tyr
 340 345 350

Ile Lys Leu Gly Lys Gly Glu Glu Lys Ser Gly Gly Arg Arg Arg Asp
 355 360 365

Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala Leu Leu Leu
 370 375 380

Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln Val Met Ile
 385 390 395 400

Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Asp Tyr Lys Thr
 405 410 415

Cys Leu Gln Glu Phe Leu Gln Thr Lys Gly Asp Val Ala Ile Asp Tyr
 420 425 430

Gln Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Gln Phe Glu Val
 435 440 445

Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu Gly Lys Ser
 450 455 460

Lys Lys Leu Ala Glu Gln Asp Ala Ala Lys Asn Ala Leu Ala Gln Leu
 465 470 475 480

Ser Glu Val

<210> 48
 <211> 483
 <212> PRT
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polipéptido
 sintético"

<400> 48
 Met Lys Gln Leu Glu Glu Leu Leu Ser Thr Ser Phe Asp Ile Gln Phe
 1 5 10 15

Asn Asp Leu Thr Leu Leu Glu Thr Ala Phe Thr His Thr Ser Tyr Ala
 20 25 30

Asn Glu His Arg Leu Leu Asn Val Ser His Asn Glu Arg Leu Glu Phe
 35 40 45

Leu Gly Asp Ala Val Leu Gln Leu Ile Ile Ser Glu Tyr Leu Phe Ala
 50 55 60

Lys Tyr Pro Lys Lys Thr Glu Gly Asp Met Ser Lys Leu Arg Ser Met
 65 70 75 80

Ile Val Arg Glu Glu Ser Leu Ala Gly Phe Ser Arg Phe Cys Ser Phe
 85 90 95

Asp Ala Tyr Ile Lys Leu Gly Lys Gly Glu Glu Lys Ser Gly Gly Arg
 100 105 110

Arg Arg Asp Thr Ile Leu Gly Asp Leu Phe Glu Ala Phe Leu Gly Ala
 115 120 125

Leu Leu Leu Asp Lys Gly Ile Asp Ala Val Arg Arg Phe Leu Lys Gln
 130 135 140

Val Met Ile Pro Gln Val Glu Lys Gly Asn Phe Glu Arg Val Lys Asp
 145 150 155 160

Tyr Lys Thr Cys Leu Gln Glu Phe Leu Gln Thr Lys Gly Asp Val Ala
 165 170 175

Ile Asp Tyr Gln Val Ile Ser Glu Lys Gly Pro Ala His Ala Lys Gln
 180 185 190

Phe Glu Val Ser Ile Val Val Asn Gly Ala Val Leu Ser Lys Gly Leu
 195 200 205

Gly Lys Ser Lys Lys Leu Ala Glu Gln Asp Ala Ala Lys Asn Ala Leu
 210 215 220

Ala Gln Leu Ser Glu Val Gly Ser Val Ser Lys Gly Glu Glu Asp Asn
 225 230 235 240

Met Ala Ile Ile Lys Glu Phe Met Arg Phe Lys Val His Met Glu Gly
 245 250 255

Ser Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Arg
 260 265 270

Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys Val Thr Lys Gly Gly
 275 280 285

Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Gln Phe Met Tyr Gly
 290 295 300

Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile Pro Asp Tyr Leu Lys
305 310 315 320

Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val Met Asn Phe Glu
325 330 335

Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser Leu Gln Asp Gly
340 345 350

Glu Phe Ile Tyr Lys Val Lys Leu Arg Gly Thr Asn Phe Pro Ser Asp
355 360 365

Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu Ala Ser Ser Glu
370 375 380

Arg Met Tyr Pro Glu Asp Gly Ala Leu Lys Gly Glu Ile Lys Gln Arg
385 390 395 400

Leu Lys Leu Lys Asp Gly Gly His Tyr Asp Ala Glu Val Lys Thr Thr
405 410 415

Tyr Lys Ala Lys Lys Pro Val Gln Leu Pro Gly Ala Tyr Asn Val Asn
420 425 430

Ile Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile Val Glu
435 440 445

Gln Tyr Glu Arg Ala Glu Gly Arg His Ser Thr Gly Gly Met Asp Glu
450 455 460

Leu Tyr Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
465 470 475 480

Lys Lys Lys

<210> 49

<211> 1423

<212> PRT

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Polipéptido sintético"

<400> 49

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

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

260					265					270					
Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr
		275					280					285			
Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln
	290					295					300				
Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln
305					310					315					320
Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser
				325					330					335	
Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr
			340					345					350		
Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His
	355					360						365			
Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu
	370					375					380				
Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly
385					390					395					400
Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys
				405					410					415	
Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu
		420						425					430		
Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser
		435					440					445			
Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg
	450					455					460				
Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu
465					470					475					480
Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg
				485					490					495	
Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile
			500					505					510		

por um sistema de vetor compreendendo um ou mais vetores compreendendo I. um primeiro elemento regulador ligado operacionalmente a (a) uma ou mais sequências guia capazes de hibridar com uma sequência alvo numa célula, e (b) pelo menos uma ou mais sequências de emparelhamento tracr, II. um segundo elemento regulador ligado operacionalmente a uma sequência que codifica enzimas que codificam uma enzima CRISPR, e III. um terceiro elemento regulador ligado operacionalmente a uma sequência tracr, em que os componentes I, II e III estão localizados no mesmo ou em diferentes vetores do sistema, em que quando transcrita, a sequência de emparelhamento tracr hibrida com a sequência tracr e a sequência guia dirige a ligação específica da sequência de um complexo CRISPR para a sequência alvo, em que o complexo CRISPR compreende a enzima CRISPR que forma um complexo com (1) a sequência guia que é hibridada com a sequência alvo, e (2) a sequência de emparelhamento tracr que é hibridada com a sequência tracr, e em que são utilizadas as sequências guia do sistema multiplex e uma única sequência tracr; e em que uma ou mais das sequências guia, tracr e de emparelhamento tracr são modificadas para melhorar a estabilidade.

De acordo com aspetos da divulgação, a modificação compreende uma estrutura secundária obtida por engenharia. Por exemplo, a modificação pode compreender uma redução da região de hibridação entre a sequência de emparelhamento tracr e a sequência tracr. Por exemplo, a modificação também pode compreender a fusão da sequência de emparelhamento tracr e a sequência tracr através de uma alça artificial. A modificação pode compreender a sequência tracr tendo um comprimento entre 40 e 120 pb. Em formas de realização da presente divulgação, a sequência tracr é

Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln
 515 520 525

Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu
 530 535 540

Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr
 545 550 555 560

Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro
 565 570 575

Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe
 580 585 590

Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe
 595 600 605

Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp
 610 615 620

Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile
 625 630 635 640

Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu
 645 650 655

Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu
 660 665 670

Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys
 675 680 685

Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys
 690 695 700

Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp
 705 710 715 720

Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile
 725 730 735

His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val
 740 745 750

Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly
 755 760 765

Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp
 770 775 780
 Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile
 785 790 795 800
 Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
 805 810 815
 Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser
 820 825 830
 Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
 835 840 845
 Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
 850 855 860
 Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
 865 870 875 880
 Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
 885 890 895
 Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
 900 905 910
 Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala
 915 920 925
 Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 930 935 940
 Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
 945 950 955 960
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
 965 970 975
 Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val
 980 985 990
 Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp
 995 1000 1005
 Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala
 1010 1015 1020

His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Val	Gly	Thr	Ala	Leu	Ile	Lys
1025						1030					1035			
Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val	Tyr	Gly	Asp	Tyr	Lys
1040						1045					1050			
Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	Lys	Ser	Glu	Gln	Glu	Ile
1055						1060					1065			
Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	Tyr	Ser	Asn	Ile	Met	Asn
1070						1075					1080			
Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	Asn	Gly	Glu	Ile	Arg	Lys
1085						1090					1095			
Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	Thr	Gly	Glu	Ile	Val	Trp
1100						1105					1110			
Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val	Arg	Lys	Val	Leu	Ser	Met
1115						1120					1125			
Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr	Glu	Val	Gln	Thr	Gly	Gly
1130						1135					1140			
Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Arg	Asn	Ser	Asp	Lys	Leu
1145						1150					1155			
Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Phe
1160						1165					1170			
Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val	Leu	Val	Val	Ala	Lys	Val
1175						1180					1185			
Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Ser	Val	Lys	Glu	Leu	Leu
1190						1195					1200			
Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser	Phe	Glu	Lys	Asn	Pro	Ile
1205						1210					1215			
Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu	Val	Lys	Lys	Asp	Leu
1220						1225					1230			
Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Glu	Leu	Glu	Asn	Gly
1235						1240					1245			
Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu	Leu	Gln	Lys	Gly	Asn

1250		1255		1260
Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala				
1265		1270		1275
Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln				
1280		1285		1290
Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile				
1295		1300		1305
Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp				
1310		1315		1320
Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp				
1325		1330		1335
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr				
1340		1345		1350
Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr				
1355		1360		1365
Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp				
1370		1375		1380
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg				
1385		1390		1395
Ile Asp Leu Ser Gln Leu Gly Gly Asp Lys Arg Pro Ala Ala Thr				
1400		1405		1410
Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys				
1415		1420		

<210> 50
 <211> 2012
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 50	
gaatgctgcc ctcagaccgc ctccctccct gtccttgtct gtccaaggag aatgaggtct	60
cactggtgga ttccggacta ccctgaggag ctggcacctg agggacaagg cccccacct	120
gccagctcc agcctctgat gaggggtggg agagagctac atgaggttgc taagaaagcc	180

tccccgaag gagaccacac agtgtgtgag gttggagtct ctagcagcgg gttctgtgcc	240
cccagggata gtctggctgt ccaggcactg ctcttgatat aaacaccacc tcctagttat	300
gaaaccatgc ccattctgcc tctctgtatg gaaaagagca tggggctggc ccgtggggtg	360
gtgtccactt taggcctgt gggagatcat gggaaaccac gcagtgggtc ataggctctc	420
tcatttacta ctcacatcca ctctgtgaag aagcgattat gatctctcct ctagaaactc	480
gtagagtccc atgtctgccg gcttcagag cctgcactcc tccaccttg cttggctttg	540
ctggggctag aggagctagg atgcacagca gctctgtgac cctttgtttg agaggaacag	600
gaaaaccacc cttctctctg gccactgtg tccctctcct gccctgccat ccccttctgt	660
gaatgttaga cccatgggag cagctggcca gaggggaccc cggcctgggg cccctaaccc	720
tatgtagcct cagtcttccc atcaggctct cagctcagcc tgagtgttga ggccccagt	780
gctgctctgg gggcctcctg agtttctcat ctgtgcccct ccctccctgg ccaggtgaa	840
ggtgtggttc cagaaccgga ggacaaagta caaacggcag aagctggagg aggaagggcc	900
tgagtccgag cagaagaaga agggctccca tcacatcaac cgggtggcga ttgccacgaa	960
gcaggccaat ggggaggaca togatgtcac ctccaatgac aagcttgcta gcggtgggca	1020
accacaaacc cagaggggca gagtgtctgt tgctgtctggc caggccccctg cgtgggcccc	1080
agctggactc tggccactcc ctggccaggc tttggggagg cctggagtca tggccccaca	1140
gggcttgaag cccggggccg ccattgacag agggacaagc aatgggctgg ctgaggcctg	1200
ggaccacttg gccttctcct cggagagcct gcctgcctgg gcggggccgc ccgccaccgc	1260
agcctcccag ctgctctccg tgtctccaat ctcccttttg ttttgatgca tttctgtttt	1320
aatttatttt ccaggcacca ctgtagttta gtgatcccca gtgtccccct tccctatggg	1380
aataataaaa gtctctctct taatgacac ggcatccagc tccagcccca gagcctgggg	1440
tggtagattc cggctctgag ggccagtggg ggctggtaga gcaaacgcgt tcagggcctg	1500
ggagcctggg gtggggact ggtggagggg gtcaagggta attcattaac tcctctcttt	1560
tgttggggga ccctggtctc tacctccagc tccacagcag gagaaacagg ctagacatag	1620
ggaagggcca tcctgtatct tgagggagga caggcccagg tctttcttaa cgtattgaga	1680
ggtgggaatc aggccaggt agttcaatgg gagagggaga gtgcttccct ctgcctagag	1740
actctggtgg cttctccagt tgaggagaaa ccagaggaaa ggggaggatt ggggtctggg	1800
ggagggaaaca ccattcacia aggtgacgg ttccagtcog aagtcgtggg cccaccagga	1860
tgctcacctg tccttgaga acgctgggc aggttgagac tgcagagaca gggcttaagg	1920
ctgagcctgc aaccagtccc cagtgaactca gggcctcctc agcccaagaa agagcaacgt	1980
gccagggccc gctgagctct tgtgttcacc tg	2012

```

<210> 51
<211> 1153
<212> PRT
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polipéptido
        sintético"

<400> 51
Met Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1          5          10          15

Lys Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
          20          25          30

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
          35          40          45

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
50          55          60

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
65          70          75          80

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
          85          90          95

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
          100          105          110

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
          115          120          125

Val Lys His Arg Gly Ile Ser Tyr Leu Asp Asp Ala Ser Asp Asp Gly
130          135          140

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
145          150          155          160

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
          165          170          175

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
          180          185          190

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
          195          200          205

```

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
 210 215 220

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
 225 230 235 240

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
 245 250 255

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
 260 265 270

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
 275 280 285

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
 290 295 300

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
 305 310 315 320

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
 325 330 335

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
 340 345 350

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
 355 360 365

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
 370 375 380

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
 385 390 395 400

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
 405 410 415

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
 420 425 430

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
 435 440 445

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
 450 455 460

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
 465 470 475 480

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
 485 490 495

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
 500 505 510

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
 515 520 525

Glu Thr Asn Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys
 530 535 540

Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln
 545 550 555 560

Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys
 565 570 575

Gln Leu Ala Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys
 580 585 590

Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser
 595 600 605

Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp
 610 615 620

Asp Ser Leu Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu
 625 630 635 640

Lys Gly Gln Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala
 645 650 655

Trp Ser Phe Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu
 660 665 670

Ser Asn Lys Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys
 675 680 685

Phe Asp Val Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg
 690 695 700

Tyr Ala Ser Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala

705					710					715				720	
His	Lys	Ile	Asp	Thr	Lys	Val	Ser	Val	Val	Arg	Gly	Gln	Phe	Thr	Ser
				725					730					735	
Gln	Leu	Arg	Arg	His	Trp	Gly	Ile	Glu	Lys	Thr	Arg	Asp	Thr	Tyr	His
			740					745					750		
His	His	Ala	Val	Asp	Ala	Leu	Ile	Ile	Ala	Ala	Ser	Ser	Gln	Leu	Asn
		755					760					765			
Leu	Trp	Lys	Lys	Gln	Lys	Asn	Thr	Leu	Val	Ser	Tyr	Ser	Glu	Asp	Gln
	770					775					780				
Leu	Leu	Asp	Ile	Glu	Thr	Gly	Glu	Leu	Ile	Ser	Asp	Asp	Glu	Tyr	Lys
785					790					795					800
Glu	Ser	Val	Phe	Lys	Ala	Pro	Tyr	Gln	His	Phe	Val	Asp	Thr	Leu	Lys
				805					810					815	
Ser	Lys	Glu	Phe	Glu	Asp	Ser	Ile	Leu	Phe	Ser	Tyr	Gln	Val	Asp	Ser
			820					825					830		
Lys	Phe	Asn	Arg	Lys	Ile	Ser	Asp	Ala	Thr	Ile	Tyr	Ala	Thr	Arg	Gln
		835					840					845			
Ala	Lys	Val	Gly	Lys	Asp	Lys	Ala	Asp	Glu	Thr	Tyr	Val	Leu	Gly	Lys
	850					855					860				
Ile	Lys	Asp	Ile	Tyr	Thr	Gln	Asp	Gly	Tyr	Asp	Ala	Phe	Met	Lys	Ile
865					870					875					880
Tyr	Lys	Lys	Asp	Lys	Ser	Lys	Phe	Leu	Met	Tyr	Arg	His	Asp	Pro	Gln
				885					890					895	
Thr	Phe	Glu	Lys	Val	Ile	Glu	Pro	Ile	Leu	Glu	Asn	Tyr	Pro	Asn	Lys
			900					905					910		
Gln	Ile	Asn	Glu	Lys	Gly	Lys	Glu	Val	Pro	Cys	Asn	Pro	Phe	Leu	Lys
		915					920					925			
Tyr	Lys	Glu	Glu	His	Gly	Tyr	Ile	Arg	Lys	Tyr	Ser	Lys	Lys	Gly	Asn
	930					935					940				
Gly	Pro	Glu	Ile	Lys	Ser	Leu	Lys	Tyr	Tyr	Asp	Ser	Lys	Leu	Gly	Asn
945					950					955					960

```

His Ile Asp Ile Thr Pro Lys Asp Ser Asn Asn Lys Val Val Leu Gln
      965                      970                      975

Ser Val Ser Pro Trp Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly
      980                      985                      990

Lys Tyr Glu Ile Leu Gly Leu Lys Tyr Ala Asp Leu Gln Phe Glu Lys
      995                      1000                      1005

Gly Thr Gly Thr Tyr Lys Ile Ser Gln Glu Lys Tyr Asn Asp Ile
      1010                      1015                      1020

Lys Lys Lys Glu Gly Val Asp Ser Asp Ser Glu Phe Lys Phe Thr
      1025                      1030                      1035

Leu Tyr Lys Asn Asp Leu Leu Leu Val Lys Asp Thr Glu Thr Lys
      1040                      1045                      1050

Glu Gln Gln Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln
      1055                      1060                      1065

Lys His Tyr Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu
      1070                      1075                      1080

Gly Gly Glu Ala Leu Ile Lys Val Leu Gly Asn Val Ala Asn Ser
      1085                      1090                      1095

Gly Gln Cys Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr
      1100                      1105                      1110

Lys Val Arg Thr Asp Val Leu Gly Asn Gln His Ile Ile Lys Asn
      1115                      1120                      1125

Glu Gly Asp Lys Pro Lys Leu Asp Phe Lys Arg Pro Ala Ala Thr
      1130                      1135                      1140

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
      1145                      1150

```

```

<210> 52
<211> 340
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<400> 52

```

entre 40 pb e o comprimento completo do *tracr*. Em certas formas de realização, o comprimento de *tracr*RNA inclui pelo menos os nucleótidos 1-67 e em algumas formas de realização, pelo menos, os nucleótidos 1-85 do *tracr*RNA do tipo selvagem. Em algumas formas de realização, podem ser usados pelo menos os nucleótidos correspondentes aos nucleótidos 1-67 ou 1-85 de *tracr*RNA de Cas9 de *S. pyogenes* do tipo selvagem. Sempre que o sistema CRISPR utiliza outras enzimas que não a Cas9, ou outras que não a SpCas9, em seguida, então podem estar presentes os nucleótidos correspondentes no *tracr*RNA relevante do tipo selvagem. Em algumas formas de realização, o comprimento de *tracr*RNA inclui não mais do que 1-67 ou 1-85 nucleótidos de *tracr*RNA do tipo selvagem. A modificação pode compreender otimização de sequências. Em certos aspectos, a otimização de sequências pode compreender a redução da incidência de sequências poli-T na sequência *tracr* e/ou de emparelhamento *tracr*. A otimização de sequências pode ser combinada com a redução da região de hibridação entre a sequência de emparelhamento *tracr* e a sequência *tracr*; por exemplo, uma sequência *tracr* de comprimento reduzido.

Num aspecto, a divulgação prevê o sistema CRISPR-Cas ou o sistema enzimático CRISPR em que a modificação compreende a redução nas sequências de poli-T nas sequências *tracr* e/ou de emparelhamento *tracr*. Em alguns aspectos da presente divulgação, estão presentes um ou mais T numa sequência poli-T da sequência de tipo selvagem relevante (ou seja, um fragmento de mais de 3, 4, 5, 6, ou mais bases T contíguas; em algumas formas de realização, um fragmento de não mais do que 10, 9, 8, 7, 6 bases T contíguas) pode ser substituído por um nucleótido que não o T, por exemplo, um A, de modo que a cadeia é dividida em fragmentos menores de

```

gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgaogtaga      120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180
atgcttaccg taacttgaaa gtatttcgat ttcttggcct tatatatctt gtggaaagga      240
cgaaacaccg ttacttaaat ctgcagaag ctacaaagat aaggcttcat gccgaaatca      300
acacctgtc attttatggc aggggtgttt cgttatttaa      340

```

```

<210> 53
<211> 360
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (288)..(317)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 53
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgaogtaga      120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180
atgcttaccg taacttgaaa gtatttcgat ttcttggcct tatatatctt gtggaaagga      240
cgaaacaccg ggttttagag ctatgctgtt ttgaatggc ccaaaacnnn nnnnnnnnnn      300
nnnnnnnnnn nnnnnnngtt ttagagctat gctgttttga atgggcccaa aacttttttt      360

```

```

<210> 54
<211> 318
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (250)..(269)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 54
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgaogtaga      120

```

```

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180
atgcttacog taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga      240
cgaaacacccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaata      300
aggctagtcc gttttttt      318

```

```

<210> 55
<211> 325
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (250)..(269)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 55
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga      120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180
atgcttacog taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga      240
cgaaacacccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaata      300
aggctagtcc gttatcattt ttttt      325

```

```

<210> 56
<211> 337
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (250)..(269)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 56
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga      120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180

```

```

atgcttaccg taacttgaaa gtatttogat ttcttggctt tatatatctt gtggaaagga      240
cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata      300
aggctagtcc gttatcaact tgaaaaagtg tttttttt                               337

```

```

<210> 57
<211> 352
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (250)..(269)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 57
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag      60
ataattggaa ttaatttgac tgtaaacaca aagatattag taaaaatac gtgacgtaga      120
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat      180
atgcttaccg taacttgaaa gtatttogat ttcttggctt tatatatctt gtggaaagga      240
cgaaacaccn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca agttaaaata      300
aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgcttttt tt              352

```

```

<210> 58
<211> 5101
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<400> 58
cgttacataa cttacggtaa atggcccgcc tggotgaccg cccaacgacc cccgccatt      60
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca      120
atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc      180
aagtacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta      240
catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac      300
catggtcgag gtgagcccca cgttctgott cactctcccc atctccccc cctcccacc      360
cccaattttg tatttattta tttttaatt attttgtca gcgatggggg cggggggggg      420
ggggggggcg gcgccaggcg gggcggggcg gggcgagggg cggggcgggg cgaggcggag      480

```

aggtgctggcg gcagccaatc agagcggcgc gctccgaaag tttcctttta tggcgaggcg	540
gcggcggcgg cgccctata aaaagcgaag cgcgcggcgg gcgggagtcg ctgcgacgct	600
gccttcgccc cgtgccccgc tccgcgcgcg cctcgcgcgc cccgccccgg ctctgactga	660
cgcggttact cccacaggtg agcggggcgg acggcccttc tcctccgggc tgtaattagc	720
tgagcaagag gtaagggttt aagggatggt tggttggtgg ggtattaatg ttttaattacc	780
tggagcacct gcctgaaatc actttttttc aggttggacc ggtgccacca tggactataa	840
ggaccacgac ggagactaca aggatcatga tattgattac aaagacgatg acgataagat	900
ggccccaag aagaagcga aggtcgggat ccacggagtc ccagcagccg acaagaagta	960
cagcatcggc ctggacatcg gcaccaactc tgtgggctgg gccgtgatca ccgacgagta	1020
caagggtgcc agcaagaaat tcaagggtgct gggcaacacc gaccggcaca gcatcaagaa	1080
gaacctgatc ggagccctgc tgttcgacag cggcgaaaca gccgaggcca cccggctgaa	1140
gagaaccgcc agaagaagat acaccagacg gaagaaccgg atctgctatc tgcaagagat	1200
cttcagcaac gagatggcca aggtggacga cagcttcttc cacagactgg aagagtcctt	1260
cctggtggaa gaggataaga agcagcagcg gcaccccatc ttccggcaaca tcgtggacga	1320
ggtggcctac caccgagaagt accccaccat ctaccacctg agaaagaaac tgggtgacag	1380
caccgacaag gccgacctgc ggctgatcta tctggccctg gccacatga tcaagttccg	1440
gggccacttc ctgatcgagg gcgacctgaa ccccgacaac agcgacgtgg acaagctgtt	1500
catccagctg gtgcagacct acaaccagct gttcggaggaa aaccccatca accccagcgg	1560
cgtggacgcc aaggccatcc tgtctgccag actgagcaag agcagacggc tggaaaatct	1620
gatcgcccag ctgcccggcg agaagaagaa tggcctgttc ggcaacctga ttgccctgag	1680
cctgggcctg acccccaact tcaagagcaa cttcgacctg gccgaggatg ccaaactgca	1740
gctgagcaag gacacctacg acgacgacct ggacaacctg ctggcccaga tgggcgacca	1800
gtacgcgcac ctgtttctgg ccgccaagaa cctgtccgac gccatcctgc tgagcgacat	1860
cctgagagtg aacaccgaga tcaccaaggc cccctgagc gcctctatga tcaagagata	1920
cgacgagcac caccaggacc tgacctgct gaaagctctc gtgcggcagc agctgcctga	1980
gaagtacaaa gagattttct tcgaccagag caagaacggc tacgcggct acattgacgg	2040
cggagccagc caggaagagt tctacaagtt catcaagccc atcctggaaa agatggacgg	2100
caccgaggaa ctgctcgtga agctgaacag agaggacctg ctgcggaagc agcggacctt	2160
cgacaacggc agcatcccc accagatcca cctgggagag ctgcacgcca ttctgcggcg	2220
gcaggaagat ttttaccat tcctgaagga caaccgggaa aagatcgaga agatcctgac	2280
cttcggcatc ccctactacg tgggcccctc ggccagggga aacagcagat tcgcctggat	2340

gaccagaaag agcgaggaaa ccatcaccoc ctggaacttc gaggaagtgg tggacaaggg	2400
cgcttcggcc cagagcttca tcgagoggat gaccaacttc gataagaacc tgcccaacga	2460
gaaggtgctg cccaagcaca gcctgctgta cgagtacttc accgtgtata acgagctgac	2520
caaagtgaaa tacgtgaccg agggaatgag aaagcccgcc ttcctgagcg gcgagcagaa	2580
aaaggccatc gtggacctgc tgttcaagac caaccggaaa gtgacctga agcagctgaa	2640
agaggactac ttcaagaaaa tcgagtgtct cgactccgtg gaaatctccg gcgtggaaga	2700
tcggttcaac gcctccctgg gcacatacca cgatctgctg aaaattatca aggacaagga	2760
cttccctggac aatgaggaaa acgaggacat tctggaagat atcgtgctga ccctgacact	2820
gtttgaggac agagagatga tcgaggaacg gctgaaaacc tatgccacc tgttcgacga	2880
caaagtgatg aagcagctga agcggcggag atacaccggc tggggcaggc tgagccggaa	2940
gctgatcaac ggcacccggg acaagcagtc cggcaagaca atcctggatt tcctgaagtc	3000
cgacggcttc gccaacagaa acttcatgca gctgatccac gacgacagcc tgacctttaa	3060
agaggacatc cagaaagccc aggtgtccgg ccagggggat agcctgcacg agcacattgc	3120
caatctggcc ggcagccccg ccattaagaa gggcatcctg cagacagtga aggtggtgga	3180
cgagctcgtg aaagtgatgg gccggcacaa gcccgagaac atcgtgatcg aaatggccag	3240
agagaaccag accaccaga agggacagaa gaacagccgc gagagaatga agcggatcga	3300
agagggcatc aaagagctgg gcagccagat cctgaaagaa cccccgtgg aaaacacca	3360
gctgcagaac gagaagctgt acctgtacta cctgcagaat gggcgggata tgtacgtgga	3420
ccaggaaactg gacatcaacc ggctgtccga ctacgatgtg gaccatatcg tgccctcagag	3480
ctttctgaag gacgactcca tcgacaacaa ggtgctgacc agaagcgaca agaaccgggg	3540
caagagcgac aacgtgccct ccgaagaggt cgtgaagaag atgaagaact actggcggca	3600
gctgctgaac gccaaactga ttaccacagag aaagtctgac aatctgacca aggccgagag	3660
agggggcctg agcgaactgg ataaggccgg cttcatcaag agacagctgg tggaaaccog	3720
gcagatcaca aagcacgtgg cacagatcct ggactcccg atgaacacta agtacgacga	3780
gaatgacaag ctgatccggg aagtgaaggt gatcaccctg aagtccaagc tgggtgtccga	3840
tttcgggaag gatttccagt tttaaaagt gcgcgagatc aacaactacc accacgccc	3900
cgacgcctac ctgaacgccg tcgtgggaac cgccctgatc aaaaagtacc ctaagctgga	3960
aagcgagttc gtgtacggcg actacaaggt gtacgacgtg cggaagatga tcgccaagag	4020
cgagcaggaa atcggaagg ctaccgccaa gtacttcttc tacagcaaca tcatgaactt	4080
tttcaagacc gagattacc tggtcaacgg cgagatccgg aagcggcctc tgatcgagac	4140
aaacggcgaa accggggaga tcgtgtggga taaggccggg gattttgcca ccgtgcggaa	4200
agtgtgagc atgccccaa tgaatatcgt gaaaaagacc gaggtgcaga caggcggctt	4260

```

cagcaaagag tctatcctgc ccaagaggaa cagcgataag ctgatcgcca gaaagaagga 4320
ctgggaccct aagaagtacg gcggcttcga cagccccacc gtggcctatt ctgtgctggt 4380
ggtggccaaa gtggaaaagg gcaagtccaa gaaactgaag agtgtgaaag agctgctggg 4440
gatcaccatc atggaaagaa gcagcttcga gaagaatccc atcgactttc tggaagccaa 4500
gggctacaaa gaagtgaaaa aggacctgat catcaagctg cctaagtact cctgtttcga 4560
gctggaaaac ggccggaaga gaatgctggc ctctgccggc gaactgcaga agggaaacga 4620
actggccctg ccctccaaat atgtgaactt cctgtacctg gccagccact atgagaagct 4680
gaagggtccc ccgaggata atgagcagaa acagctgttt gtggaacagc acaagcacta 4740
cctggacgag atcatcgagc agatcagcga gttctccaag agagtgatcc tggccgacgc 4800
taatctggac aaagtgcgtg ccgcctacaa caagcacccg gataagccca tcagagagca 4860
ggccgagaat atcatccacc tgtttaccct gaccaatctg ggagcccctg ccgccttcaa 4920
gtactttgac accaccatcg accggaagag gtacaccagc accaaagagg tgctggacgc 4980
caccctgatc caccagagca tcaccggcct gtacgagaca cggatcgacc tgtctcagct 5040
gggaggcgac tttctttttc ttagcttgac cagctttctt agtagcagca ggacgcttta 5100
a 5101

```

```

<210> 59
<211> 137
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 59
nnnnnnnnnn nnnnnnnnnn gtttttgtac totcaagatt tagaaataaa tcttgcagaa 60
gctacaaaga taaggcttca tgccgaaatc aacacctgt catTTTTatgg caggggtgtt 120
tcgttatTTta atTTTTt 137

```

```

<210> 60
<211> 123
<212> DNA
<213> Sequência Artificial

<220>
<221> source

```

<223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>

<221> base_modificada

<222> (1)..(20)

<223> a, c, t, g, outro ou desconhecido

<400> 60

nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

ggttcattgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 61

<211> 110

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>

<221> base_modificada

<222> (1)..(20)

<223> a, c, t, g, outro ou desconhecido

<400> 61

nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

ggttcattgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt 110

<210> 62

<211> 137

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>

<221> base_modificada

<222> (1)..(20)

<223> a, c, t, g, outro ou desconhecido

<400> 62

nnnnnnnnnn nnnnnnnnnn gttattgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaaaga taaggcttca tgccgaaatc aacacootgt ctttttatgg caggggtgtt 120

tcgttattta atttttt 137

<210> 63
 <211> 123
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 63
 nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caaagataag 60
 gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120
 ttt 123

<210> 64
 <211> 110
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 64
 nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caaagataag 60
 gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt 110

<210> 65
 <211> 137
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 65

```

nnnnnnnnnn nnnnnnnnnn gttattgtac tctcaagatt tagaaataaa tcttgcagaa      60
gctacaatga taaggcttca tgccgaaatc aacaccctgt ctttttatgg caggggtgtt      120
tcgttattta atttttt                                     137

```

```

<210> 66
<211> 123
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 66
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caatgataag      60
gcttcatgcc gaaatcaaca cctgtcatt ttatggcagg gtgttttcgt tatttaattt      120
ttt                                                         123

```

```

<210> 67
<211> 110
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

```

```

<400> 67
nnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaat gcagaagcta caatgataag      60
gcttcatgcc gaaatcaaca cctgtcatt ttatggcagg gtgttttttt      110

```

```

<210> 68
<211> 107
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

```

```

<220>
<221> base modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 68
nnnnnnnnnn nnnnnnnnnn gtttttagagc tgtggaaaca cagcgagtta aaataaggct      60
tagtcogtac tcaacttgaa aagggtggcac cgattcgggtg tttttttt                  107

<210> 69
<211> 4263
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial:Polinucleótido
        sintético"

<400> 69
atgaaaaggc cggcggccac gaaaaaggcc ggccaggcaa aaaagaaaaa gaccaagccc      60
tacagcatcg gcctggacat cggcaccaat agcgtgggct gggccgtgac caccgacaac      120
tacaagggtg ccagcaagaa aatgaagggtg ctgggcaaca cctccaagaa gtacatcaag      180
aaaaacctgc tgggctgtgt gctgttcgac agcggcatta cagccgaggg cagacgggtg      240
aagagaaccg ccagacggcg gtacaccggg cggagaaaca gaatcctgtg tctgcaagag      300
atcttcagca ccgagatggc taccctggac gacgccttct tccagcggct ggacgacagc      360
ttcctggtgc ccgacgacaa gcgggacagc aagtacccca tcttcggcaa cctggtggaa      420
gagaaggcct accacgacga gttccccacc atctaccacc tgagaaagta cctggccgac      480
agcaccaaga aggccgacct gagactggtg tatctggccc tggccacat gatcaagtac      540
cggggccact tctgatcga gggcgagttc aacagcaaga acaacgacat ccagaagaac      600
ttccaggact tcttgacac ctacaacgcc atcttcgaga ggcacctgtc cctggaaaac      660
agcaagcagc tgggaagagat cgtgaaggac aagatcagca agctggaaaa gaaggaccgc      720
atcctgaagc tgttccccgg cgagaagaac agcggaatct tcagcgagtt tctgaagctg      780
atcgtgggca accaggccga cttcagaaag tgcttcaacc tggacgagaa agccagcctg      840
cacttcagca aagagagcta cgacgaggac ctggaaaccc tgctgggata tatcggcgac      900
gactacagcg acgtgttcct gaaggccaag aagctgtacg acgctatcct gctgagcggc      960
ttcctgaccg tgaccgacaa cgagacagag gcccactga gcagcgccat gattaagcgg      1020
tacaacgagc acaaagagga tctggctctg ctgaaagagt acatccggaa catcagcctg      1080
aaaacctaca atgaggtgtt caaggacgac accaagaacg gctaogccgg ctacatcgac      1140
ggcaagacca accaggaaga tttctatgtg tacctgaaga agctgctggc cgagttcgag      1200

```

T com cada fragmento com 4, ou menos do que 4 (por exemplo, 3 ou 2) T contíguos. Podem ser utilizadas para a substituição outras bases que não A, por exemplo C ou G, ou nucleótidos que não ocorrem naturalmente ou nucleótidos modificados. Se a cadeia de T está envolvida na formação de uma forquilha (ou alça de haste, do inglês *hairpin*), então é vantajoso que a base complementar para a base que não é T seja alterada para complementar o nucleótido que não é T. Por exemplo, se a base que não é T é um A, então o seu complemento pode ser mudado para um T, por exemplo, para preservar ou ajudar na preservação da estrutura secundária. Por exemplo, 5'-TTTTT pode ser alterado para se tornar em 5'-TTTAT e o complementar 5'-AAAAA pode ser alterado para 5'-ATAAA.

Num aspeto, a divulgação prevê o sistema CRISPR-Cas ou o sistema enzimático CRISPR em que a modificação compreende a adição de uma sequência de finalização poli-T. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a adição de uma sequência de finalização poli-T em sequências tracr e/ou de emparelhamento tracr. Num aspeto, a divulgação prevê o sistema CRISPR-Cas ou o sistema enzimático CRISPR em que a modificação compreende a adição de uma sequência de finalização poli-T na sequência guia. A sequência de finalização poli-T pode compreender 5 bases contíguas T, ou mais do que 5 bases.

Num aspeto, a presente divulgação prevê o sistema CRISPR-Cas ou o sistema enzimático CRISPR em que a modificação compreende a alteração de alças e/ou forquilhas. Num aspeto a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a

```

ggggccgact actttctgga aaaaatcgac cgcgaggatt tcctgcggaa gcagcggacc 1260
ttcgacaacg gcagcatccc ctaccagatc catctgcagg aaatgcgggc catcctggac 1320
aagcaggcca agttctaccc attcctggcc aagaacaaag agcggatoga gaagatcctg 1380
accttcgca tcccttacta cgtgggcccc ctggccagag gcaacagcga ttttgccctg 1440
tccatccgga agcgcaatga gaagatcacc ccctggaact tcgaggacgt gatcgacaaa 1500
gagtccagcg ccgaggcctt catcaaocgg atgaccagct tcgacctgta cctgcccgag 1560
gaaaagggtg tgcccaagca cagcctgctg tacgagacat tcaatgtgta taacgagctg 1620
accaaagtgc ggtttatcgc caggtctatg cgggactacc agttcctgga ctccaagcag 1680
aaaaaggaca tcgtgcccgt gtacttcaag gacaagcggg aagtgaccga taaggacatc 1740
atcgagtacc tgcacgccat ctacggctac gatggcatcg agctgaaggg catcgagaag 1800
cagttcaact ccagcctgag cacataccac gacctgctga acattatcaa cgacaaagaa 1860
tttctggacg actccagcaa cgaggccatc atogaagaga tcatccacac cctgaccatc 1920
tttgaggacc gcgagatgat caagcagcgg ctgagcaagt tcgagaacat cttcgacaaag 1980
agcgtgctga aaaagctgag cagacggcac tacaccggct ggggcaagct gagcgccaag 2040
ctgatcaacg gcatccggga cgagaagtcc ggcaacacaa tcctggacta cctgatcgac 2100
gacggcatca gcaaccggaa cttcatgcag ctgatccacg acgacgccct gagcttcaag 2160
aagaagatcc agaaggccca gatcatcggg gacgaggaca agggcaacat caaagaagtc 2220
gtgaagtccc tgcccggcag ccccgccatc aagaaggga tcctgcagag catcaagatc 2280
gtggacgagc tcgtgaaagt gatgggcggc agaaagcccg agagcatcgt ggtggaaatg 2340
gctagagaga accagtacac caatcagggc aagagcaaca gccagcagag actgaagaga 2400
ctggaaaagt ccctgaaaga gctgggcagc aagattctga aagagaatat ccctgccaaag 2460
ctgtccaaga tcgacaacaa cgccctgcag aacgaccggc tgtacctgta ctacctgcag 2520
aatggcaagg acatgtatac aggcgacgac ctggatatcg accgcctgag caactacgac 2580
atcgaccata ttatccccc ggccttcctg aaagacaaca gcattgacaa caaagtctg 2640
gtgtcctccg ccagcaaccg cggcaagtcc gatgatgtgc ccagcctgga agtcgtgaaa 2700
aagagaaaga cttctggta tcagctgctg aaaagcaagc tgattagcca gaggaagttc 2760
gacaacctga ccaaggccga gagaggcggc ctgagccctg aagataaggc cggcttcac 2820
cagagacagc tggtggaac ccggcagatc accaagcacg tggccagact gctggatgag 2880
aagtttaaca acaagaagga cgagaacaac cgggcccgtg ggaccgtgaa gatcatcacc 2940
ctgaagtcca ccctggtgtc ccagttccgg aaggacttcg agctgtataa agtcgcgcag 3000
atcaatgact ttcaaccgc ccacgacgc tacctgaatg ccgtggtggc ttccgcctg 3060
ctgaagaagt accctaagct ggaacccgag ttcgtgtacg gcgactacc caagtacaac 3120

```



```

tctttcagag agcggaagtc cgccaccgag aaggtgtact tctactccaa catcatgaat 3180
atctttaaga agtccatctc cctggccgat ggcagagtga tgcagcggcc cctgatcgaa 3240
gtgaacgaag agacaggcga gagcgtgtgg aacaaagaaa gcgacctggc caccgtgctg 3300
cggtgtctga gttatcctca agtgaatgtc gtgaagaagg tgaagaaca gaaccacggc 3360
ctggatcggg gcaagcccaa gggcctgttc aacgccaaac tgtccagcaa gcctaagccc 3420
aactccaacg agaatctcgt gggggocaaa gagtacctgg acctaaaga gtacggcgga 3480
tacgcgggca tctccaatag ctccaccgtg ctctgtgaagg gcacaatcga gaaggcgct 3540
aagaaaaaga tcacaaacgt gctggaattt caggggatct ctatcctgga ccggatcaac 3600
taccggaagg ataagctgaa ctttctgtct gaaaaaggct acaaggacat tgagctgatt 3660
atcgagctgc ctaagtactc cctgttcgaa ctgagcgacg gctccagacg gatgctggcc 3720
tccatcctgt ccaccaacaa caagcggggc gagatccaca agggaaacca gatcttcctg 3780
agccagaaat ttgtgaaact gctgtaccac gccaaagcga tctccaacac catcaatgag 3840
aaccacggga aatacgtgga aaaccacaag aaagagtttg aggaactgtt ctactacatc 3900
ctggagttca acgagaacta tgtgggagcc aagaagaacg gcaaactgct gaactccgcc 3960
ttccagagct ggcagaacca cagcatcgac gagctgtgca gctccttcac cggcctacc 4020
ggcagcgagc ggaagggaact gtttgagctg acctccagag gctctgcgc cgactttgag 4080
ttcctgggag tgaagatccc ccggtacaga gactacacc cctctagtct gctgaaggac 4140
gccaccctga tccaccagag cgtgaccggc ctgtacgaaa cccggatcga cctggctaag 4200
ctgggcgagg gaaagcgtcc tgctgtact aagaaagctg gtcaagctaa gaaaaagaaa 4260
taa 4263

```

<210> 70

<211> 84

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 70

```
ggaaccattc ataacagcat agcaagttat aataaggcta gtccgttatc aacttgaaaa 60
```

```
agtggcaccg agtgggtgct tttt 84
```

<210> 71

<211> 36

<212> DNA

<213> Sequência Artificial

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 71
gttatagagc tatgctgtta tgaatggtcc caaaac 36

<210> 72
<211> 84
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 72
ggaaccattc aatacagcat agcaagttaa tataaggcta gtccgttata aacttgaaaa 60
agtggcaccg agtcggtgct tttt 84

<210> 73
<211> 36
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 73
gtattagagc tatgctgtat tgaatggtcc caaaac 36

<210> 74
<211> 103
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 74
nnnnnnnnnn nnnnnnnnnn gtttttagagc tagaaatagc aagttaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 75
<211> 103

```

```

<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 75
nnnnnnnnnn nnnnnnnnnn gtattagagc tagaaatagc aagttaatat aaggctagtc      60
cgttatcaac ttgaaaaagt ggcaccgagt cgggtgctttt ttt                        103

<210> 76
<211> 123
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 76
nnnnnnnnnn nnnnnnnnnn gtttttagagc tatgctgttt tggaacaaa acagcatagc      60
aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cgggtgctttt      120
ttt                                                                123

<210> 77
<211> 123
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 77
nnnnnnnnnn nnnnnnnnnn gtattagagc tatgctgtat tggaacaat acagcatagc      60

```

aagttaatat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt 120
 ttt 123

<210> 78
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 78
 gtcacctcca atgactaggg 20

<210> 79
 <211> 984
 <212> PRT
 <213> Campylobacter jejuni

<400> 79
 Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
 1 5 10 15

Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
 20 25 30

Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
 35 40 45

Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg
 50 55 60

Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
 65 70 75 80

Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
 85 90 95

Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu
 100 105 110

Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg
 115 120 125

Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala
 130 135 140

Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln
 145 150 155 160

Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu
 165 170 175

Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu
 180 185 190

Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe
 195 200 205

Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu
 210 215 220

Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser
 225 230 235 240

His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro
 245 250 255

Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile
 260 265 270

Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys
 275 280 285

Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu
 290 295 300

Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu
 305 310 315 320

Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys
 325 330 335

Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu
 340 345 350

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
 355 360 365

Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
 370 375 380

Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
 385 390 395 400

Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
 405 410 415

Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
 420 425 430

Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr
 435 440 445

Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn
 450 455 460

Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
 465 470 475 480

Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
 485 490 495

Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys
 500 505 510

Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
 515 520 525

Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
 530 535 540

Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
 545 550 555 560

Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu
 565 570 575

Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
 580 585 590

Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala
 595 600 605

Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
 610 615 620

Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
 625 630 635 640

Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp
 645 650 655

Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln
 660 665 670

Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser

675	680	685
Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His 690 695 700		
Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser 705 710 715 720		
Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser 725 730 735		
Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys 740 745 750		
Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp 755 760 765		
Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser 770 775 780		
Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln 785 790 795 800		
Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys 805 810 815		
Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg 820 825 830		
Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro 835 840 845		
Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val 850 855 860		
Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu 865 870 875 880		
Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile 885 890 895		
Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe 900 905 910		
Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe 915 920 925		

Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
 930 935 940

Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
 945 950 955 960

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
 965 970 975

Arg Gln Arg Glu Asp Phe Lys Lys
 980

<210> 80

<211> 91

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 80

tataatctca taagaaattt aaaaaggac taaaataaag agtttgcggg actctgcggg 60

gttacaatcc cctaaaaccg cttttaaaat t 91

<210> 81

<211> 36

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 81

attttaccat aaagaaattt aaaaaggac taaaac 36

<210> 82

<211> 95

<212> RNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>

<221> base_modificada

<222> (1)..(20)

<223> a, c, u, g, outro ou desconhecido

<400> 82

nnnnnnnnnn nnnnnnnnnn guuuuagucc cgaaagggac uaaaaaaaag aguuugcggg 60
acucugcggg guuacaaucc ccuaaaaccg guuuu 95

<210> 83
<211> 69
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 83
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaagc 60
cguuuuuuuu 69

<210> 84
<211> 69
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 84
gacaucgaug uccuccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaagc 60
cguuuuuuuu 69

<210> 85
<211> 69
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 85
gaguccgagc agaagaaga guuuuagagc uagaaauagc aaguuaaaau aaggcuaagc 60
cguuuuuuuu 69

<210> 86
<211> 69
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

previsão de um mínimo de duas forquilhas na sequência guia. Num aspeto a divulgação prevê o sistema CRISPR-Cas ou o sistema enzimático CRISPR em que a modificação compreende a previsão de uma forquilha formada por complementaridade entre a sequência tracr e a de emparelhamento tracr (repetição direta). Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a previsão de uma forquilha ou mais forquilhas adicionais na, ou na, direção da extremidade 3' da sequência tracrRNA. Por exemplo, uma forquilha pode ser formada fornecendo sequências autocomplementares no interior da sequência tracrRNA unidas por uma alça de tal modo que uma forquilha é formada no auto enrolamento (*folding*). Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a previsão de forquilhas adicionais adicionadas ao extremo 3' da sequência guia. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a extensão da extremidade 5' da sequência guia. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a previsão de uma ou mais forquilhas na extremidade 5' da sequência guia. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a anexação da sequência (5'-AGGACGAAGTCCTAA) à extremidade 5' da sequência guia. Outras sequências adequadas para a formação de forquilhas serão conhecidas dos peritos na técnica, e podem ser usadas em certos aspetos da divulgação. Em alguns aspetos da divulgação, são previstas pelo menos 2, 3, 4, 5, ou mais forquilhas adicionais. Em alguns aspetos da divulgação, são previstas não mais do que 10, 9, 8, 7, 6 forquilhas

```

<400> 86
ggggccgaga uggguguuc guuuuagagc uagaaaagc aaguaaaaau aaggcuaguc      60
cguuuuuuuu                                                                69

<210> 87
<211> 69
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 87
guggcgagag gggccgagau guuuuagagc uagaaaagc aaguaaaaau aaggcuaguc      60
cguuuuuuuu                                                                69

<210> 88
<211> 76
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 88
gucaccucca augacuaggg guuuuagagc uagaaaagc aaguaaaaau aaggcuaguc      60
cguuaucauu uuuuuu                                                        76

<210> 89
<211> 76
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 89
gacaucgaug uccuccccau guuuuagagc uagaaaagc aaguaaaaau aaggcuaguc      60
cguuaucauu uuuuuu                                                        76

<210> 90
<211> 76
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

```

<400> 90
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucauu uuuuuu 76

<210> 91
<211> 76
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 91
ggggccgaga ugggguuuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucauu uuuuuu 76

<210> 92
<211> 76
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 92
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucauu uuuuuu 76

<210> 93
<211> 88
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 93
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu guuuuuuu 88

<210> 94
<211> 88
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido"

sintético"

<400> 94
gacauccgaug uccuccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguaaucaac uugaaaaagu guuuuuuu 88

<210> 95
<211> 88
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
sintético"

<400> 95
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguaaucaac uugaaaaagu guuuuuuu 88

<210> 96
<211> 88
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
sintético"

<400> 96
ggggccgaga ugggguuguu guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguaaucaac uugaaaaagu guuuuuuu 88

<210> 97
<211> 88
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
sintético"

<400> 97
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguaaucaac uugaaaaagu guuuuuuu 88

<210> 98
<211> 103
<212> RNA
<213> Sequência Artificial

<220>
<221> source

```

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 98
gucaccucca augacuaggg guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc      60

cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu                          103

<210> 99
<211> 103
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 99
gacauagcug uccuccccau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc      60

cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu                          103

<210> 100
<211> 103
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 100
gaguccgagc agaagaagaa guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc      60

cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu                          103

<210> 101
<211> 103
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 101
ggggccgaga uggguuguuc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc      60

cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu                          103

<210> 102
<211> 103
<212> RNA
<213> Sequência Artificial

<220>

```

```

<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
sintético"

<400> 102
guggcgagag gggccgagau guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu uuu 103

<210> 103
<211> 102
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
sintético"

<400> 103
gttttagagc tatgctgttt tgaatggtcc caaaacggaa ggcctgagt ccgagcagaa 60
gaagaagttt tagagctatg ctgttttgaa tggccccaa ac 102

<210> 104
<211> 100
<212> DNA
<213> Homo sapiens

<400> 104
cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag 60
aagaagggct cccatcacat caacoggtgg cgcattgcc 100

<210> 105
<211> 50
<212> DNA
<213> Homo sapiens

<400> 105
agctggagga ggaagggcct ggtccgagc agaagaagaa gggctccac 50

<210> 106
<211> 30
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
sintético"

<400> 106
gaguccgagc agaagaagaa guuuuagagc 30

<210> 107
<211> 49

```

```

<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 107
agctggagga ggaagggcct ggtccgagc agaagagaag ggctcccat      49

<210> 108
<211> 53
<212> DNA
<213> Homo sapiens

<400> 108
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccatca cat      53

<210> 109
<211> 52
<212> DNA
<213> Homo sapiens

<400> 109
ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at      52

<210> 110
<211> 54
<212> DNA
<213> Homo sapiens

<400> 110
ctggaggagg aagggcctga gtccgagcag aagaagaag ggctcccatc acat      54

<210> 111
<211> 50
<212> DNA
<213> Homo sapiens

<400> 111
ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat      50

<210> 112
<211> 47
<212> DNA
<213> Homo sapiens

<400> 112
ctggaggagg aagggcctga gcccgagcag aagggtccc atcacat      47

<210> 113
<211> 66
<212> DNA
<213> Sequência Artificial

<220>

```



```

<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<220>
<221> source
<223> /nota="Descrição de Molécula com Combinação DNA/RNA:
        Oligonucleótido sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 113
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguaaaau aaggctagtc      60
cguuuu                                         66

<210> 114
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 114
gaguccgagc agaagaagaa                                         20

<210> 115
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 115
gacaucgaug uccuccccau                                         20

<210> 116
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 116
gucaccucca augacuaggg                                         20

```

<210> 117
 <211> 20
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 117
 auuggguguu cagggcagag

20

<210> 118
 <211> 20
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 118
 guggcgagag gggccgagau

20

<210> 119
 <211> 20
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 119
 ggggccgaga uuggguguuc

20

<210> 120
 <211> 20
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 120
 gugccauuag cuaaaugcau

20

<210> 121
 <211> 20
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source

```

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 121
guaccaccca caggugccag
20

<210> 122
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 122
gaaagccucu gggccaggaa
20

<210> 123
<211> 48
<212> DNA
<213> Homo sapiens

<400> 123
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccat
48

<210> 124
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 124
gaguccgagc agaagaagau
20

<210> 125
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 125
gaguccgagc agaagaagua
20

<210> 126
<211> 20
<212> RNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 126
gaguccgagc agaagaacaa                                20

<210> 127
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 127
gaguccgagc agaagaugaa                                20

<210> 128
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 128
gaguccgagc agaaguagaa                                20

<210> 129
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 129
gaguccgagc agaugaagaa                                20

<210> 130
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 130
gaguccgagc acaagaagaa                                20

```

adicionais. Num aspeto, a divulgação prevê um sistema de CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende duas forquilhas. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende três forquilhas. Num aspeto, a divulgação prevê um sistema de CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende um máximo de cinco forquilhas.

Num aspeto, a divulgação prevê um sistema de CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a previsão de reticulação (*cross linking*), ou previsão de um ou mais nucleótidos modificados na sequência de polinucleótidos. Nucleótidos modificados e/ou reticulação podem ser fornecidos em qualquer ou todas as sequências tracr, de emparelhamento tracr, e/ou guia, e/ou na sequência codificante das enzimas, e/ou nas sequências de vetores. As modificações podem incluir a inclusão de pelo menos um nucleótido que não ocorre naturalmente, ou um nucleótido modificado, ou seus análogos. Os nucleótidos modificados podem ser modificados na região da ribose, do fosfato, e/ou da base. Os nucleótidos modificados podem incluir análogos 2'-O-metilo, análogos 2'-desoxi, ou análogos 2'-fluoro. A base dos ácidos nucleicos pode ser modificada, por exemplo, usando uma base de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos unidos por ponte (BNA) pode também ser possível. Outros exemplos de bases modificadas incluem, mas não estão limitados a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina.

Deve entender-se que qualquer uma ou todas as modificações acima pode ser fornecida de forma separada ou em

```

<210> 131
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 131
gaguccgagg agaagaagaa                                20

<210> 132
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 132
gaguccgugc agaagaagaa                                20

<210> 133
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 133
gagucggagc agaagaagaa                                20

<210> 134
<211> 20
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 134
gagaccgagc agaagaagaa                                20

<210> 135
<211> 24
<212> DNA
<213> Sequência Artificial

```

```

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 135
aatgacaagc ttgctagcgg tggg
24

<210> 136
<211> 39
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 136
aaaacggaag ggctgagtc cgagcagaag aagaagttt
39

<210> 137
<211> 39
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 137
aaacaggggc cgagattggg tgttcagggc agaggtttt
39

<210> 138
<211> 38
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 138
aaaacggaag ggctgagtc cgagcagaag aagaagtt
38

<210> 139
<211> 40
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 139
aacggaggga ggggcacaga tgagaaactc agggttttag
40

```

```

<210> 140
<211> 38
<212> DNA
<213> Homo sapiens

<400> 140
agcccttctt cttctgctcg gactcaggcc ettcctcc 38

<210> 141
<211> 40
<212> DNA
<213> Homo sapiens

<400> 141
cagggaggga ggggcacaga tgagaaactc aggaggcccc 40

<210> 142
<211> 80
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 142
ggcaatggcg caccggttga tgtgatggga gcccttctag gaggcccca gagcagccac 60
tggggcctca acactcaggc 80

<210> 143
<211> 98
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 143
ggacgaaaca ccggaaccat tcaaaacagc atagcaagtt aaaataaggc tagtccgtta 60
tcaacttgaa aaagtggcac cgagtcggtg cttttttt 98

<210> 144
<211> 186
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<400> 144

```


ggacgaaaca cggtagtat taagtattgt tttatggctg ataaatttct ttgaatttct 60
 ccttgattat ttgttataaa agttataaaa taatcttgtt ggaaccattc aaaacagcat 120
 agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 180
 tttttt 186

<210> 145
 <211> 46
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(19)
 <223> a, c, u, g, outro ou desconhecido

<400> 145
 nnnnnnnnnn nnnnnnnnng uuauuguacu cucaagauuu auuuuu 46

<210> 146
 <211> 91
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 146
 guuacuuasa ucuugcagaa gcuacaaaga uaaggcuuca ugccgaaauac aacacccugu 60
 cauuuuauugy caggguguuu ucguuuuuu a 91

<210> 147
 <211> 70
 <212> DNA
 <213> Homo sapiens

<400> 147
 ttttctagtg ctgagtttct gtgactcctc tacattctac ttctctgtgt ttctgtatac 60
 tacctcctcc 70

<210> 148
 <211> 122
 <212> DNA
 <213> Homo sapiens

<400> 148
 ggaggaaggg cctgagtcgg agcagaagaa gaagggtcc catcacatca accggtggcg 60

cattgccacg aagcaggcca atggggagga catcgatgtc acctocaatg actagggtgg 120

gc 122

<210> 149

<211> 48

<212> RNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>

<221> base_modificada

<222> (3)..(32)

<223> a, c, u, g, outro ou desconhecido

<400> 149

acnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnguuvuaga gcuaugcu 48

<210> 150

<211> 67

<212> DNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>

<221> source

<223> /nota="Descrição de Molécula com Combinação DNA/RNA: Oligonucleótido sintético"

<400> 150

agcauagcaa guaaaaauaa ggctaguccg uaucaacu u gaaaaagugg caccgagucg 60

gugcuuu 67

<210> 151

<211> 62

<212> RNA

<213> Sequência Artificial

<220>

<221> source

<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>

<221> base_modificada

<222> (1)..(20)

<223> a, c, u, g, outro ou desconhecido

<400> 151
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cg 62

<210> 152
 <211> 73
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 152
 tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagttt tagagctatg 60
 ctgttttgaa tgg 73

<210> 153
 <211> 99
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, u, g, outro ou desconhecido

<400> 153
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu 99

<210> 154
 <211> 127
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 154
 guuuuugac ucuaagauu uaaguaacug uacaacguua cuuaaaucuu gcagaagcua 60
 caaagauaag gcuucaugcc gaaaucaaca ccugugcauu uuauggcagg guguuuucgu 120
 uauuuuaa 127

<210> 155

```

<211> 56
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 155
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt taagtaactg tacaac      56

<210> 156
<211> 91
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 156
gttacttaaa tcttgcagaa gctacaaaga taagggttca tgccgaaatc aacaccctgt      60
cattttatgg cagggtgttt tcgttattta a                                     91

<210> 157
<211> 134
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, t, g, outro ou desconhecido

<400> 157
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt taaggaaact aaatcttgca      60
gaagctacaa agataaggct tcatgccgaa atcaacaccc tgcatttta tggcagggtg      120
ttttcgttat ttaa                                                         134

<210> 158
<211> 131
<212> DNA
<213> Sequência Artificial

```

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 158
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa ttttgcagaa 60
 gctacaasaga taaggcttca tgccgaaatc aacaccctgt ctttttatgg cagggtgttt 120
 tcgttatatta a 131

<210> 159
 <211> 125
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 159
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatg aaaatcttgc agaagctaca 60
 aagataaggc ttcattgccga aatcaacacc ctgtcatttt atggcagggt gttttcgtta 120
 tttaa 125

<210> 160
 <211> 112
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 160
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctgaaaaga agctacaaag ataaggcttc 60

atgccgaaat caacaccctg tcattttatg gcagggtgtt ttcgttattt aa 112

<210> 161
 <211> 107
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 161
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tgaaaagcta caaagataag gtttcattgcc 60
 gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaa 107

<210> 162
 <211> 108
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 162
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttcagaa 60
 gctacaaaga taaggcttca tgcggaatc aacaccctgt catTTTTat 108

<210> 163
 <211> 86
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 163
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60
 gctacaaaaga taaggcttca tgccga 86

<210> 164
 <211> 79
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 164
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60
 gctacaaaaga taaggcttc 79

<210> 165
 <211> 73
 <212> DNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, t, g, outro ou desconhecido

<400> 165
 nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60
 gctacaaaaga taa 73

<210> 166
 <211> 125
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 166
 guuuuagucc cuuuuuuuuu uucuuuauagg uaaaauuuaa aucucauaag aaauuuuaaa 60

combinação, num dado sistema CRISPR-Cas ou num sistema enzimático CRISPR. Tal sistema pode incluir uma, duas, três, quatro, cinco, ou mais das referidas modificações.

Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR, em que a enzima CRISPR é uma enzima do sistema CRISPR de tipo II, por exemplo, uma enzima Cas9. Num aspeto, a presente divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a enzima CRISPR compreende menos que mil aminoácidos, ou menos que quatro mil aminoácidos. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a enzima Cas9 é StCas9 ou StlCas9, ou a enzima Cas9 é uma enzima Cas9 de um organismo selecionado a partir do grupo que consiste no género *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* ou *Corynebacter*. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a enzima CRISPR é uma nuclease que dirige a clivagem de ambas as cadeias no local da sequência alvo.

Num aspeto, a presente divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que o primeiro elemento regulador é um promotor da polimerase III. Num aspeto, a divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que o segundo elemento regulador é um promotor da polimerase II.

Num aspeto, a presente divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a sequência guia compreende pelo menos quinze nucleótidos.

agggacuaaa auaaagaguu ugcggggacuc ugccggggua caaucccccua aaaccgcuuu 120
uaaaa 125

<210> 167
<211> 91
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 167
guuacuuuaa ucuugcagaa gcuacaaaga uaaggcuuca ugccgaauc aacaccugu 60
cauuuuauagg caggguuuu uoguuaauua a 91

<210> 168
<211> 56
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 168
gggacucaac caagucuuuc guuuuuguac ucucaagauu uaaguaacug uacaac 56

<210> 169
<211> 147
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 169
gggacucaac caagucuuuc guuuuuguac ucucaagauu uaaguaacug uacaacguua 60
cuuaaaucuu gcagaagcua caaagauaag gcuucaugcc gaaaucaaca ccugucuuu 120
uuauaggcagg guguuuucgu uuuuuua 147

<210> 170
<211> 70
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 170
 cuugcagaag cuacaaagau aaggcuucau gccgaaauca acaccuguc auuuuauaggc 60
 aggguguuuu 70

<210> 171
 <211> 42
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 171
 gggacucaac caagucauuc guuuuuguac ucucaagauu ua 42

<210> 172
 <211> 112
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 172
 gggacucaac caagucauuc guuuuuguac ucucaagauu uacuugcaga agcuacaaag 60
 auaaggcuuc augccgaaau caacaccug ucauuuuuug gcaggguguu uu 112

<210> 173
 <211> 116
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 173
 gggacucaac caagucauuc guuuuuguac ucucaagauu uagaaacuug cagaagcuac 60
 aaagauaagg cuucaugccg aaaucaacac ccugucauuu uauugcaggg uguuuu 116

<210> 174
 <211> 116
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 174

gggacucaac caagucauuc guuuuuguac ucucaagauu uagaaacuug cagaagcuac 60
 aaagauaagg cuucaugccg aaaucaacac ccugucuuu uauggcaggg uguuuu 116

<210> 175
 <211> 102
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 175
 gggacucaac caagucauuc guuuuuguag aaauacaaag auaaggcuuc augccgaaau 60
 caacacccug ucauuuuuug gcaggguuguu uucguuuuu aa 102

<210> 176
 <211> 102
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<400> 176
 gggacucaac caagucauuc guuuuuguag aaauacaaag auaaggcuuc augccgaaau 60
 caacacccug ucauuuuuug gcaggguuguu uucguuuuu aa 102

<210> 177
 <211> 57
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 177
 gggacucaac caagucauuc guuuuuguag aaauacaaag auaaggcuuc augccga 57

<210> 178
 <211> 57
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Oligonucleótido sintético"

<400> 178
 gggacucaac caagucauuc guuuuuguag aaauacaaag auaaggcuuc augccga 57

```

<210> 179
<211> 23
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 179
gtggtgtcac gctcgtcgtt tgg                                     23

<210> 180
<211> 23
<212> DNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<400> 180
tccagtctat taattgttgc cgg                                     23

<210> 181
<211> 64
<212> DNA
<213> Homo sapiens

<400> 181
caagaggctt gaggtaggaga ggagtgccgc cgaggcgggg cggggcgggg cgtggagctg   60
ggct                                           64

<210> 182
<211> 99
<212> RNA
<213> Sequência Artificial

<220>
<221> source
<223> /nota="Descrição de Sequência Artificial: Oligonucleótido
        sintético"

<220>
<221> base_modificada
<222> (1)..(20)
<223> a, c, u, g, outro ou desconhecido

<400> 182
nnnnnnnnnn nnnnnnnnnn guauuagagc uagaaauagc aaguuaauau aaggcuaguc   60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu                               99

```

<210> 183
 <211> 119
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, u, g, outro ou desconhecido

<400> 183
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaaacaaa acagcauagc 60
 aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu 119

<210> 184
 <211> 119
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, u, g, outro ou desconhecido

<400> 184
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguau uggaaacaa u acagcauagc 60
 aaguuaauau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu 119

<210> 185
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 185
 tagcgggtaa gc 12

<210> 186
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 186
 tcggtgacat gt 12

<210> 187

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 187
 actccccgta gg 12

<210> 188
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 188
 actgcgtggt aa 12

<210> 189
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 189
 acgtcgctg at 12

<210> 190
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 190
 taggtcgacc ag 12

<210> 191
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 191
 ggcgttaatg at 12

<210> 192
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 192
 tgtcgcatgt ta 12

<210> 193
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 193
 atggaaacgc at 12

<210> 194

<211> 12
<212> DNA
<213> Homo sapiens

<400> 194
gccgaattcc tc 12

<210> 195
<211> 12
<212> DNA
<213> Homo sapiens

<400> 195
gcatggtacg ga 12

<210> 196
<211> 12
<212> DNA
<213> Homo sapiens

<400> 196
cggtactctt ac 12

<210> 197
<211> 12
<212> DNA
<213> Homo sapiens

<400> 197
gcctgtgccg ta 12

<210> 198
<211> 12
<212> DNA
<213> Homo sapiens

<400> 198
tacggttaagt cg 12

<210> 199
<211> 12
<212> DNA
<213> Homo sapiens

<400> 199
cacqaaatta cc 12

<210> 200
<211> 12
<212> DNA
<213> Homo sapiens

<400> 200
aaccaagata cg 12

<210> 201

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 201
 gagtcgatac gc 12

<210> 202
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 202
 gtctcacgat cg 12

<210> 203
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 203
 tcgtcgggtg ca 12

<210> 204
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 204
 actccgtagt ga 12

<210> 205
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 205
 caggacgtcc gt 12

<210> 206
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 206
 tcgtatccct ac 12

<210> 207
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 207
 tttcaaggcc gg 12

<210> 208

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 208
 cgccggtgga at 12

<210> 209
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 209
 gaaccggtcc ta 12

<210> 210
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 210
 gattcatcag cg 12

<210> 211
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 211
 acaccggtct tc 12

<210> 212
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 212
 atcgtgcctt aa 12

<210> 213
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 213
 gcgtcaatgt tc 12

<210> 214
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 214
 ctccgtatct cg 12

<210> 215

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 215
 ccgattcctt cg 12

<210> 216
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 216
 tgcgcctcca gt 12

<210> 217
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 217
 taacgtcgga gc 12

<210> 218
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 218
 aaggtcgccc at 12

<210> 219
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 219
 gtcggggact at 12

<210> 220
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 220
 ttcgagcgat tt 12

<210> 221
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 221
 tgagtcgtcg ag 12

<210> 222

Num aspeto, a presente divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a modificação compreende a sequência otimizada tracr e/ou RNA de sequência guia otimizada e/ou estrutura dobra da sequência tracr e/ou sequência ou sequências de emparelhamento tracr e/ou estruturas secundárias para estabilização da sequência tracr e/ou uma sequência tracr com uma região reduzida de emparelhamento de bases e/ou elementos de RNA fundidos de sequência tracr; e/ou, em sistema multiplex existem dois RNAs compreendendo um marcador e compreendendo uma pluralidade de guias ou um RNA que compreende uma pluralidade de quiméricos.

Em aspetos da presente divulgação a arquitetura de RNA quimérico é ainda otimizada de acordo com os resultados de estudos de mutagénese. No RNA quimérico com duas ou mais forquilhas, as mutações na repetição direta proximal para estabilizar a forquilha podem resultar na ablação da atividade do complexo CRISPR. Mutações na repetição direta distal para reduzir ou estabilizar a forquilha podem não ter efeito sobre a atividade do complexo CRISPR. A randomização de sequências na região da protuberância entre as repetições nas extremidades proximal e distal podem reduzir significativamente a atividade do complexo CRISPR. As mudanças individuais de pares de bases ou a randomização de sequências na região adaptadora entre forquilhas podem resultar na perda completa de atividade do complexo CRISPR. A estabilização da forquilha das forquilhas distais que seguem a primeira forquilha após a sequência guia pode resultar numa manutenção ou melhoria da atividade do complexo CRISPR. Por conseguinte, em formas de realização preferidas da divulgação, a arquitetura do RNA quimérico pode ser adicionalmente otimizada através da geração de um

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 222
 tttacgcaga gg 12

<210> 223
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 223
 aggaagtatc gc 12

<210> 224
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 224
 actcgatacc at 12

<210> 225
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 225
 cgctacatag ca 12

<210> 226
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 226
 ttcataaccg gc 12

<210> 227
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 227
 ccaaacggtt aa 12

<210> 228
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 228
 cgattccttc gt 12

<210> 229

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 229
 cgtcatgaat aa 12

<210> 230
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 230
 agtggcgatg ac 12

<210> 231
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 231
 cccctaaggc ac 12

<210> 232
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 232
 gccaacccgc ac 12

<210> 233
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 233
 tgggacaccg gt 12

<210> 234
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 234
 ttgactgcgg cg 12

<210> 235
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 235
 actatgcgta gg 12

<210> 236

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 236
 tcacccaaag cg 12

<210> 237
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 237
 gcaggacgtc cg 12

<210> 238
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 238
 acaccgaaaa cg 12

<210> 239
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 239
 cggtgtattg ag 12

<210> 240
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 240
 cacgaggtat gc 12

<210> 241
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 241
 taaagcgacc cg 12

<210> 242
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 242
 cttagtcggc ca 12

<210> 243

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 243
 cgaaaacgtg gc 12

<210> 244
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 244
 cgtgccctga ac 12

<210> 245
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 245
 tttaccatcg aa 12

<210> 246
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 246
 cgtageccatg tt 12

<210> 247
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 247
 cccaaacggt ta 12

<210> 248
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 248
 gcgttatcag aa 12

<210> 249
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 249
 tcgatggtaa ac 12

<210> 250

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 250
 cgactttttg ca 12

<210> 251
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 251
 tcgacgactc ac 12

<210> 252
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 252
 acgcgtcaga ta 12

<210> 253
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 253
 cgtacggcac ag 12

<210> 254
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 254
 ctatgccgtg ca 12

<210> 255
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 255
 cgcgtcagat at 12

<210> 256
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 256
 aagatcggtg gc 12

<210> 257

<211> 12
 <212> DNA
 <213> Homo sapiens

<400> 257
 cttcgcaagg ag 12

<210> 258
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 258
 gtcgtggact ac 12

<210> 259
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 259
 ggtcgtcatc aa 12

<210> 260
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 260
 gttaacagcg tg 12

<210> 261
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 261
 tagctaaccg tt 12

<210> 262
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 262
 agtaaaggcg ct 12

<210> 263
 <211> 12
 <212> DNA
 <213> Homo sapiens

<400> 263
 ggtaatttcg tg 12

<210> 264

<211> 147
 <212> RNA
 <213> Sequência Artificial

<220>
 <221> source
 <223> /nota="Descrição de Sequência Artificial: Polinucleótido sintético"

<220>
 <221> base_modificada
 <222> (1)..(20)
 <223> a, c, u, g, outro ou desconhecido

<400> 264
 nnnnnnnnnn nnnnnnnnnn guuuuaguac ucuguaauuu uagguaugag guagacgaaa 60
 auuguacuua uaccuaaaau uacagaaucu acuaaaacaa ggcaaaaugc cguguuuauu 120
 uogucaacuu guuggcgaga uuuuuuu 147

Lisboa, 2 de Dezembro de 2015

REIVINDICAÇÕES

1. Um sistema de vetor relacionado com Repetições Palindrômicas Curtas Agrupadas e Regularmente Espaçadas (CRISPR) - sistema associado CRISPR (Cas) (CRISPR-Cas) de origem não natural e obtido por engenharia, compreendendo um ou mais vetores, que compreendem:

- a) um primeiro elemento regulador ligado operacionalmente a uma sequência de nucleótidos que codifica uma sequência de polinucleótidos do sistema CRISPR-Cas que compreendem uma sequência guia, um RNA tracr, e uma sequência de emparelhamento tracr, em que a sequência guia hibridiza com uma ou mais sequências alvo nos loci de polinucleótido numa célula eucariótica,
- b) um segundo elemento regulador, ligado operativamente a uma sequência de nucleótidos que codifica uma proteína Cas9 de Tipo II,
- c) um molde recombinante

em que os componentes (a), (b) e (c) estão localizados em vetores do sistema iguais ou diferentes, em que o sistema compreende um ou mais sinais de localização nuclear (NLS) expressos com a sequência de nucleótidos que codifica a proteína Cas9,

em que a sequência guia tem como alvo um ou mais loci de polinucleótidos numa célula eucariótica e a proteína Cas9 cliva este um ou mais loci de polinucleótido, em que a sequência dos loci de um ou mais polinucleótidos é modificada.

2. Um sistema de vetor CRISPR-Cas de tipo II manipulado por engenharia, de ocorrência não natural, de acordo com a reivindicação 1,

em que a proteína Cas9 está mutada no que diz respeito a um tipo de proteína selvagem Cas9 correspondente de tal modo que a proteína mutada é uma *nickase* que não tem a capacidade de clivar uma cadeia de polinucleótido alvo, em que as sequências guia são dirigidas a um ou mais loci de polinucleótidos numa célula eucariótica e a proteína Cas9 cliva apenas uma cadeia do loci de polinucleótidos, em que a sequência de um ou mais loci de polinucleótidos é modificada.

3. O sistema de acordo com a reivindicação 2, em que a proteína Cas9 compreende uma ou mais mutações nos domínios catalíticos de RuvC I, RuvC II ou RuvC III.

4. O sistema de acordo com a reivindicação 2, em que a proteína Cas9 compreende uma mutação selecionada a partir do grupo que consiste em D10A, H840A, N854A e N863A com referência à numeração das posições de uma proteína Cas9 de *Streptococcus pyogenes* (SpCas9).

5. O sistema de acordo com qualquer uma das reivindicações 1-4, em que o molde recombinante (c) está contido num vetor separado dos componentes (a) e (b).

6. O sistema de acordo com qualquer uma das reivindicações 1-4, em que o molde recombinante (c) é um componente de um vetor que compreende o componente (a) e/ou (b).

7. O sistema de acordo com qualquer uma das reivindicações 1 a 6, em que os componentes (a) e (b) são componentes do mesmo vetor.

8. O sistema de acordo com qualquer uma das reivindicações anteriores em que o molde recombinante é complementar a uma porção de um polinucleótido compreendendo a sequência alvo.

9. O sistema de acordo com qualquer uma das reivindicações anteriores, em que o sistema CRISPR-Cas compreende dois ou mais sinais de localização nuclear expressos com a sequência de nucleótidos que codifica a proteína Cas9.

10. O sistema de acordo com a reivindicação 9, em que pelo menos um NLS encontra-se no terminal amino ou perto deste na proteína Cas9 e/ou pelo menos um NLS encontra-se no terminal carboxílico da proteína Cas9 ou perto deste.

11. O sistema de acordo com a reivindicação 10, em que pelo menos um NLS está localizado no terminal amino na proteína Cas9 ou perto deste e pelo menos um NLS encontra-se no terminal carboxílico da proteína Cas9 ou perto deste.

12. O sistema de acordo com qualquer reivindicação precedente, em que os vetores são vetores virais.

13. O sistema de acordo com a reivindicação 12, em que os vetores virais são vetores retrovirais, lentivirais, adenovirais, adeno-associados, ou herpes simplex.

14. O sistema de acordo com qualquer uma das reivindicações anteriores, em que as sequências de polinucleótidos do

RNA quimérico menor que pode ser benéfico para as opções de administração terapêutica e outros usos e isto pode ser alcançado através da alteração da repetição direta distal, de modo a reduzir ou estabilizar a forquilha. Em outras formas de realização preferidas da presente divulgação, a arquitetura de RNA quimérico pode ser adicionalmente otimizada através da estabilização de uma ou mais das forquilhas distais. A estabilização das forquilhas pode incluir modificações nas sequências adequadas para formar forquilhas. Em alguns aspetos da divulgação, são fornecidas pelo menos, 2, 3, 4, 5, ou mais forquilhas adicionais. Em alguns aspetos da divulgação, são fornecidas não mais do que 10, 9, 8, 7, 6 forquilhas adicionais. Em alguns aspetos da divulgação a estabilização pode ser devida à reticulação e outras modificações. As modificações podem incluir a inclusão de pelo menos um nucleótido que não ocorre naturalmente, ou um nucleótido modificado, ou seus análogos. Os nucleótidos modificados podem ser modificado na região da ribose, fosfato, e/ou base. Os nucleótidos modificados podem incluir análogos 2'-O-metilo, análogos 2'-desoxi, ou análogos 2'-fluoro. A base dos ácidos nucleicos pode ser modificada, por exemplo, usando uma base de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos unidos por ponte (BNA) pode também ser possível. Outros exemplos de bases modificadas incluem, mas não estão limitados a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina.

Num aspeto a presente divulgação prevê um sistema CRISPR-Cas ou um sistema enzimático CRISPR em que a enzima CRISPR tem codões otimizados para a expressão numa célula eucariótica.

sistema CRISPR-Cas compreendem uma sequência guia fundida com uma sequência de transativação cr (tracr).

15. O sistema de acordo com qualquer reivindicação precedente, em que a sequência de polinucleótidos do sistema CRISPR-Cas é um RNA quimérico compreendendo a sequência guia, a sequência tracr, e uma sequência de emparelhamento tracr.

16. O sistema de acordo com qualquer reivindicação precedente, em que a célula eucariótica é uma célula de mamífero ou uma célula humana.

17. O sistema de acordo com qualquer reivindicação precedente, em que a proteína Cas9 tem codões otimizados para a expressão numa célula eucariótica.

18. O sistema de acordo com qualquer reivindicação precedente, em que a sequência guia inclui uma sequência tracr compreendendo 50 ou mais nucleótidos de comprimento.

19. A utilização do sistema de acordo com qualquer uma das reivindicações 1 a 18 para engenharia de genoma, desde que a utilização não compreenda um processo para a modificação da identidade genética da linha germinal de seres humanos, e desde que a referida utilização não seja um método para o tratamento do corpo humano ou animal por cirurgia ou terapia.

20. Utilização do sistema de acordo com qualquer uma das reivindicações 1 a 18 na produção de um animal transgênico não humano ou planta transgênica.

21. O sistema de acordo com qualquer uma das reivindicações 1 a 18 para uso em terapia genética ou edição de genoma.

Lisboa, 2 de Dezembro de 2015

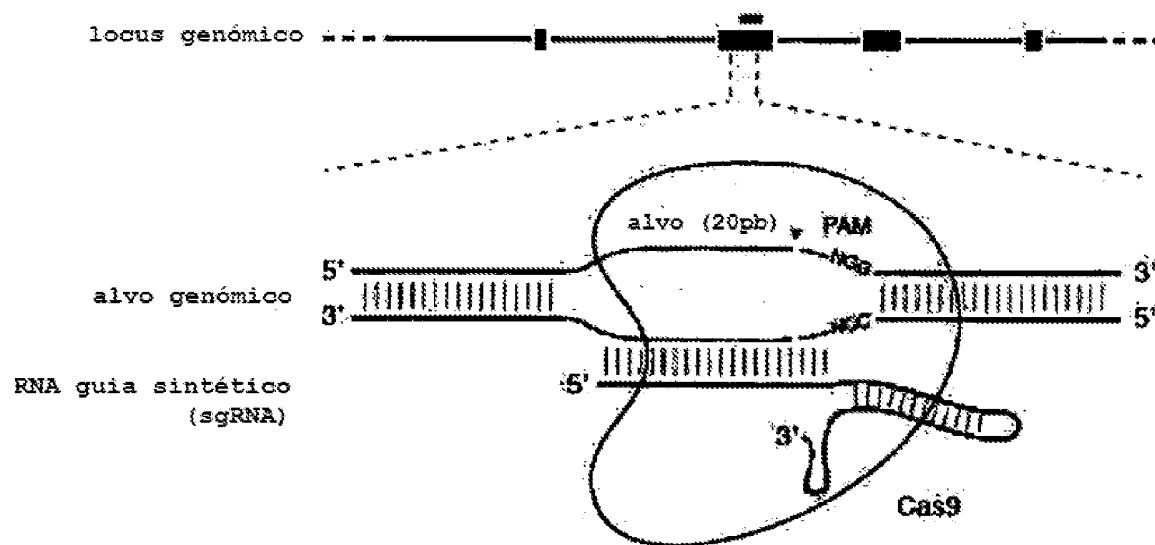


FIG. 1

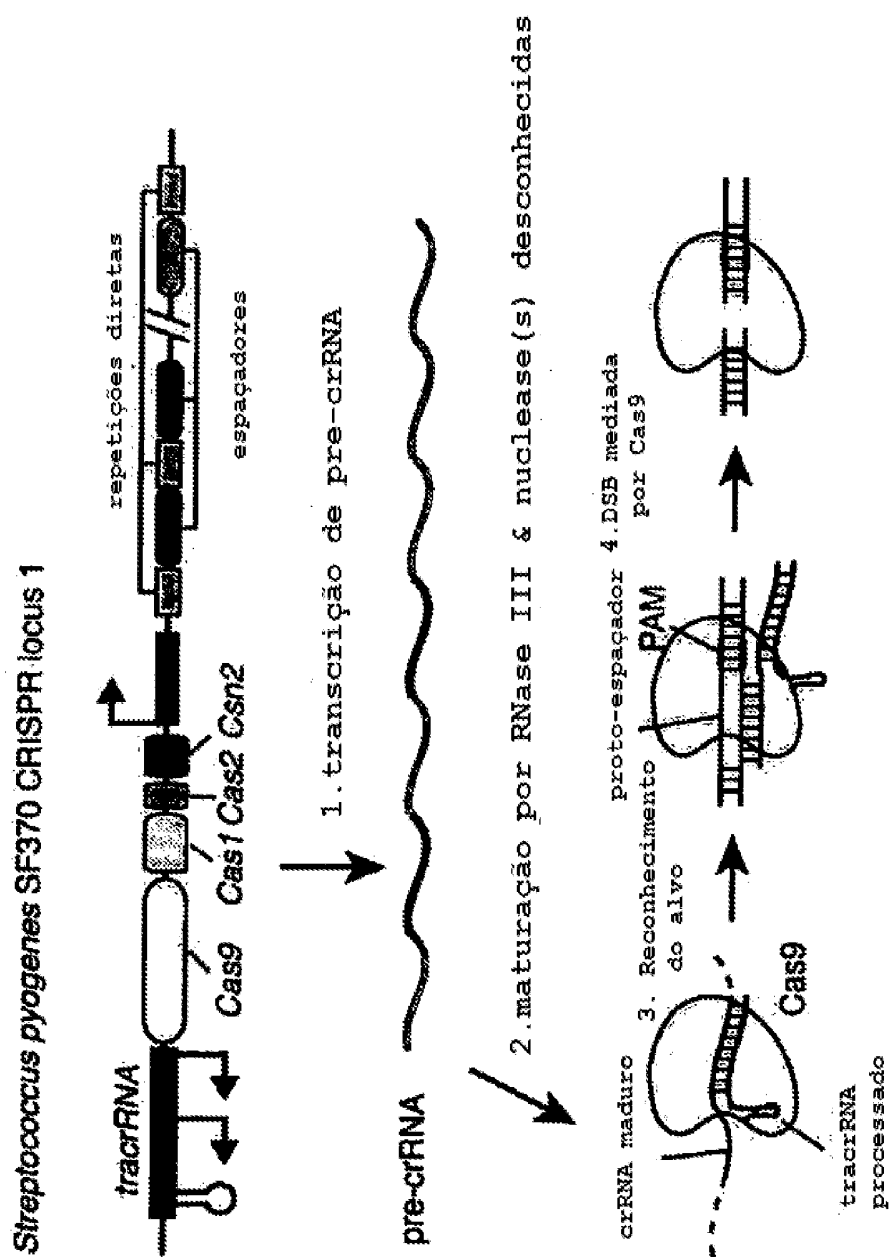


FIG. 2A

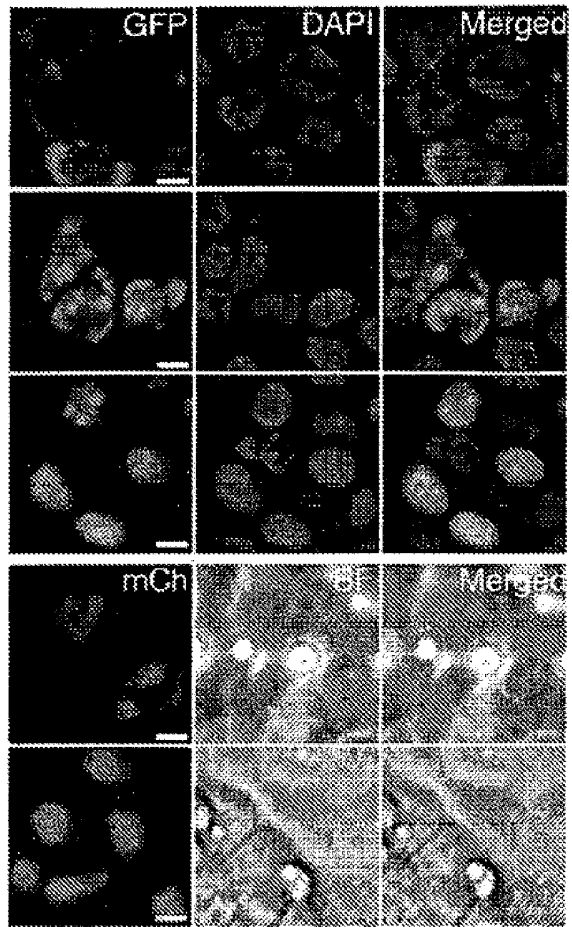
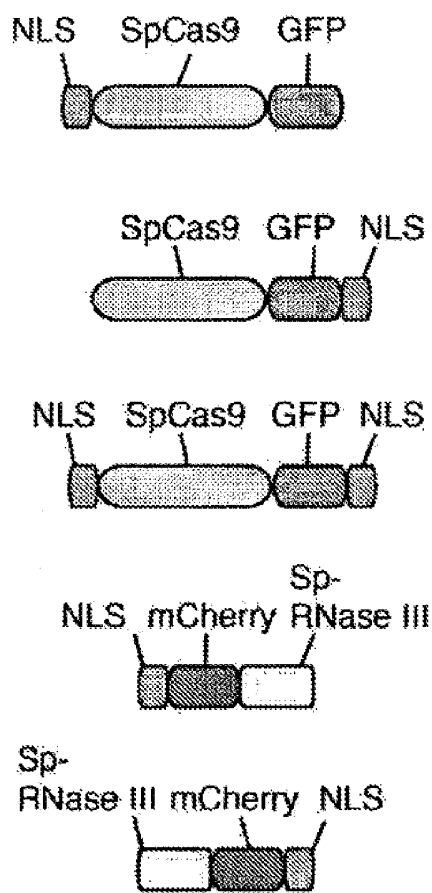
B**FIG. 2B**



FIG. 2C

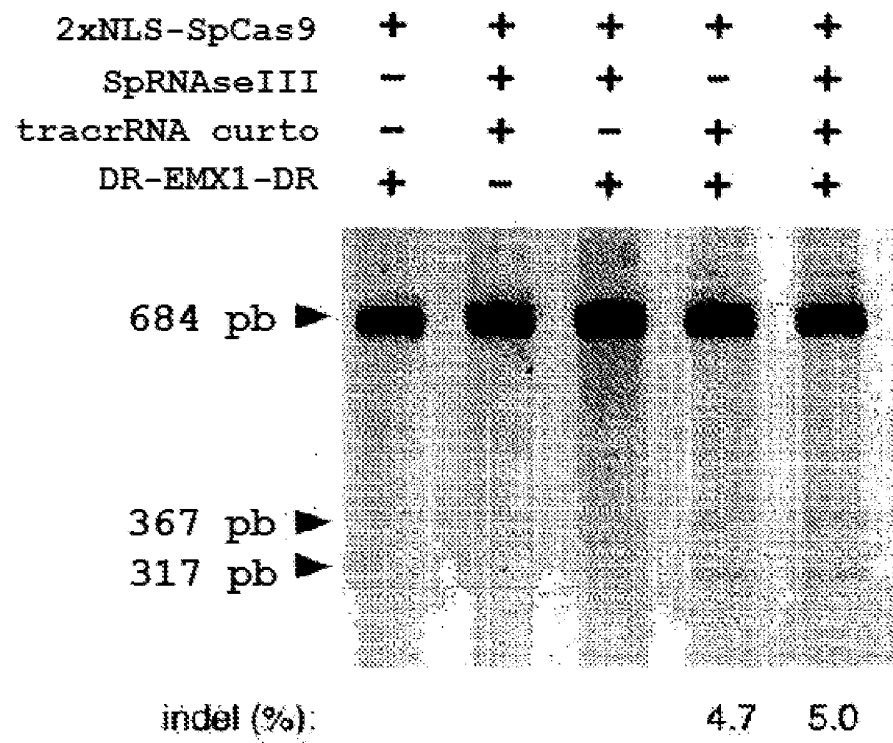
D

FIG. 2D

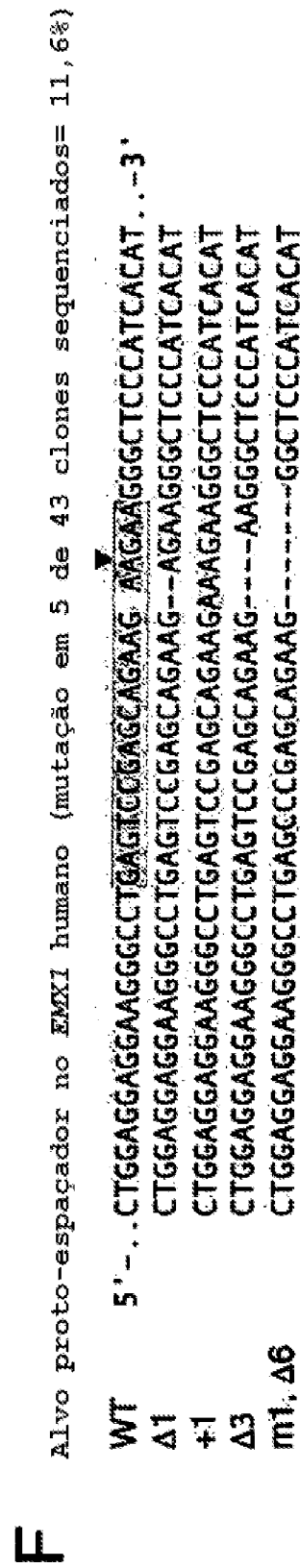
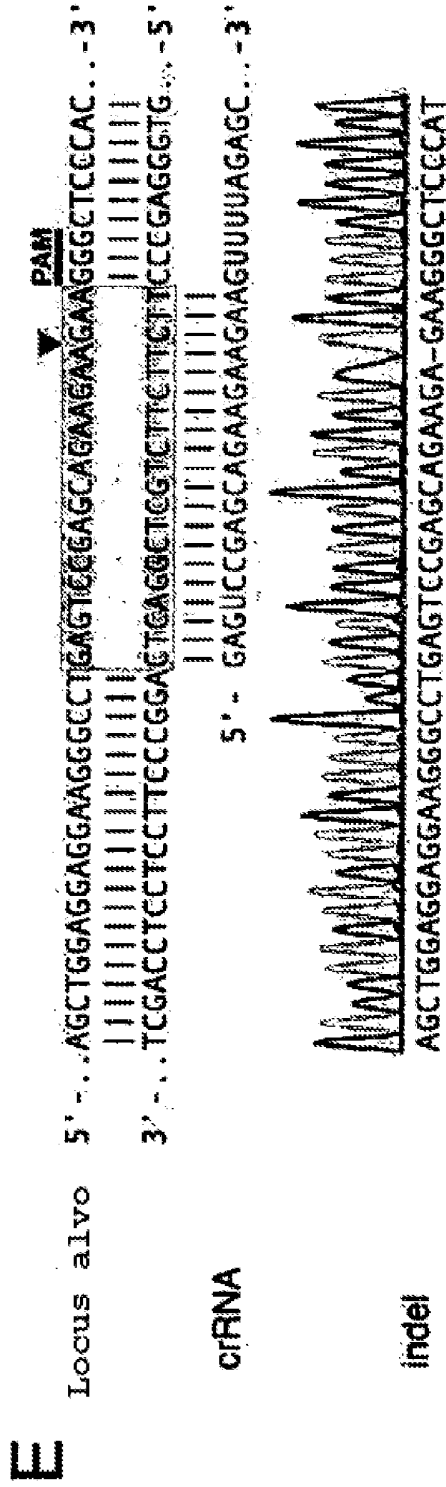


FIG. 2E-F

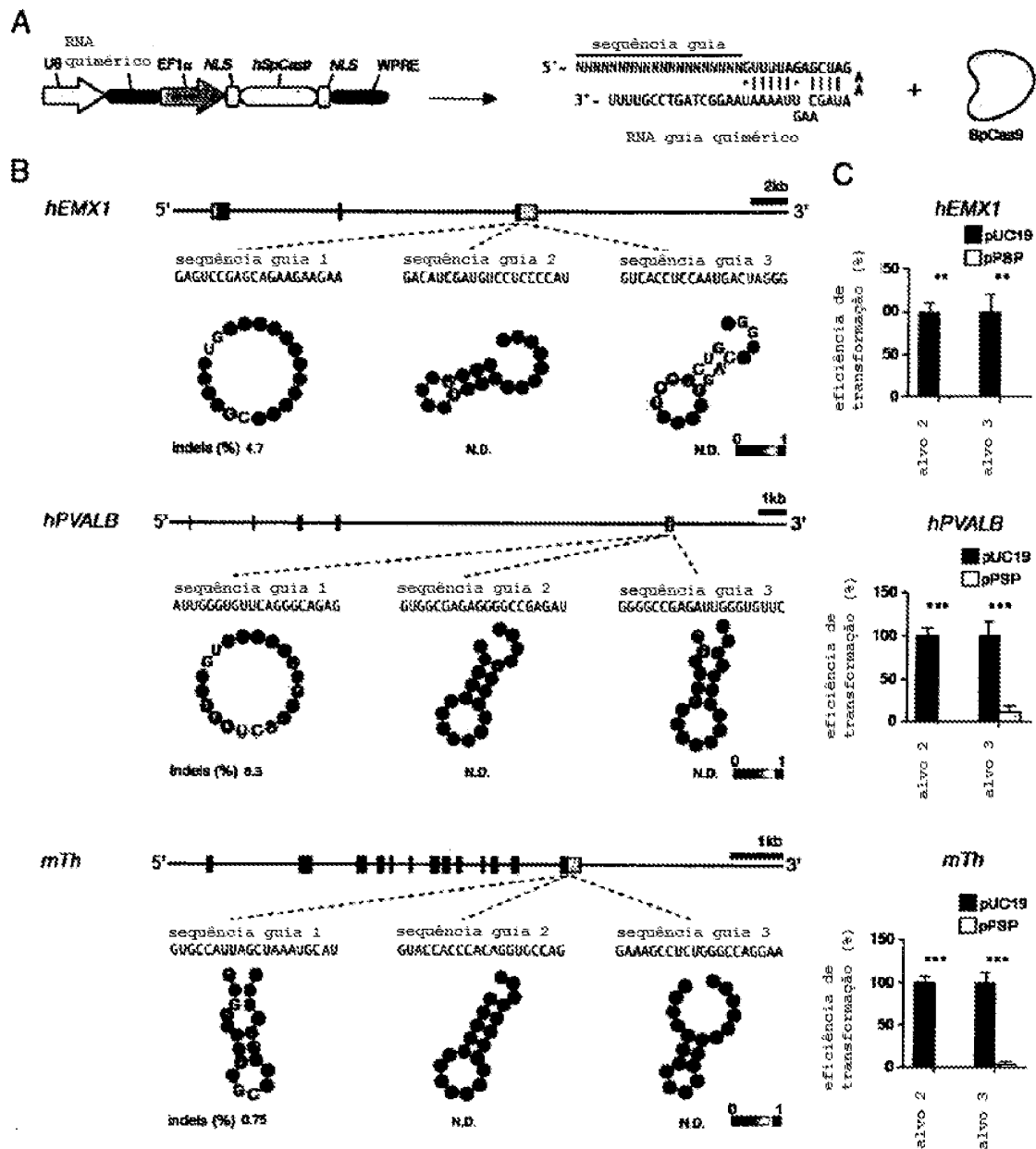


FIG. 3A-C

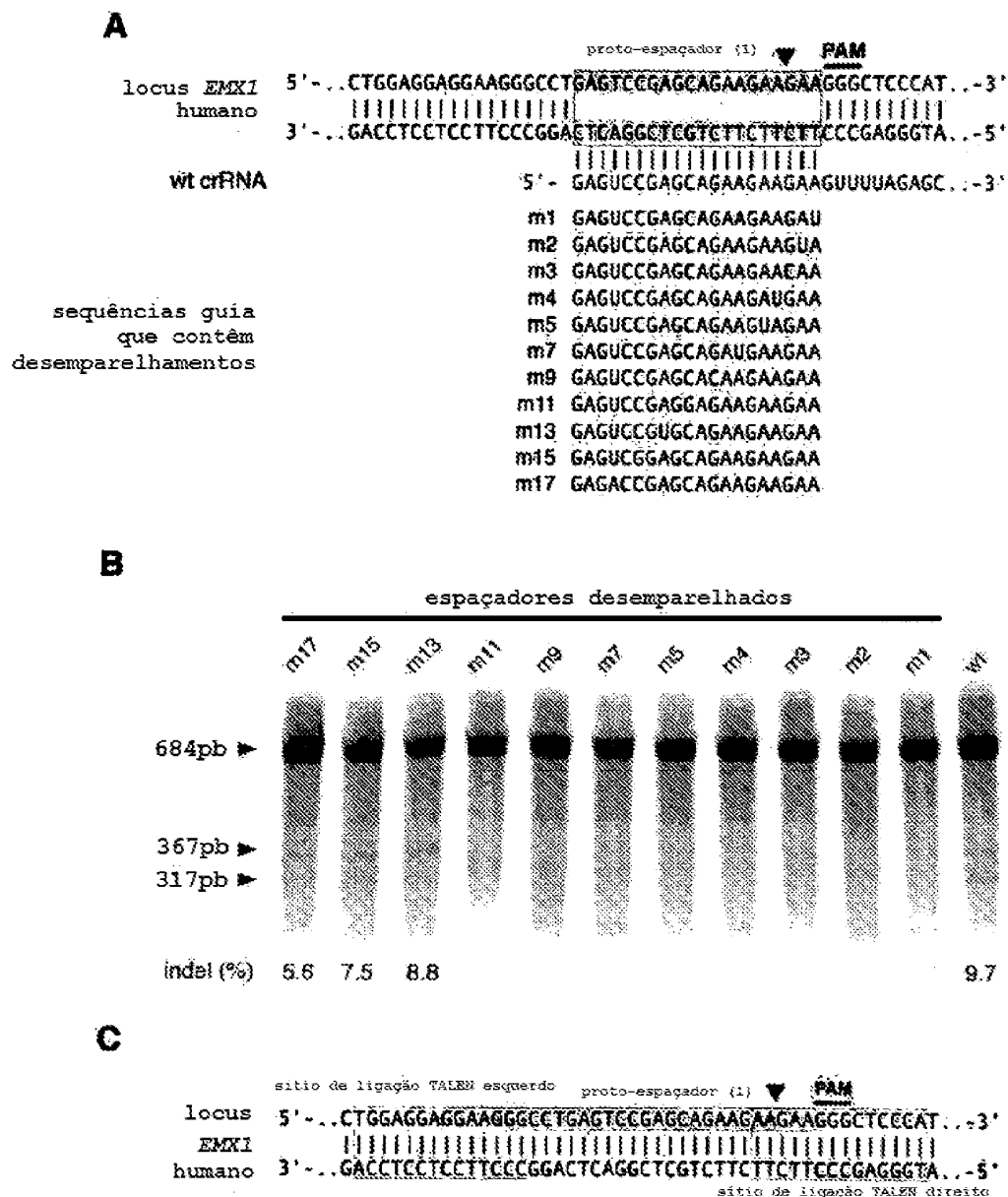


FIG. 4A-C

Deste modo, em alguns aspectos da presente divulgação, o comprimento necessário do tracrRNA numa construção da presente divulgação, por exemplo, uma construção quimérica, não tem necessariamente de ser fixo, e em alguns aspectos da divulgação, pode ser entre 40 e 120 pb, e em alguns aspectos da divulgação até ao comprimento total do tracr, por exemplo, em alguns aspectos da divulgação, até à extremidade 3' do tracr como pontuado pelo sinal de terminação da transcrição no genoma bacteriano. Em certas formas de realização, o comprimento do tracrRNA inclui pelo menos os nucleótidos 1-67 e em algumas formas de realização, pelo menos, os nucleótidos 1-85 do tracrRNA do tipo selvagem. Em algumas formas de realização, podem ser usados pelo menos os nucleótidos correspondentes aos nucleótidos 1-67 ou 1-85 do tracrRNA Cas9 de *S. pyogenes* do tipo selvagem. Quando o sistema CRISPR utiliza outras enzimas que não a Cas9, ou outras que não a SpCas9, então podem estar presentes os nucleótidos correspondentes no tracrRNA do tipo selvagem relevante. Em algumas formas de realização, o comprimento do tracrRNA inclui não mais do que 1-67 ou 1-85 nucleótidos do tracrRNA do tipo selvagem com respeito à otimização de sequência (por exemplo, redução de sequências poli-T), por exemplo, quanto a cadeias de T internos para o emparelhamento tracr (repetição direta) ou tracrRNA, em alguns aspectos da divulgação, um ou mais T presentes numa sequência poli-T da sequência de tipo selvagem relevante (ou seja, um fragmento de mais de 3, 4, 5, 6, ou mais bases T contíguas; em algumas formas de realização, um fragmento de não mais do que 10, 9, 8, 7, 6 bases T contíguas) podem ser substituídos com um nucleótido que não o T, por exemplo, um A, de modo que a cadeia é dividida em fragmentos menores de T com cada fragmento com 4, ou menos do que 4 (por exemplo, 3 ou 2) T contíguos. Se a cadeia de

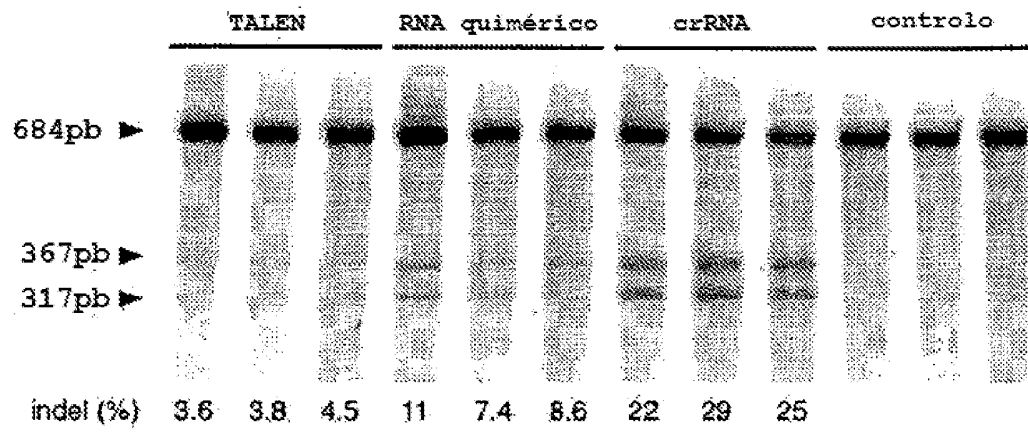
D

FIG. 4D

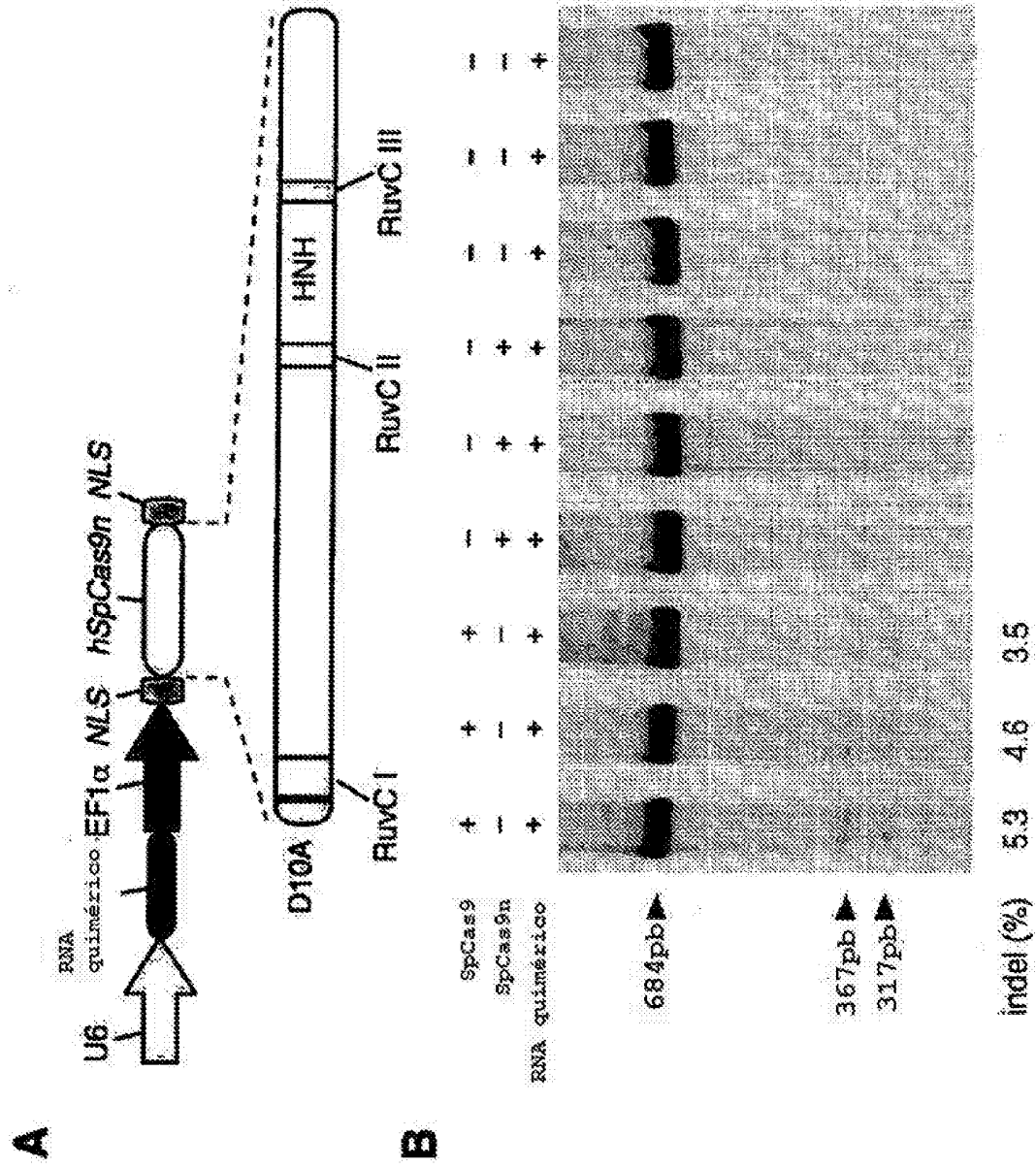


FIG. 5A-B

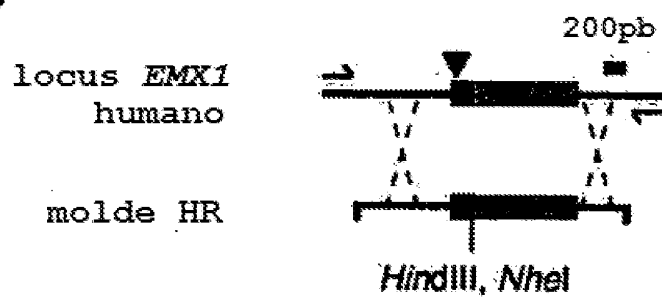
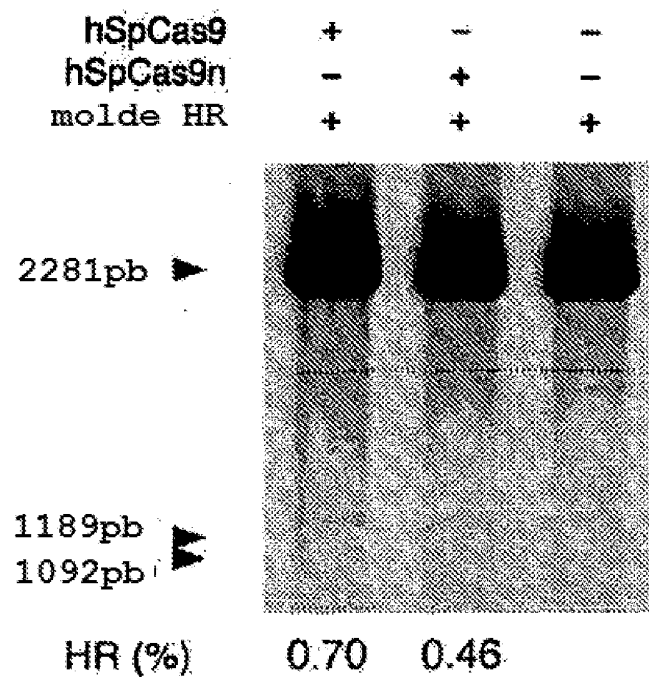
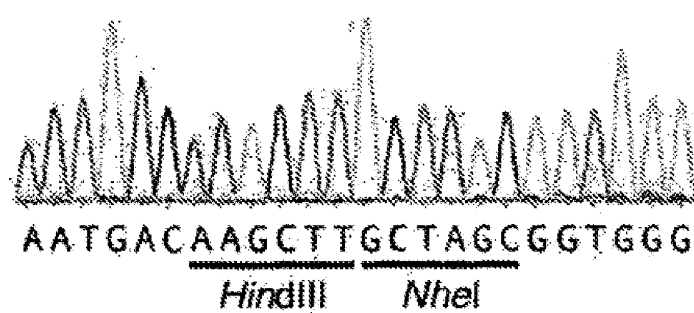
C**D****E**

FIG. 5C-E

Desenho do ay de cRNA

locus EMX1 humano

5'...AAACGGAGGGCTGAGTCCGAGCAGAAGAAGTT...AACGGAGGGAGGGGCACAGATGAGAACTCAGGGTTTTAG...-3'

5'...AGCCCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC...CAGGGAGGGAGGGGGCAGAGATGACAAACTCAGGGAGGCCCC...-3'

3'...TCGGGAGAGAGAGAGAGAGGAGGCTGAGTCCGGGAGGAGG...GTCTCTCCCTCCCGGTCCTACTCTTTGAGTCTCTGGGGG...-5'

PAM

proto-espaçador (8)

junção da deleção

118pb

resultado deleção

FIG. 5G

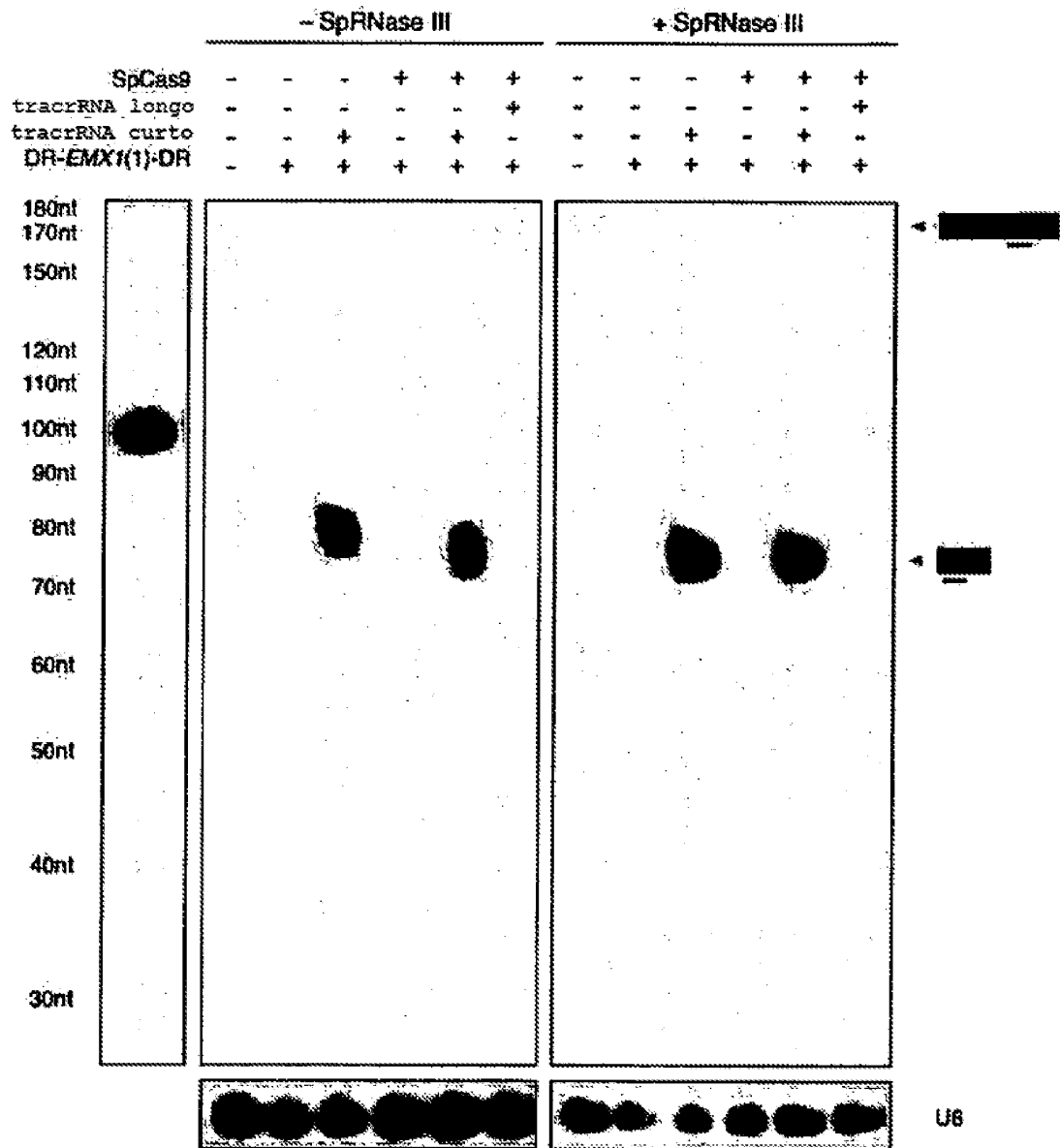
C

FIG. 6C

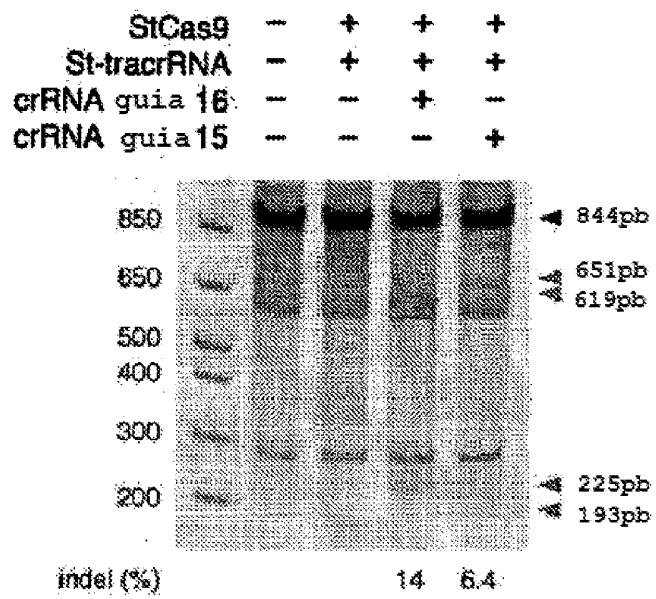


FIG. 7D

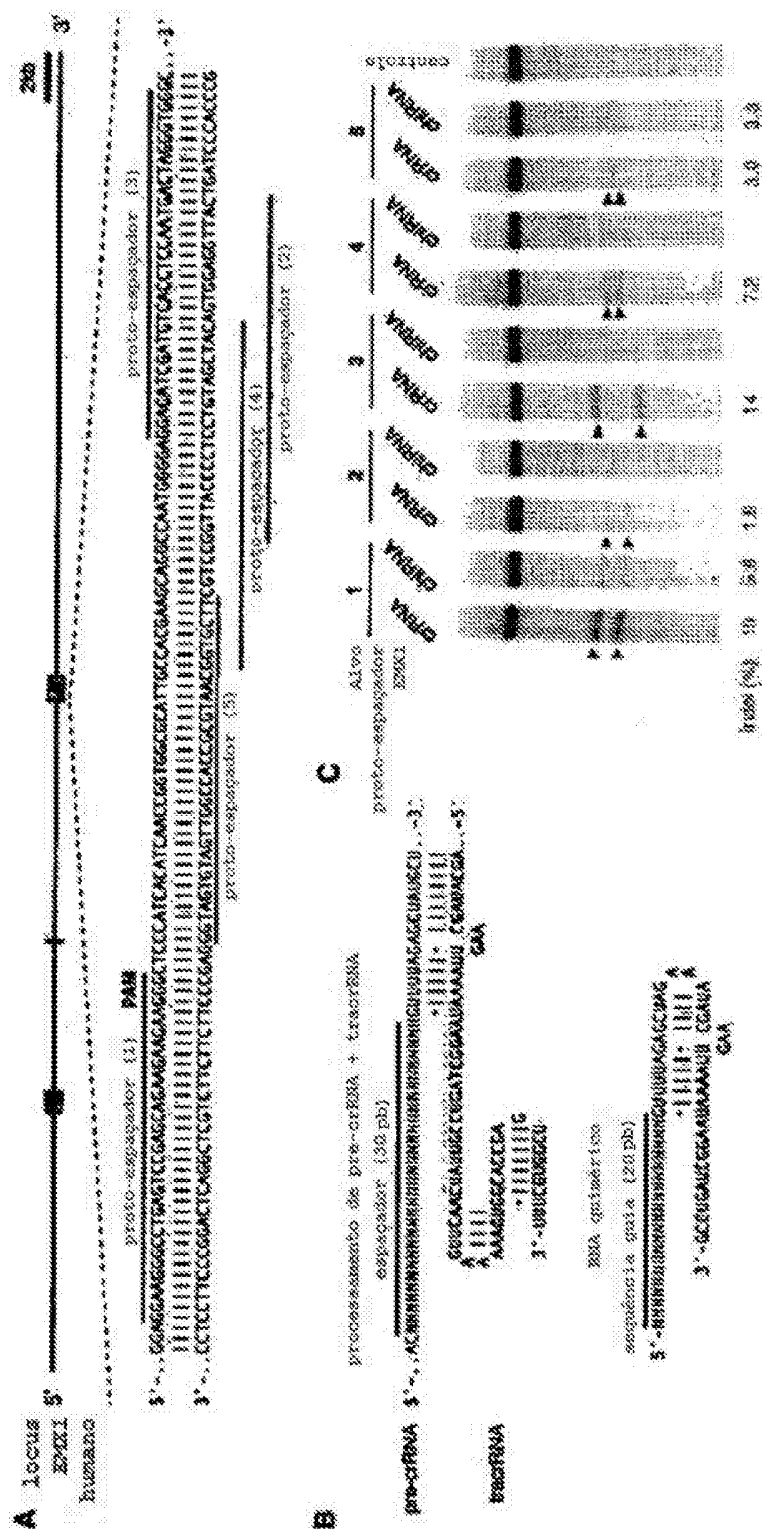


FIG. 8A-C

T está envolvida na formação de uma forquilha (ou alça de haste), então é vantajoso que a base complementar para a base que não é T seja alterada para complementar o nucleótido que não é T. Por exemplo, se a base que não é T é um A, então o seu complemento pode ser mudado para um T, por exemplo, para preservar ou ajudar na preservação da estrutura secundária. Por exemplo, 5'-TTTTT pode ser alterado para se tornar em 5'-TTTAT e o complementar 5'-AAAAA pode ser alterado para 5'-ATAAA. Em relação à presença de sequências de terminação poli-T nos transcritos tracr + de emparelhamento tracr, por exemplo, um terminador poli-T (TTTTT ou mais), em alguns aspectos da divulgação é vantajoso que tais sejam adicionadas ao extremo do transcrito, quer seja na forma de dois RNA (tracr e de emparelhamento tracr) ou de RNA única guia. No que diz respeito a alças e forquilhas nos transcritos tracr e de emparelhamento tracr, em alguns aspectos da divulgação é vantajoso que um mínimo de duas forquilhas estejam presentes no RNA guia quimérico. Uma primeira forquilha pode ser a forquilha formada pelo complemento entre a sequência tracr e de emparelhamento tracr (repetição direta). Uma segunda forquilha pode ser na extremidade 3' da sequência tracrRNA, e isto pode fornecer uma estrutura secundária para a interação com a Cas9. Podem ser adicionadas forquilhas adicionais à extremidade 3' do RNA guia, por exemplo, em alguns aspectos da divulgação para aumentar a estabilidade do RNA guia. Além disso, a extremidade 5' do RNA guia, em alguns aspectos da divulgação, pode ser prolongada. Em alguns aspectos da divulgação, pode-se considerar 20 pb na extremidade 5' como uma sequência guia. A zona 5' pode ser prorrogada. Uma ou mais forquilhas pode ser fornecida na zona 5', por exemplo, em alguns aspectos da divulgação, isto pode também melhorar

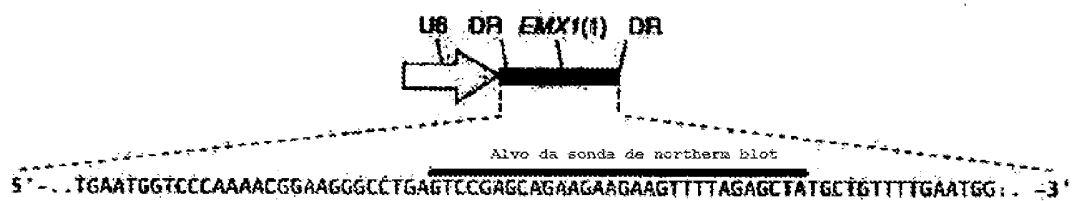
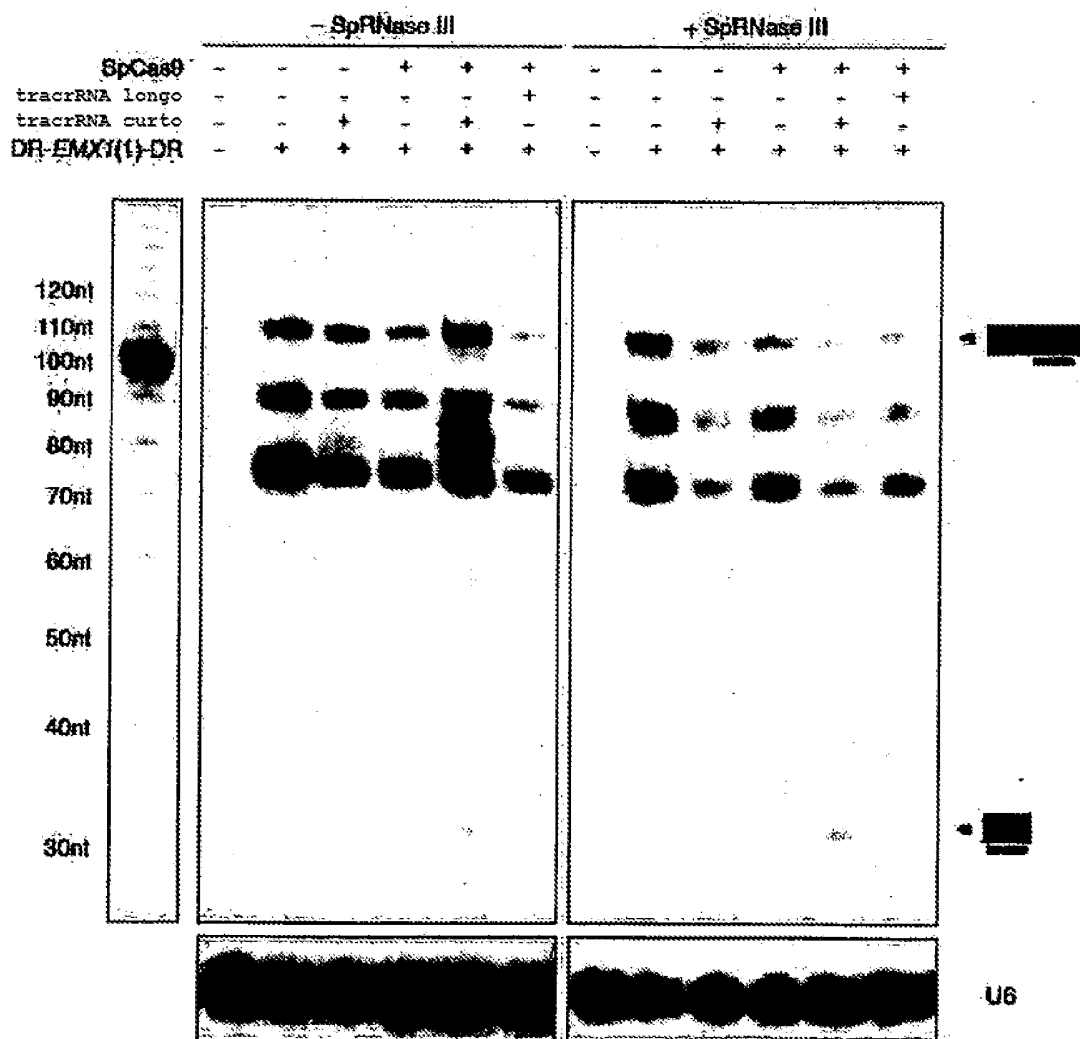
A**B**

FIG. 9A-B

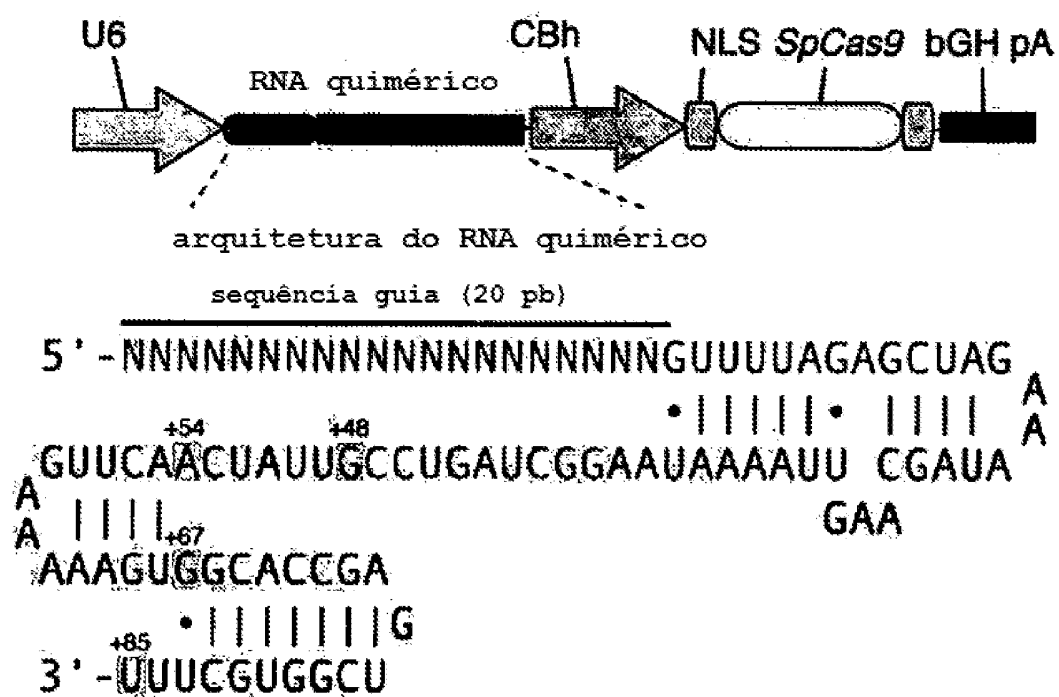
a

FIG. 10A

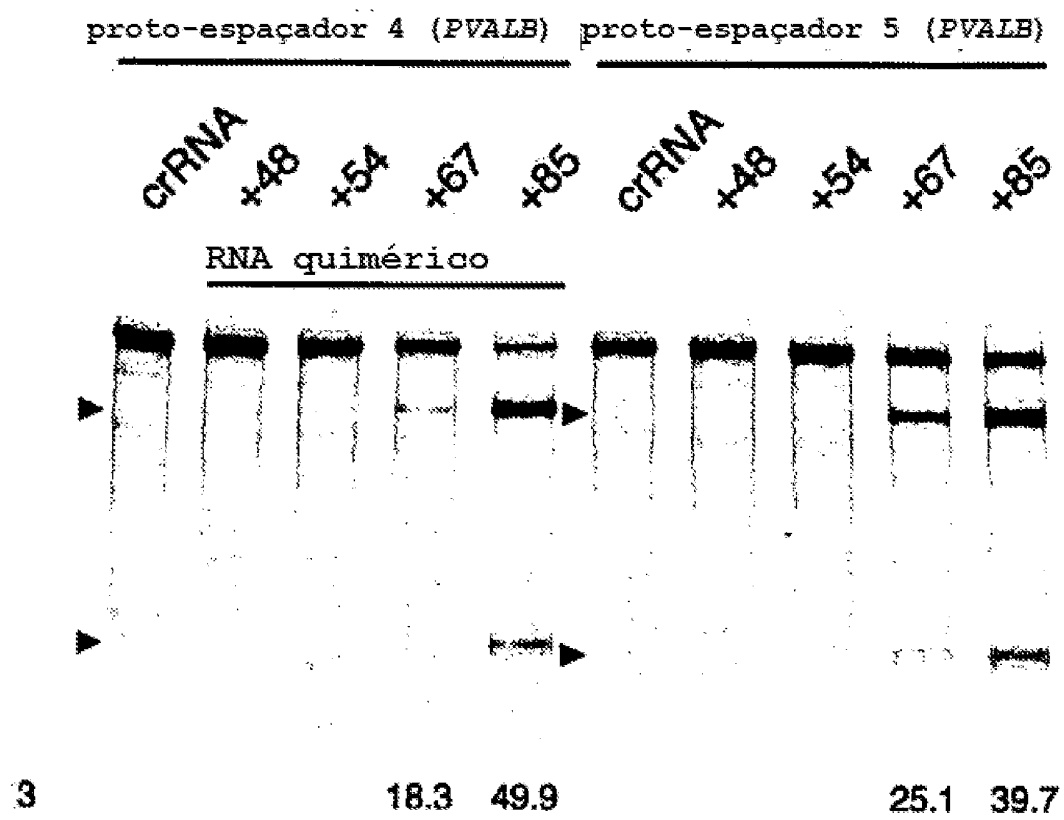
C

FIG. 10C

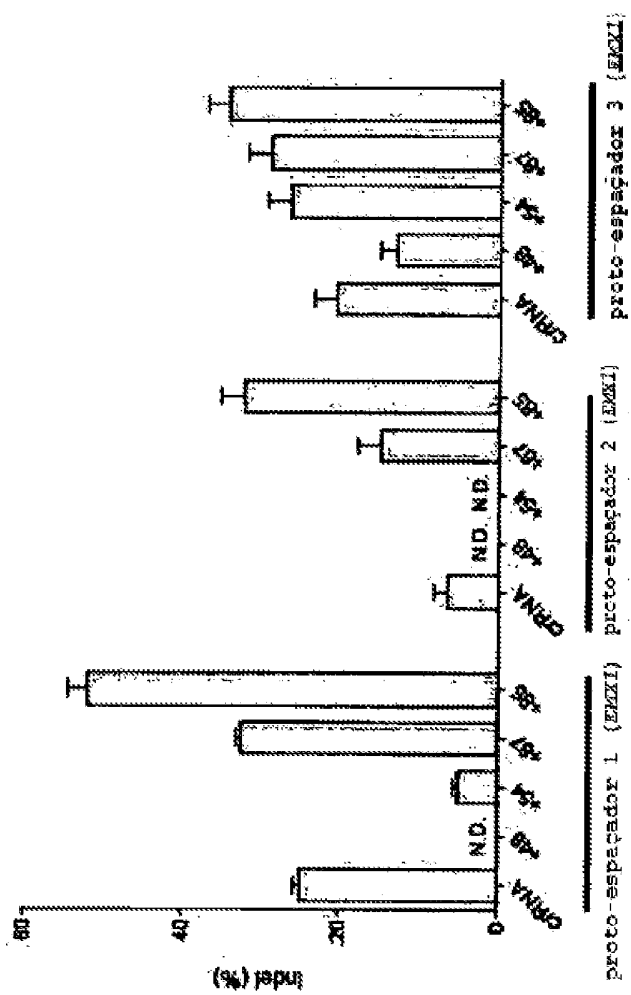
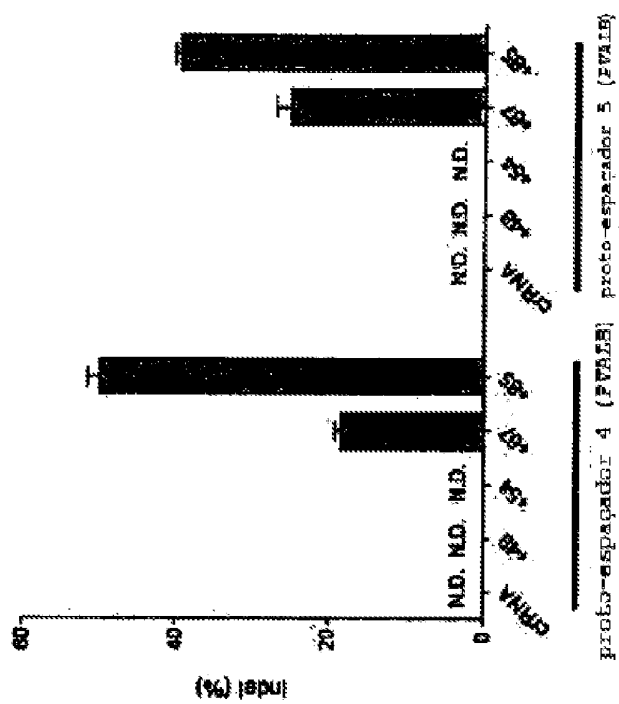
a**b**

FIG. 11A-B

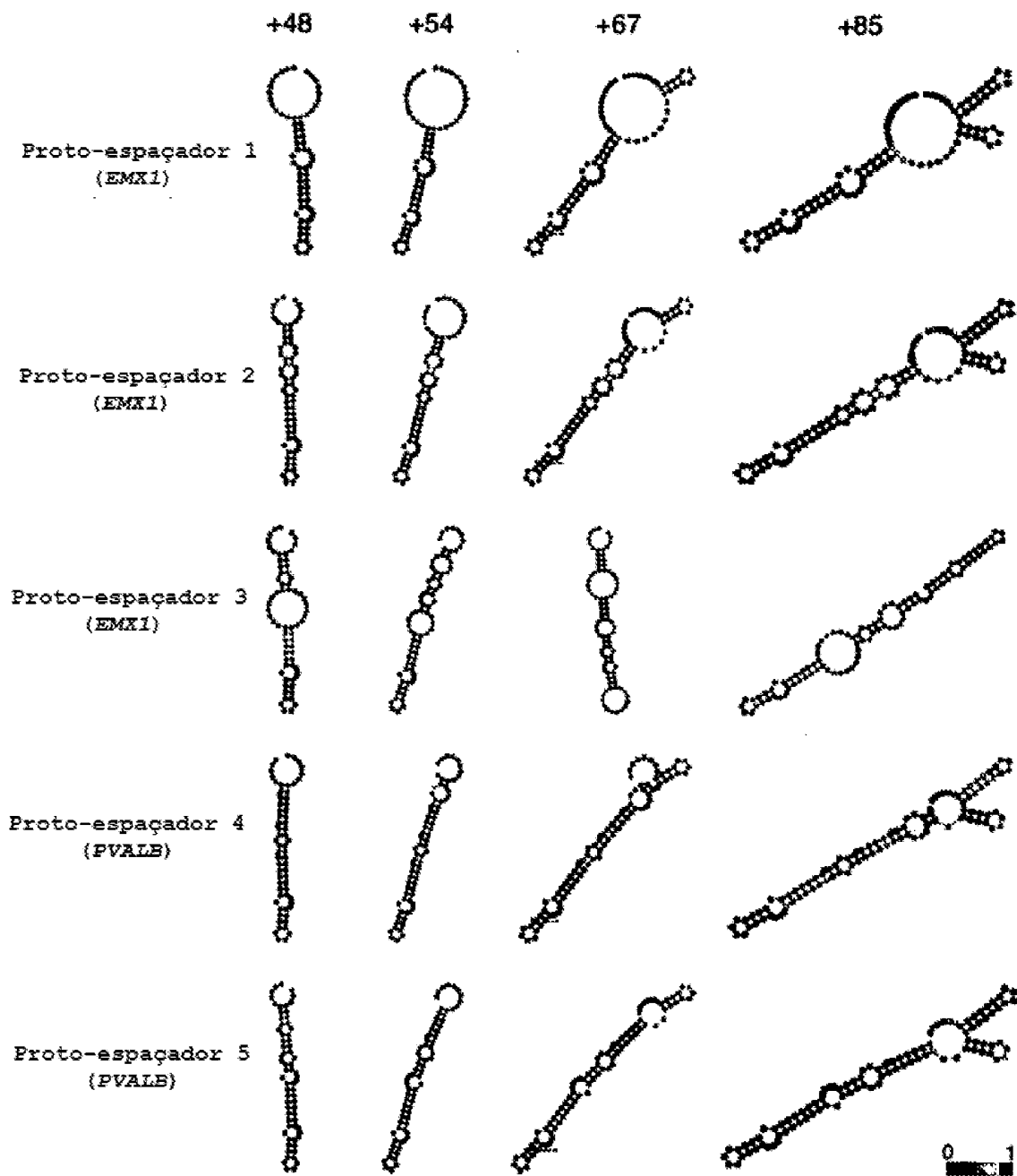


FIG. 12

Hypothetical protein - Proteína hipotética
CRISPR-associated protein, Csn1 family - Proteína associada a CRISPR, Família Csn1
CRISPR-associated endonuclease Csn1 family protein- Proteína Família Csn1,
endonuclease associada a CRISPR
Csn1-like CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype
Rs-D17] - Proteína associada a CRISPR de tipo Csn1 [bactérias não cultivadas do
fototipo Rs-D17 de Térmitas do grupo 1]
Csn1 family CRISPR-associated protein - Família Csn1 da proteína associada a CRISPR
Csn1-like CRISPR-associated protein - Proteína associada a CRISPR de tipo Csn1
Conserved hypothetical protein - Proteína hipotética conservada
Putative CRISPR-associated (Cas) protein - Proteína putativa associada a CRISPR
(Cas)
Exopolyphosphatase - Exopolifosfatase (PPX)
Restriction endonuclease - Endonuclease de restrição
Putative HNH endonuclease domain protein - Proteína de domínio putativo
endonucleases HNH
McrA/HNH-like CRISPR-associated nuclease Csn1 - Csn1 nuclease associada a CRISPR do
tipo McrA/HNH
HNH endonuclease - Endonuclease HNH

CRISPR-associated protein - Proteína asociada a CRISPR
 CRISPR-associated protein, Csn1 family - Proteína asociada a CRISPR,
 Família Csn1
 CRISPR-protein - Proteína CRISPR
 CRISPR-associated CasE family protein - Família de proteínas CasE associada
 a CRISPR
 CRISPR-associated protein Cas9/Csn1, subtype II/Nmemei - Proteína asociada
 a CRISPR Cas9/Csn1, subtipo II/Nmemei

Putative cytoplasmic protein - Proteína citoplasmática putativa
cytoplasmic protein - Proteína citoplasmática
CRISPR-Associated Protein Csn1 - Csn1 Proteína Associada a CRISPR
CRISPR-system-like protein - Proteína tipo sistema CRISPR
CRISPR associated protein - Proteína associada CRISPR
Possible CRISPR associated protein - Possível Proteína associada CRISPR

CRISPR-system related protein - Proteína relacionada com o sistema CRISPR
CRISPR-system family protein - Família de proteínas do sistema CRISPR
Hypothetical protein MoviS_00710 - Proteína MoviS_00710 hipotética
Membrane protein - Proteína de membrana
CRISPR-associated large protein (provisional), putative - Proteína grande associada a CRISPR (provisória), putativa
Putative CRISPR-associated protein, Csx12 family - Proteína putativa associada a CRISPR, família Csx12
CRISPR-associated protein, Csx12 family - Proteína associada a CRISPR, família Csx12

a estabilidade do RNA guia. Em alguns aspectos da divulgação, pode ser fornecida uma forquilha específica adicionando a sequência (5'-AGGACGAAGTCCTAA) à extremidade 5' da sequência guia, e, em alguns aspectos da divulgação, isto pode ajudar a melhorar a estabilidade. Outras sequências adequadas para a formação de forquilhas serão conhecidas dos peritos na técnica, e podem ser usadas em certos aspectos da divulgação. Em alguns aspectos da divulgação, são previstas pelo menos 2, 3, 4, 5, ou mais forquilhas adicionais. Em alguns aspectos da divulgação, são previstas não mais do que 10, 9, 8, 7, 6 forquilhas adicionais. O referido anteriormente também fornece os aspectos da presente divulgação que envolvem a estrutura secundária em sequências guia. Em alguns aspectos da divulgação, pode haver reticulação e outras modificações, por exemplo, para melhorar a estabilidade. As modificações podem incluir a inclusão de pelo menos um nucleótido não ocorre naturalmente, ou um nucleótido modificado, ou seus análogos. Os nucleótidos modificados podem ser modificado na região da ribose, fosfato, e/ou base. Os nucleótidos modificados podem incluir análogos 2'-O-metilo, análogos 2'-desoxi, ou análogos 2'-fluoro. A base dos ácidos nucleicos pode ser modificada, por exemplo, usando uma base de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos unidos por ponte (BNA) pode também ser possível. Outros exemplos de bases modificadas incluem, mas não estão limitados a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina. Tais modificações ou reticulação podem estar presentes na sequência guia ou outras sequências adjacentes à sequência guia.



FIG. 14A

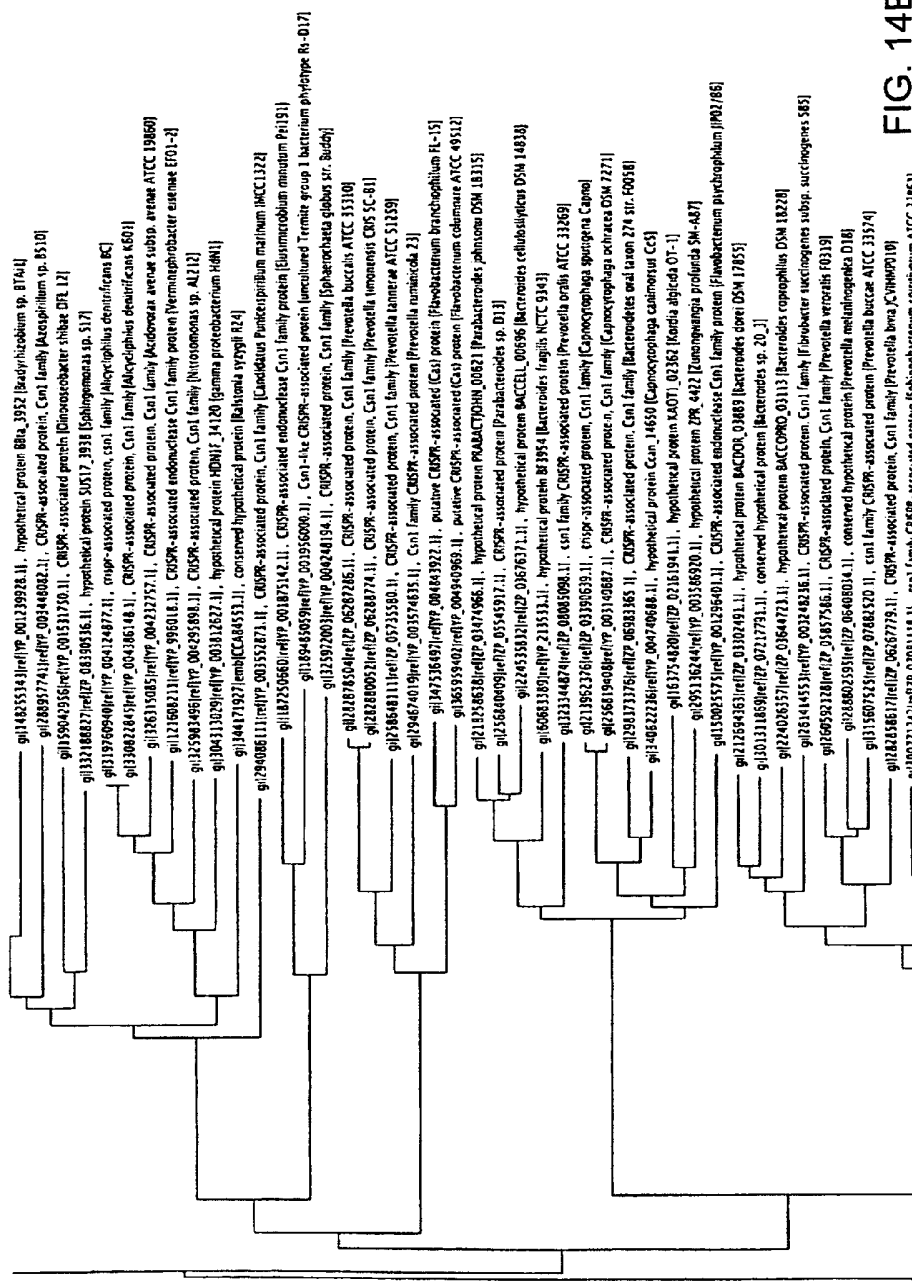


FIG. 14B

putative CRISPR-associated (Cas) protein - Proteína putativa (Cas)
 asociada a CRISPR

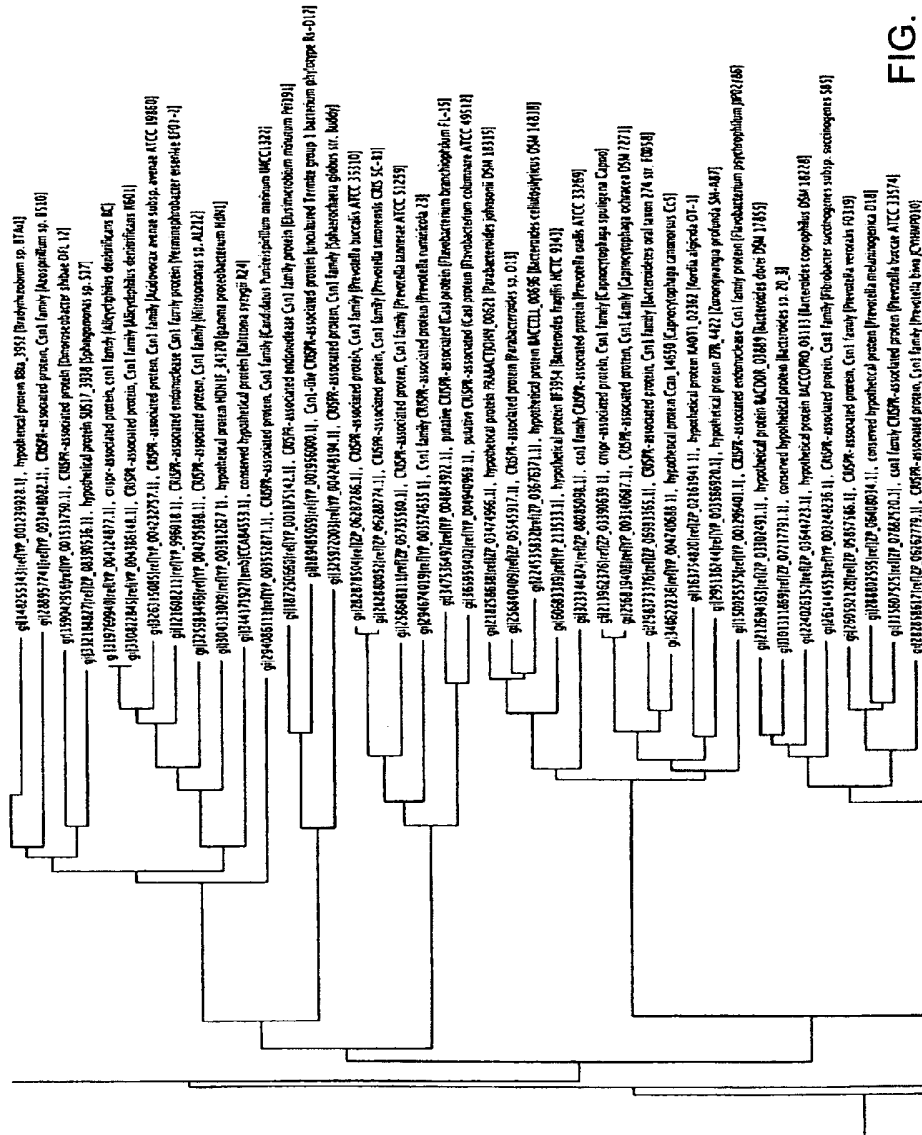


FIG. 14C

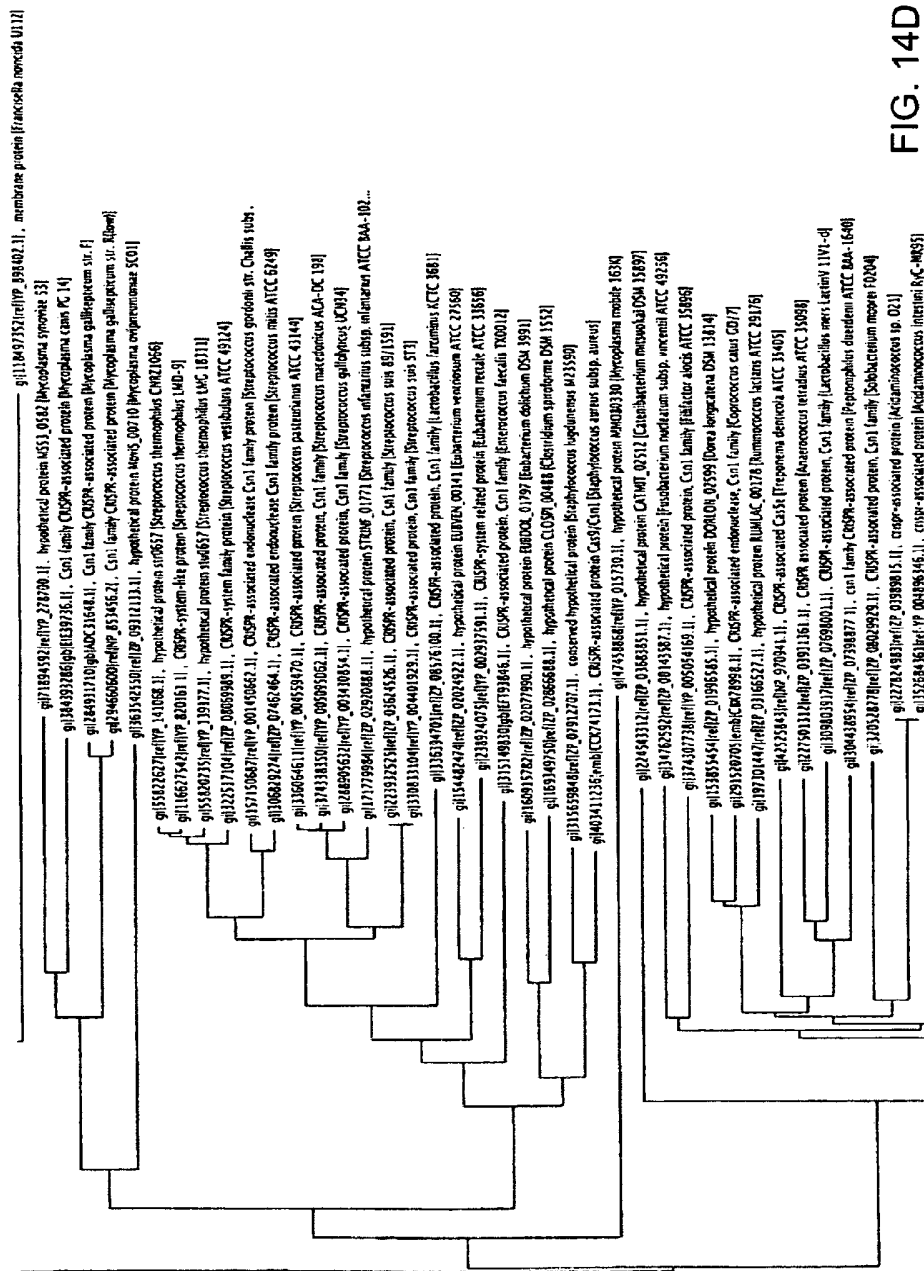


FIG. 14D

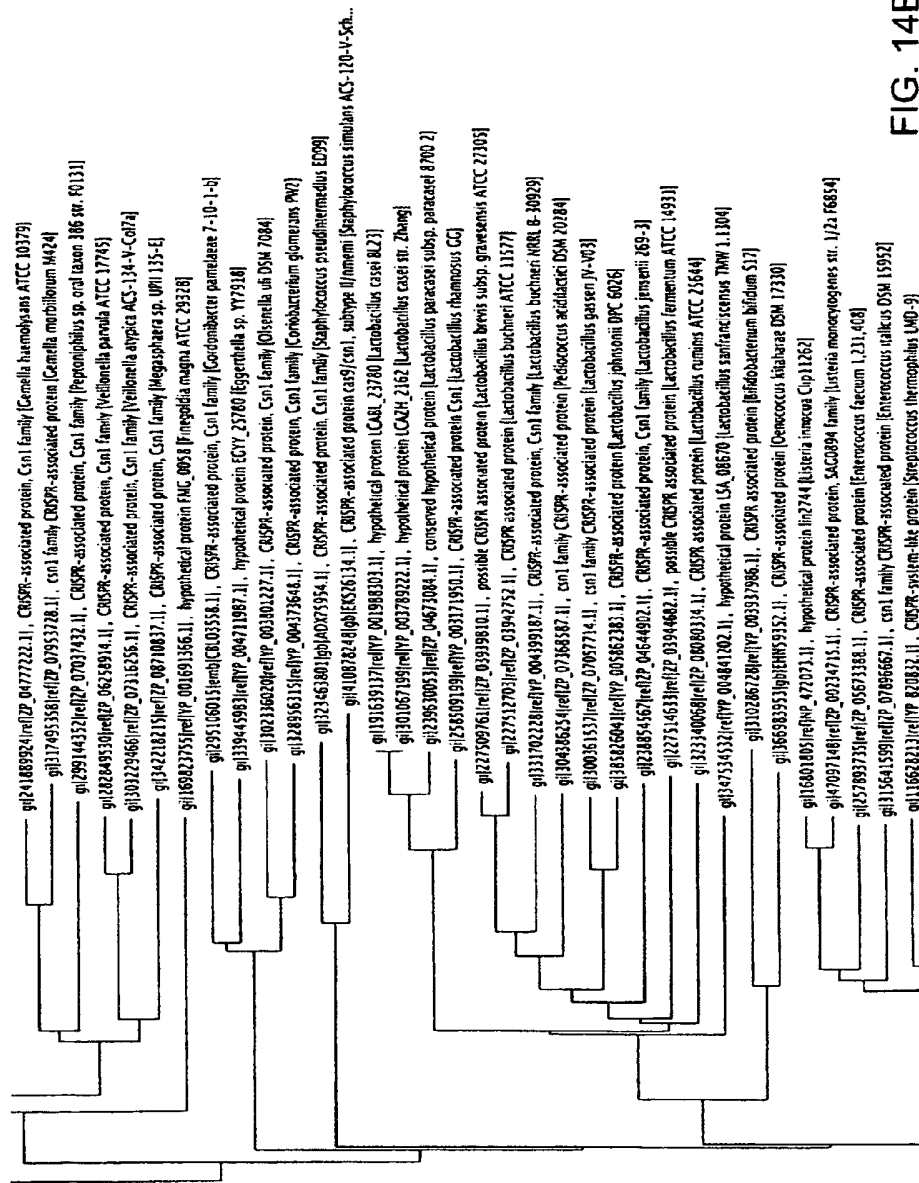


FIG. 14E

CRISPR-associated protein, SAG0894 family - Proteína asociada a CRISPR, familia SAG0894

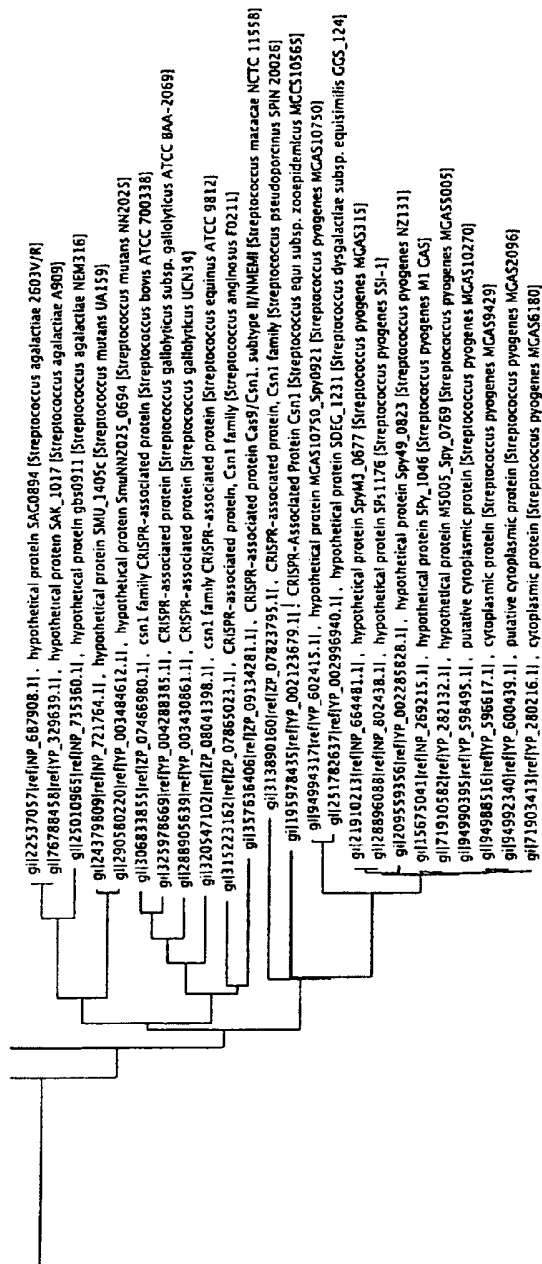


FIG. 14F

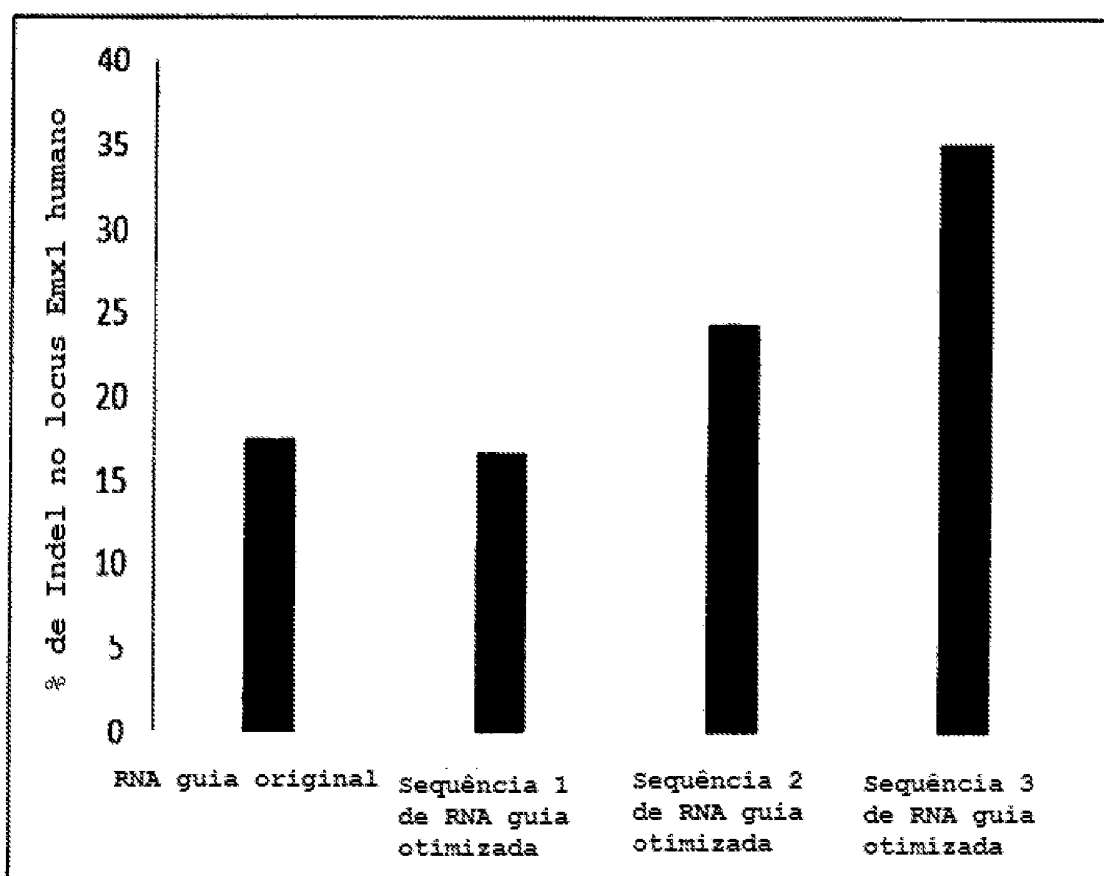
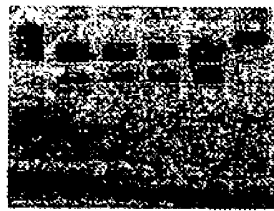


FIG. 15

FIG. 16

Ensaio com
StlCas9

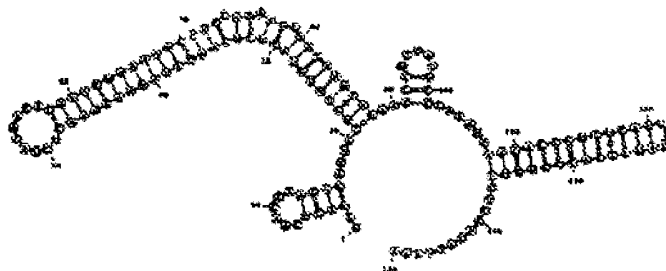
Marcador 1 2 3 4 5



Linha 1: tracrRNA total + espaçador_DR total

tracrRNA total: GUUACUUAUAUCUUSCAGAAGCUACAAGAUAAAGGCUACAUGCCGAAAUCAACACCCUGUCALUUUAUGGCAGGUGUUAUUUGUUAU

espaçador_DR total: GGGACUCACCAAGUCALUUGGUUUUGUACUCUCAAGAUUAAGUAAUGUACAAC



Linha 2: tracrRNA maduro + spacer_DR maduro

tracrRNA maduro: CUUSCAGAAGCUACAAGAUAAAGGCUACAUGCCGAAAUCAACACCCUGUCALUUUAUGGCAGGUGUUAU

spacer_DR maduro: GGGACUCACCAAGUCALUUGGUUUUGUACUCUCAAGAUUAAGUAAUGUACAAC

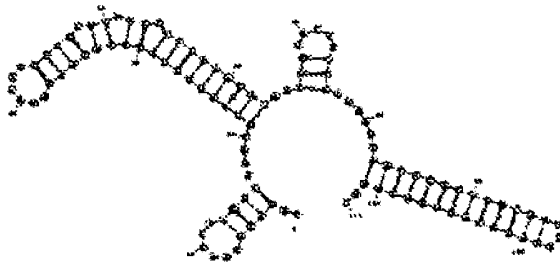


FIG. 18A

Descrição Breve das Figuras

As novas funcionalidades da presente divulgação são apresentadas com detalhe nas reivindicações anexas. Uma melhor compreensão das características e vantagens da presente divulgação será obtida através da referência à seguinte descrição detalhada que apresenta formas de realização ilustrativas, em que os princípios da divulgação são utilizados, e os desenhos que acompanham, das quais:

Figura 1 mostra um modelo esquemático do sistema CRISPR. A nuclease Cas9 de *Streptococcus pyogenes* (amarelo) está direcionada para o DNA genómico através de um RNA guia sintético (sgRNA) que consiste de uma sequência guia de 20 nt (azul) e um suporte (vermelho). Os pares de bases da sequência guia com o DNA alvo (azul), diretamente a montante de um motivo adjacente proto-espaçador 5'-NGG necessário (PAM; magenta), e Cas9 medeia uma quebra de cadeia dupla (DSB) ~ 3 pb a montante do PAM (triângulo vermelho).

Figura 2A-F ilustra exemplificativamente um sistema CRISPR, um possível mecanismo de ação, um exemplo de uma adaptação para a expressão em células eucarióticas, e os resultados dos ensaios de avaliação de localização nuclear e atividade CRISPR.

Figura 3A-C ilustra exemplificativamente uma cassete de expressão para a expressão de elementos do sistema CRISPR em células eucarióticas, previsão de estruturas de sequências guia exemplo, e a atividade do sistema CRISPR como medida em células eucarióticas e procarióticas.

FIG. 18B

a

não metilado
 GTGGTGTACAGCTCGTCGTTTGG
 A AAC
 tratado
 com M.SssI Me Me Me
 GTGGTGTACAGCTCGTCGTTTGG
 A

não metilado
 TCCAGTCTATTAATTGTTGCCGG
 A
 tratado
 com M.SssI Me
 TCCAGTCTATTAATTGTTGCCGG
 A

Sítio alvo 1

Sítio alvo 2

b

metilação: - +		- +		- +	
sgRNA 1		sgRNA 2			

FIG. 19A-B

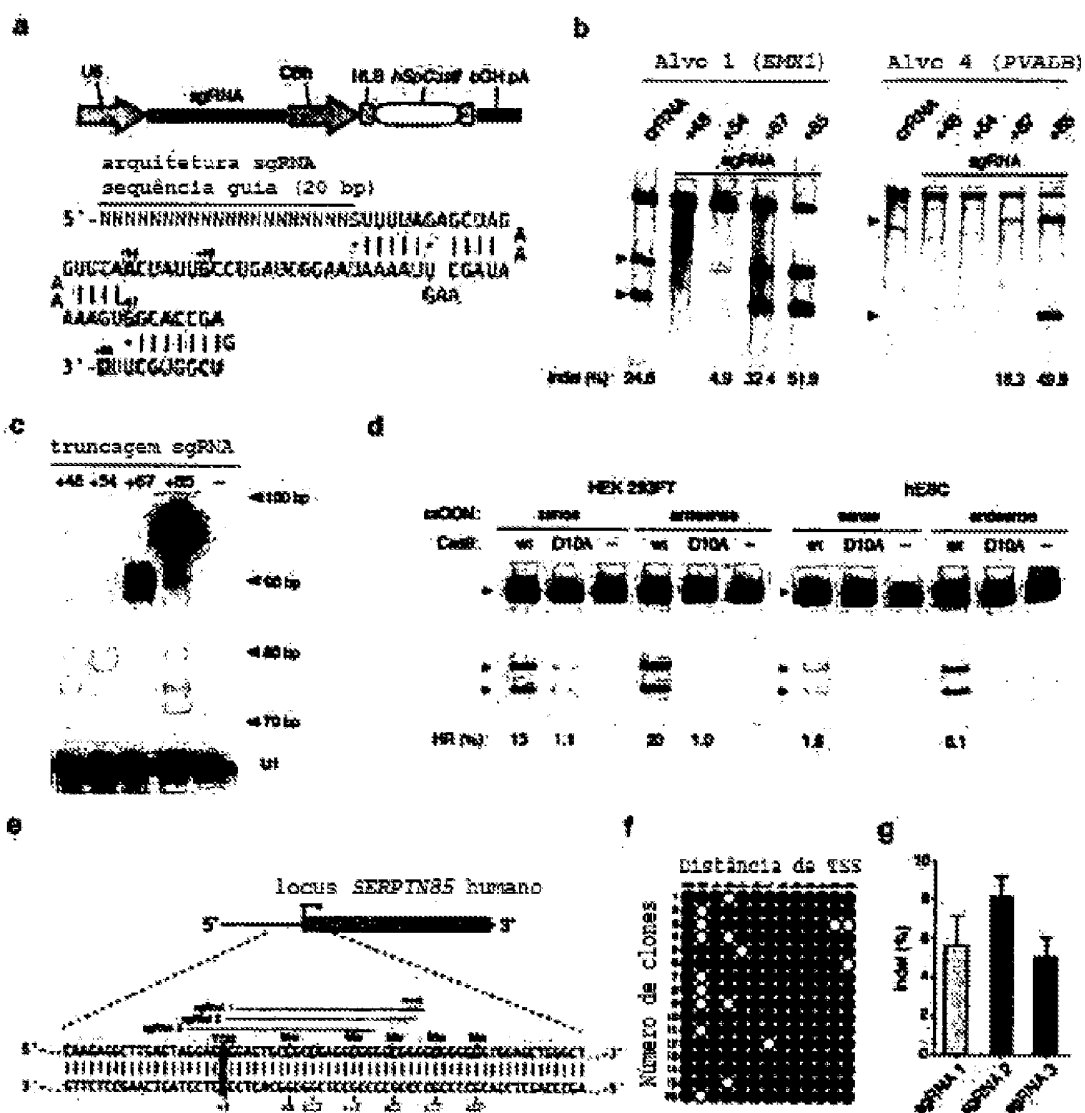


FIG. 20A-G

05/22/2025 - 6:31 PM

[illegible]

99A-G-1634-1635

2011-12-15

Figure 1. The study area.

SUUCANCUUUGCCUEBGGABARAUU COASACGACABACA

॥ श्रीगणेशाय नमः ॥

1. The first part of the document is a list of references. The references are listed in a vertical column on the left side of the page. The references are:

- 1. The first part of the document is a list of references. The references are listed in a vertical column on the left side of the page. The references are:

573

[illegible][illegible][illegible]

62-111747-1

पुष्पकविः

Figure 1

3. **Urgency**

RNA químico

U6

EMX1	EMX2	EMX1	EMX2	EMX1	EMX1
alvo 1	alvo 2	alvo 3	alvo 1	alvo 2	alvo 3

desenho
30R17A



Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100																																																																																																																																																																																								
Population (millions)	11.0	11.2	11.4	11.6	11.8	12.0	12.2	12.4	12.6	12.8	13.0	13.2	13.4	13.6	13.8	14.0	14.2	14.4	14.6	14.8	15.0	15.2	15.4	15.6	15.8	16.0	16.2	16.4	16.6	16.8	17.0	17.2	17.4	17.6	17.8	18.0	18.2	18.4	18.6	18.8	19.0	19.2	19.4	19.6	19.8	20.0	20.2	20.4	20.6	20.8	21.0	21.2	21.4	21.6	21.8	22.0	22.2	22.4	22.6	22.8	23.0	23.2	23.4	23.6	23.8	24.0	24.2	24.4	24.6	24.8	25.0	25.2	25.4	25.6	25.8	26.0	26.2	26.4	26.6	26.8	27.0	27.2	27.4	27.6	27.8	28.0	28.2	28.4	28.6	28.8	29.0	29.2	29.4	29.6	29.8	30.0	30.2	30.4	30.6	30.8	31.0	31.2	31.4	31.6	31.8	32.0	32.2	32.4	32.6	32.8	33.0	33.2	33.4	33.6	33.8	34.0	34.2	34.4	34.6	34.8	35.0	35.2	35.4	35.6	35.8	36.0	36.2	36.4	36.6	36.8	37.0	37.2	37.4	37.6	37.8	38.0	38.2	38.4	38.6	38.8	39.0	39.2	39.4	39.6	39.8	40.0	40.2	40.4	40.6	40.8	41.0	41.2	41.4	41.6	41.8	42.0	42.2	42.4	42.6	42.8	43.0	43.2	43.4	43.6	43.8	44.0	44.2	44.4	44.6	44.8	45.0	45.2	45.4	45.6	45.8	46.0	46.2	46.4	46.6	46.8	47.0	47.2	47.4	47.6	47.8	48.0	48.2	48.4	48.6	48.8	49.0	49.2	49.4	49.6	49.8	50.0	50.2	50.4	50.6	50.8	51.0	51.2	51.4	51.6	51.8	52.0	52.2	52.4	52.6	52.8	53.0	53.2	53.4	53.6	53.8	54.0	54.2	54.4	54.6	54.8	55.0	55.2	55.4	55.6	55.8	56.0	56.2	56.4	56.6	56.8	57.0	57.2	57.4	57.6	57.8	58.0	58.2	58.4	58.6	58.8	59.0	59.2	59.4	59.6	59.8	60.0	60.2	60.4	60.6	60.8	61.0	61.2	61.4	61.6	61.8	62.0	62.2	62.4	62.6	62.8	63.0	63.2	63.4	63.6	63.8	64.0	64.2	64.4	64.6	64.8	65.0	65.2	65.4	65.6	65.8	66.0	66.2	66.4	66.6	66.8	67.0	67.2	67.4	67.6	67.8	68.0	68.2	68.4	68.6	68.8	69.0	69.2	69.4	69.6	6

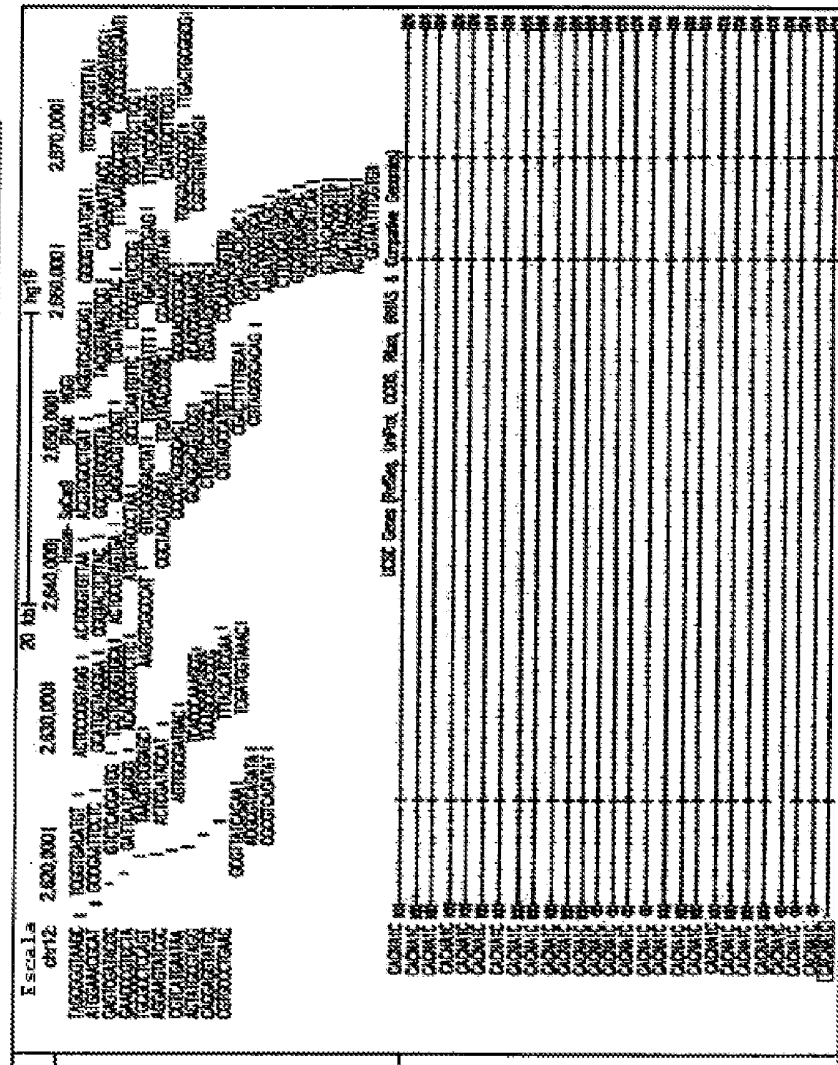
FIG. 21A-B

MOVE

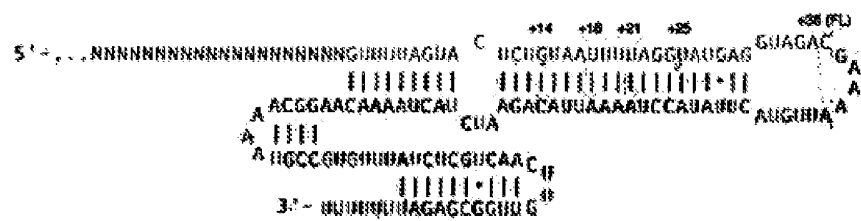
chrt2.2614.008-008-2675,748 61,741 bp.

MICROBIUM POSITIVO. SIMBOLIZADO JUNE OU JANEIRO DE PROCELA

chrt2 (p11.3)



A

**B**

alvo 1

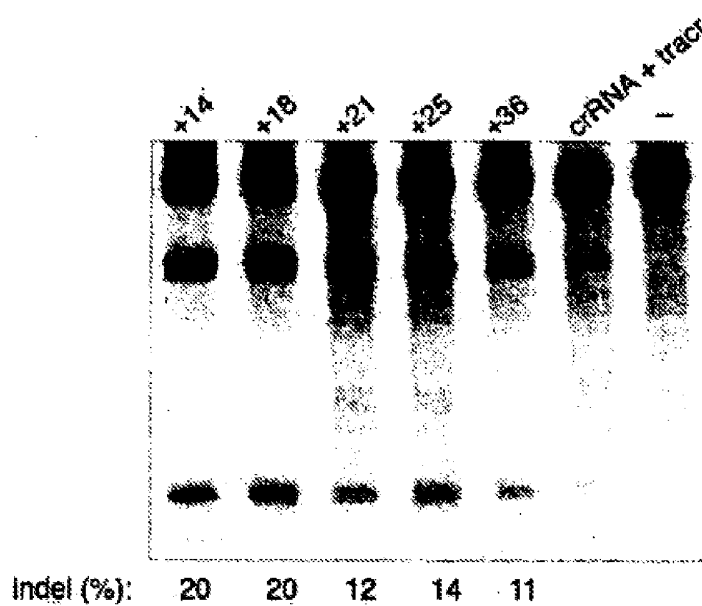


FIG. 23A-B

Figura 4A-D mostra os resultados de uma avaliação da especificidade SpCas9 para um exemplo de alvo.

Figura 5A-G ilustra exemplificativamente um sistema de vetor e resultados para a sua utilização em dirigir a recombinação homóloga em células eucarióticas.

Figura 6A-C ilustra uma comparação de diferentes transcritos tracrRNA para orientação de genes mediada por Cas9.

Figura 7A-D ilustra exemplificativamente um sistema CRISPR, um exemplo de adaptação para a expressão em células eucarióticas, e os resultados de testes que avaliam a atividade CRISPR.

Figura 8A-C ilustra exemplificativamente as manipulações de um sistema CRISPR para direcionamento de loci genômicos em células de mamífero.

Figura 9A-B ilustra os resultados de uma análise de transferência de Northern do processamento de crRNA em células de mamífero.

Figura 10A-C ilustra uma representação esquemática de RNA quiméricos e os resultados dos ensaios SURVEYOR para a atividade do sistema CRISPR em células eucarióticas.

Figura 11A-B ilustra uma representação gráfica dos resultados SURVEYOR de ensaios para a atividade do sistema CRISPR em células eucarióticas.

Figura 12 ilustra exemplificativamente as estruturas secundárias previstas para os RNA quiméricos compreendendo uma sequência guia, sequência de emparelhamento tracr, e sequência tracr.

Figura 13 é uma árvore filogenética de genes Cas.

Figura 14A-F mostra a análise filogenética revelando cinco famílias Cas9, incluindo três grupos de Cas9 grandes (~ 1400 aminoácidos) e dois grupos de Cas9 pequenas (~ 1100 aminoácidos).

Figura 15 mostra um gráfico que representa a função de diferentes RNA guia otimizados.

Figura 16 mostra a sequência e estrutura de diferentes RNA guia quiméricos.

Figura 17 mostra a estrutura de co-dobra do tracrRNA e repetição direta.

Figura 18 A e B mostra dados da otimização RNA guia quimérico St1Cas9 *in vitro*.

Figura 19A-B apresenta a clivagem de alvos não metilados ou metilados por um lisado celular de SpCas9.

Figura 20A-G mostra a otimização da arquitetura do RNA guia para edição do genoma de mamíferos mediada por SpCas9. (a) Esquema de um vetor de expressão bicistrônico (PX330) para um RNA guia individual dirigido pelo promotor U6 (sgRNA) e Cas9 de *Streptococcus pyogenes* com codão otimizado para humano dirigida pelo promotor CBh (hSpCas9) usados para

todas as experiências subsequentes. O sgRNA consiste numa sequência guia de 20 nt (azul) e suporte (vermelho), truncada em várias posições, tal como indicado. (b) Ensaio SURVEYOR para indels (inserções-deleções) mediadas por SpCas9 no loci *EMX1* e *PVALB* humanos. As setas indicam os fragmentos esperados de SURVEYOR ($n = 3$). (c) Análise de transferência de Northern para as quatro arquiteturas de truncagem sgRNA, com U1 como controlo da quantidade aplicada (*loading control*). (d) Tanto o tipo selvagem (wt) como o mutante da *nickase* (enzima que corta uma cadeia) (D10A) de SpCas9 promoveram a inserção de um local *HindIII* no gene *EMX1* humano. Os oligonucleótidos de cadeia simples (ssODNs), orientados quer na direção sense ou anti-sense em relação à sequência do genoma, foram utilizados como moldes de recombinação homóloga. (e) Representação esquemática do locus *SERPINB5* humano. sgRNAs e PAMs são indicados por barras coloridas acima sequência; metilcitosina (Me) são destacadas (rosa) e numeradas relativamente ao local de iniciação da transcrição (TSS, +1). (f) Estado de metilação do *SERPINB5* em 16 clones determinado por tratamento com bissulfito e posterior sequenciação. Círculos preenchidos, CpG metilado; círculos abertos, CpG não metilado. (g) Eficiência da modificação por três sgRNAs tendo como alvo a região metilada do *SERPINB5*, determinada por sequenciação profunda ($n = 2$). As barras de erro indicam intervalos Wilson (Métodos On-line).

Figura 21A-B apresenta a otimização adicional da arquitetura do sgRNA CRISPR-Cas. (a) Esquema das quatro arquiteturas adicionais do sgRNA, I-IV. Cada uma consiste numa sequência guia de 20 nt (azul) unida à repetição direta (DR, cinzento), que hibrida com o tracrRNA (vermelho). O híbrido DR-tracrRNA é truncado nas posições

+12 ou +22, conforme indicado, com uma alça de haste artificial GAAA. As posições de truncagem no tracrRNA são numeradas de acordo com o local de início da transcrição previamente descrito para o tracrRNA. As arquiteturas II e IV do sgRNA apresentam mutações dentro das suas regiões poli-U, o que pode servir como terminadores de transcrição precoce. (b) Ensaio SURVEYOR para indels mediadas por SpCas9 no locus *EMX1* humano para os locais alvo 1-3. As setas indicam os fragmentos SURVEYOR esperados ($n = 3$).

Figura 22 ilustra a visualização de alguns locais alvo no genoma humano.

Figura 23A-B Apresenta (A) um esquema do sgRNA e (B) a análise SURVEYOR de cinco variantes de sgRNA para SaCas9 para uma arquitetura truncada ideal com uma eficiência de clivagem mais elevada.

As figuras neste documento são apenas para fins ilustrativos e não são necessariamente desenhadas à escala.

Descrição Detalhada da Invenção

Os termos "polinucleótido", "nucleótido", "sequência de nucleótidos", "ácido nucleico" e "oligonucleótido" são utilizados alternadamente. Eles referem-se a uma forma polimérica de nucleótidos de qualquer comprimento, quer ribonucleótidos ou desoxirribonucleótidos, ou os seus análogos. Os polinucleótidos podem ter qualquer estrutura tridimensional, e podem desempenhar qualquer função, conhecida ou desconhecida. Os seguintes são exemplos não limitativos de polinucleótidos: regiões codantes e não-codantes de um gene ou fragmentos de gene, loci (locus)

definidos a partir de análise de ligação (ou linkage), exões, intrões, RNA mensageiro (mRNA), RNA de transferência, RNA ribossômico, RNA de interferência curto (siRNA), RNA de forquilha curta (shRNA, *short hairpin RNA*), microRNA (miRNA), ribozimas, cDNA, polinucleótidos recombinantes, polinucleótidos ramificados, plasmídeos, vetores, DNA isolado de qualquer sequência, RNA isolado de qualquer sequência, sondas de ácido nucleico, e primers. Um polinucleótido pode compreender um ou mais nucleótidos modificados, tais como nucleótidos metilados e análogos de nucleótidos. Se presentes, as modificações da estrutura de nucleótidos podem ser transmitidas antes ou após montagem do polímero. A sequência de nucleótidos pode ser interrompida por componentes não nucleótidos. Um polinucleótido pode ser adicionalmente modificado após polimerização, tal como por conjugação com um componente de marcação.

De acordo com aspetos da invenção os termos "RNA quimérico", "RNA guia quimérico", "RNA guia", "RNA guia individual" e "RNA guia sintético" são utilizadas indistintamente e referem-se a sequência de polinucleótido que compreende a sequência guia, a sequência tracr e a sequência de emparelhamento tracr. O termo "sequência guia" refere-se à sequência de cerca de 20 pb dentro do RNA guia que especifica o local alvo e podem ser utilizados de forma intercambiável o termo "guia" ou "espaçador". O termo "sequência de emparelhamento tracr" também pode ser utilizado intermutavelmente com o termo "repetição ou repetições diretas".

Como usado na presente divulgação, o termo "tipo selvagem" é um termo da técnica que é compreendido pelos peritos e

significa a forma típica de um organismo, estirpe, gene ou uma característica tal como ocorre na natureza distinta de formas mutantes ou variantes.

Como usado na presente divulgação, o termo "variante" deve ser utilizado para fazer referência à apresentação de qualidades que têm um padrão que se desvia do que ocorre na natureza.

Os termos de "origem não natural" ou "modificado por engenharia" são utilizados indistintamente e indicam o envolvimento da mão do homem. Os termos, quando se referem a moléculas de ácidos nucleicos ou polipéptidos significam que a molécula de ácido nucleico ou o polipéptido é, pelo menos, substancialmente livre de, pelo menos, um outro componente com o qual eles são naturalmente associados na natureza e como se encontram na natureza.

"Complementaridade" refere-se à capacidade de um ácido nucleico para formar ligação ou ligações de hidrogénio com outra sequência de ácido nucleico quer pelo emparelhamento de bases Watson-Crick tradicional ou por outros tipos não tradicionais. A percentagem de complementaridade indica a percentagem de resíduos de uma molécula de ácido nucleico que podem formar pontes de hidrogénio (por exemplo, emparelhamento de bases de Watson-Crick) com uma segunda sequência de ácido nucleico (por exemplo, 5, 6, 7, 8, 9, 10 de 10 tendo uma complementaridade de 50%, 60%, 70%, 80%, 90%, e 100%). "Perfeitamente complementar" significa que todos os resíduos contíguos de uma sequência de ácido nucleico irão estabelecer ligações de hidrogénio com o mesmo número de resíduos contíguos de uma segunda sequência de ácido nucleico. "Substancialmente complementar", tal

como usado na presente divulgação, refere-se a um grau de complementaridade, que é, pelo menos, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, ou 100% em relação a uma região de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, ou mais nucleótidos, ou refere-se a dois ácidos nucleicos que hibridam sob condições rigorosas.

Tal como usado na presente divulgação, "condições rigorosas" para hibridação referem-se a condições sob as quais um ácido nucleico possuindo complementaridade com uma sequência alvo predominantemente hibrida com a sequência alvo, e substancialmente não hibrida com a sequência não alvo. As condições rigorosas são geralmente dependentes da sequência, e variam dependendo de um número de fatores. Em geral, quanto mais longa a sequência, maior é a temperatura à qual a sequência hibrida especificamente com a sua sequência alvo. Exemplos não limitativos de condições rigorosas são descritos em detalhe em Tijssen (1993), *Laboratory Techniques in Biochemistry and Molecular Biology- Hybridization With Nucleic Acid Probes Part I*, Capítulo II "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, Nova Iorque.

"Hibridização" refere-se a uma reação em que um ou mais polinucleótidos reagem para formar um complexo que é estabilizado através de ligações de hidrogénio entre as bases dos resíduos de nucleótidos. A ligação de hidrogénio pode ocorrer por emparelhamento de base de Watson Crick, ligação Hoogsteen, ou de qualquer outra forma específica de sequência. O complexo pode compreender duas cadeias que formam uma estrutura dupla, três ou mais cadeias que formam

um complexo multi cadeia, uma única cadeia que auto hibrida, ou qualquer combinação destas. A reação de hibridação pode constituir um passo de um processo mais vasto, tal como a iniciação de PCR, ou a clivagem de um polinucleótido por uma enzima. Uma sequência capaz de hibridar com uma determinada sequência é designada por "complementar" da sequência dada.

Tal como usado na presente divulgação, "estabilização" ou "aumento da estabilidade" no que diz respeito aos componentes do sistema CRISPR relaciona-se com a segurança e estabilização da estrutura da molécula. Isto pode ser conseguido pela introdução de uma ou mutações, incluindo alterações de um único ou múltiplos pares de bases, o aumento do número de forquilhas, reticulação, quebra de fragmentos específicos de nucleótidos e outras modificações. As modificações podem incluir a inclusão de pelo menos um nucleótido que não ocorre naturalmente, ou um nucleótido modificado, ou um análogo dos mesmos. Os nucleótidos modificados podem ser modificados na região da ribose, do fosfato, e/ou da base. Os nucleótidos modificados podem incluir análogos 2'-O-metilo, análogos 2'-desoxi, ou análogos 2'-fluoro. A base dos ácidos nucleicos pode ser modificada, por exemplo, usando uma base de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos unidos por ponte (BNA) pode também ser possível. Outros exemplos de bases modificadas incluem, mas não estão limitados a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina. Essas modificações podem ser aplicadas a qualquer componente do sistema CRISPR. Numa forma de realização preferida, estas modificações são feitas para os

componentes de RNA, por exemplo, o RNA guia ou a sequência de polinucleótidos quimérica.

Tal como usado na presente divulgação, "expressão" refere-se ao processo pelo qual um polinucleótido é transcrito a partir de um molde de DNA (tal como em um mRNA ou outro transcrito de RNA) e/ou o processo pelo qual um mRNA transcrito é subsequentemente traduzido em péptidos, polipéptidos, ou proteínas. Os transcritos e polipéptidos codificados podem ser coletivamente referidos como "produto do gene." Se o polinucleótido é derivado do DNA genómico, a expressão pode incluir o *splicing* do mRNA numa célula eucariótica.

Os termos "polipéptido", "péptido" e "proteína" são utilizados indistintamente na presente divulgação para referência a polímeros de aminoácidos de qualquer comprimento. O polímero pode ser linear ou ramificado, pode compreender os aminoácidos modificados e pode ser interrompido por não aminoácidos. Os termos englobam também um polímero de aminoácidos que foi modificado; por exemplo, formação de ligação dissulfureto, glicosilação, lipidação, acetilação, fosforilação ou qualquer outra manipulação, tal como conjugação com um componente de marcação. Tal como utilizado na presente divulgação, o termo "aminoácido" inclui aminoácidos naturais e/ou não naturais ou sintéticos, incluindo glicina e ambos os isómeros óticos D ou L, e análogos de aminoácidos e peptidomiméticos.

Os termos "sujeito", "indivíduo" e "paciente" são aqui utilizados indistintamente para referência a um vertebrado, preferencialmente a um mamífero, mais preferencialmente a um ser humano. Os mamíferos incluem, mas não estão

limitados a, murídeos, símios, humanos, animais de criação, animais para desporto e animais de estimação. São também englobados os tecidos, células e a suas linhagens de uma entidade biológica obtida *in vivo* ou cultivada *in vitro*. Em algumas formas de realização, um sujeito pode ser um animal invertebrado, por exemplo, um inseto ou um nematode; enquanto que em outras formas de realização, um sujeito pode ser uma planta ou um fungo.

Os termos "agente terapêutico", "agente com capacidade terapêutica" ou "agente para tratamento" são utilizados indistintamente e referem-se a uma molécula ou um composto que confere um efeito benéfico depois da administração a um sujeito. O efeito benéfico inclui a capacidade de determinação de diagnóstico; melhoria de uma doença, sintomas, distúrbios ou condição patológica; reduzir ou prevenir o aparecimento de uma doença, sintomas, distúrbios ou condição; e, geralmente, contrariando uma doença, sintomas, distúrbios ou condição patológica.

Na presente divulgação, "tratamento" ou "tratar", ou "atenuar" ou "melhorar" são utilizados alternadamente. Estes termos referem-se a uma abordagem para obtenção de resultados benéficos ou desejados, que incluem, mas não limitados a um benefício terapêutico e/ou um benefício profilático. Por benefício terapêutico se faz referência a qualquer melhoria ou efeito terapeuticamente relevante sobre uma ou mais doenças, condições ou sintomas em tratamento. Para o benefício profilático, as composições podem ser administradas a um sujeito em risco de desenvolver uma doença particular, condição ou sintoma, ou a um sujeito que apresente um ou mais dos sintomas

fisiológicos de uma doença, embora a doença, condição ou sintoma pode não ter sido ainda manifestada.

O termo "quantidade eficaz" ou "quantidade terapeuticamente eficaz" refere-se à quantidade de um agente que é suficiente para obter resultados benéficos ou desejados. A quantidade terapeuticamente eficaz pode variar dependendo de um ou mais dos seguintes: do sujeito e da condição da doença a ser tratada, do peso e da idade do sujeito, a gravidade da condição da doença, do modo de administração e semelhantes, os quais podem ser prontamente determinadas por um normal perito na técnica. O termo também se aplica a uma dose que irá fornecer uma imagem para detecção por qualquer um dos métodos de imagiologia descritos na presente divulgação. A dose específica pode variar de acordo com um ou mais dos seguintes: o agente escolhido em particular, o regime de dosagem a ser seguido, se este é administrado em combinação com outros compostos, tempo de administração, o tecido a ser visualizado, e o sistema físico de administração que é utilizado.

A componente prática da presente invenção emprega, a menos que indicado de outro modo, técnicas convencionais de imunologia, de bioquímica, química, biologia molecular, microbiologia, biologia celular, genómica e de DNA recombinante, que estão dentro da perícia da técnica. Consulte Sambrook, Fritsch e Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2ª edição (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel, et al. Eds., (1987)); a série METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames e G.R. Taylor eds. (1995)), Harlow e Lane, eds. (1988) ANTIBODIES,

A LABORATORY MANUAL, e ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

Vários aspetos da presente invenção referem-se a sistemas de vetores que compreendem um ou mais vetores, ou vetores como tal. Os vetores podem ser desenhados para a expressão de transcritos CRISPR (por exemplo, transcritos de ácidos nucleicos, proteínas ou enzimas) em células procarióticas ou eucarióticas. Por exemplo, os transcritos CRISPR pode ser expressos em células bacterianas tais como de *Escherichia coli*, células de inseto (usando vetores de expressão de baculovírus), células de leveduras, ou células de mamíferos. As células hospedeiras adequadas são discutidas mais detalhe em Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego; Califórnia (1990). Como alternativa, o vetor de expressão recombinante pode ser transcrito e traduzido *in vitro*, por exemplo usando sequências reguladoras do promotor T7 e a polimerase T7.

Os vetores podem ser introduzidos e propagados num procariota. Em algumas formas de realização, um procariota é utilizado para amplificar cópias de um vetor para ser introduzido numa célula eucariótica, ou como um vetor intermédio na produção de um vetor para ser introduzido numa célula eucariótica (por exemplo, amplificação de um plasmídeo como parte de um sistema de empacotamento de vetor viral). Em algumas formas de realização, um procariota é utilizado para amplificar cópias de um vetor e expressar um ou mais ácidos nucleicos, de modo a fornecer uma fonte de uma ou mais proteínas para entrega a uma célula hospedeira ou organismo hospedeiro. A expressão de proteínas em procariotas é mais frequentemente realizada em

Escherichia coli com vetores contendo promotores constitutivos ou induzíveis que orientam a expressão de proteínas de fusão, quer ou de não fusão. Os vetores de fusão adicionam um número de aminoácidos a uma proteína nele codificado, tal como um fragmento amino-terminal da proteína recombinante. Tais vetores de fusão podem servir um ou mais fins, tais como: (i) para aumentar a expressão de proteína recombinante; (ii) para aumentar a solubilidade da proteína recombinante; e (iii) auxiliar na purificação da proteína recombinante atuando como um ligando na purificação por afinidade. Muitas vezes, em vetores de expressão de fusão, um local de clivagem proteolítica é introduzido entre a junção da região de fusão e a proteína recombinante para permitir a separação da proteína recombinante a partir da região de fusão subsequentemente à purificação da proteína de fusão. Tais enzimas, e as suas sequências de reconhecimento relacionadas, incluem Fator Xa, trombina e enteroquinase. Exemplos de vetores de expressão de fusão incluem pGEX (Pharmacia Biotech Inc; Smith e Johnson, 1988. *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) e pRIT5 (Pharmacia, Piscataway, N.J.) que fundem a glutathione S-transferase (GST), proteína de ligação à maltose E, ou proteína A, respetivamente, com a proteína alvo recombinante.

Exemplos de vetores adequados de expressão de *E. coli* de não fusão induzíveis incluem pTrc (Amrann *et al.*, (1988) *Gene* 69: 301-315) e pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

Em algumas formas de realização, um vetor é um vetor de expressão de levedura. Exemplos de vetores para expressão

na levedura *Saccharomyces cerevisiae* incluem pYepSec1 (Baldari, et al., 1987. *EMBO J.* 6:229-234), pMFa (Kuijan e Herskowitz, 1982. *Cell* 30:933-943), pJRY88 (Schultz et al., 1987. *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Califórnia), e picZ (InVitrogen Corp., San Diego, Califórnia).

Em algumas formas de realização, um vetor dirige a expressão da proteína em células de inseto usando vetores de expressão de baculovírus. Os vetores de baculovírus disponíveis para a expressão de proteínas em células de inseto em cultura (por exemplo, células SF9) incluem a série pAc (Smith, et al., 1983. *Mol. Cell. Biol.* 3:2156-2165) e a série pVL (Lucklow e Summers, 1989. *Virology* 170:31-39).

Em algumas formas de realização, um vetor é capaz de dirigir a expressão de uma ou mais sequências em células de mamífero usando um vetor de expressão de mamífero. Exemplos de vetores de expressão de mamíferos incluem pCDM8 (Seed, 1987. *Nature* 329:840) e pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6:187-195). Quando usado em células de mamífero, as funções de controlo do vetor de expressão são tipicamente fornecidas por um ou mais elementos reguladores. Por exemplo, promotores utilizados normalmente são derivados de polioma, adenovírus 2, citomegalovírus, vírus símio 40, e outros revelados na presente divulgação e conhecidos na técnica. Para outros sistemas de expressão adequados, tanto para células procarióticas e eucarióticas ver, por exemplo, capítulos 16 e 17 de Sambrook, et al., *MOLECULAR CLONING: A LABORATORY MANUAL*. 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Em algumas formas de realização, o vetor de expressão de mamífero recombinante é capaz de dirigir a expressão do ácido nucleico preferencialmente num tipo de célula em particular (por exemplo, elementos reguladores específicos de tecido são usados para expressar o ácido nucleico). Os elementos reguladores específicos de tecido são conhecidos na técnica. Exemplos não limitantes de promotores específicos adequados de tecido incluem o promotor da albumina (específico do fígado; Pinkert *et al.*, 1987. *Genes Dev.* 1:268-277), promotores específicos de linfóides (Calame e Eaton, 1988. *Adv. Immunol.* 43:235-275), em particular promotores de receptores de células T (Winoto e Baltimore, 1989, *EMBO J.* 8:729-733) e imunoglobulinas (Baneiji, *et al.*, 1983. *Cell* 33:729-740; Queen e Baltimore, 1983. *Cell* 33:741-748), promotores específicos de neurónios (por exemplo, o promotor de neurofilamentos; Byrne e Ruddle, 1989, *Proc. Natl. Acad. Sci. USA* 86:5473-5477), promotores específicos do pâncreas (Edlund, *et al.*, 1985. *Science* 230:912-916), e promotores específicos de glândula mamária (por exemplo, promotor de soro de leite; Patente U.S. No. 4,873,316 e Publicação de Pedido Europeu No. 264,166). Os promotores regulados pelo desenvolvimento também estão incluídos, por exemplo, os promotores *hox* de murino (Kessel e Gruss, 1990. *Science* 249:374-379) e o promotor de α -fetoproteína (Campes e Tilghman, 1989. *Genes Dev.* 3:537-546).

Em algumas formas de realização, um elemento regulador está ligado de forma operativa a um ou mais elementos de um sistema CRISPR, de modo a conduzir a expressão de um ou mais elementos do sistema de CRISPR. Em geral, CRISPR (Repetições Palindrômicas Curtas Agrupadas e Regularmente Espaçadas), também conhecidos como SPIDR (Repetições

Diretas de Espaçador Intercalar; do inglês, *SPacer Interspersed Direct Repeats*), constituem uma família de loci de DNA que são usualmente específicos para uma espécie bacteriana em particular. O locus CRISPR compreende uma classe distinta de repetições intercaladas de sequência curta (SSR) que foi reconhecida em *E. coli* (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; e Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), e os genes associados. Foram identificadas SSR intercaladas similares em *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, e *Mycobacterium tuberculosis* (Ver, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; e Mojica et al., Mol. Microbiol., 17:85-93 [1995]). Os loci CRISPR diferem tipicamente de outras SSR pela estrutura das repetições, que foram denominadas repetições curtas espaçadas regularmente (SRSR) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; e Mojica et al., Mol. Microbiol., 36:244-246 [2000]). Em geral, as repetições são elementos curtos que ocorrem em clusters que estão espaçados regularmente pela intervenção única de sequências com um comprimento substancialmente constante (Mojica et al., [2000], supra). Embora as sequências de repetição sejam altamente conservadas entre estirpes, o número de repetições intercaladas e as sequências das regiões espaçadoras normalmente diferem de estirpe para estirpe (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). Foram identificados loci CRISPR em mais de 40 procariotas (Ver por exemplo, Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; e Mojica et al., [2005]), incluindo, mas não limitados a *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*,

Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomyces, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema, e Thermotoga.

De um modo geral, o "sistema CRISPR" refere-se coletivamente aos transcritos e outros elementos envolvidos na expressão de ou dirigindo a atividade de genes associados a CRISPR ("Cas"), incluindo sequências que codificam um gene Cas, uma sequência *tracr* (CRISPR de transativação) (por exemplo, *tracrRNA* ou um *tracrRNA* parcial ativo), uma sequência de emparelhamento *tracr* (que inclui uma "repetição direta" e uma repetição direta parcial processada por *tracrRNA* no contexto de um sistema CRISPR endógeno), uma sequência guia (também referido como um "espaçador" no contexto de um sistema CRISPR endógena), ou de outras sequências e transcritos de um locus CRISPR. Em algumas formas de realização, um ou mais elementos de um sistema CRISPR é derivado de um sistema CRISPR de tipo I, tipo II, ou tipo III. Em algumas formas de realização, um ou mais elementos de um sistema CRISPR é derivado de um organismo em particular que compreende um sistema CRISPR endógeno, tal como *Streptococcus pyogenes*. Em geral, um sistema CRISPR é caracterizado por elementos que promovem a formação de um complexo CRISPR no local de uma sequência alvo (também referido como um proto-espaçador no contexto de um sistema endógeno CRISPR). No contexto da formação de

um complexo CRISPR, "a sequência alvo" refere-se a uma sequência em que uma sequência guia é desenhada para ter complementaridade, em que a hibridação entre uma sequência alvo e uma sequência guia promove a formação de um complexo CRISPR. A complementaridade total não é necessariamente requerida, desde que haja complementaridade suficiente para levar à hibridação e promover a formação de um complexo CRISPR. A sequência alvo pode compreender qualquer polinucleótido, tais como polinucleótidos de DNA ou RNA. Em algumas formas de realização, uma sequência alvo está localizada no núcleo ou citoplasma de uma célula. Em algumas formas de realização, a sequência alvo pode estar dentro de um organelo de uma célula eucariótica, por exemplo, mitocôndria ou cloroplasto. Uma sequência ou molde que pode ser utilizado para recombinação no locus alvo compreendendo as sequências alvo é denominado de um "molde de edição" ou "polinucleótido de edição" ou "sequência de edição". Em aspectos da presente invenção, um molde de polinucleótido exógeno pode ser referido como um molde de edição. Num aspecto da invenção, a recombinação é recombinação homóloga.

Tipicamente, no contexto de um sistema CRISPR endógeno, a formação de um complexo CRISPR (compreendendo uma sequência guia hibridada com uma sequência alvo e em complexo com uma ou mais proteínas Cas) resulta na clivagem de uma ou ambas as cadeias na, ou próximo da, sequência alvo (por exemplo, dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, ou mais pares de bases). Sem se pretender ficar limitado pela teoria, a sequência tracr, que pode compreender ou consistir na totalidade ou numa porção de uma sequência de tipo selvagem tracr (por exemplo, aproximadamente ou mais do que aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85,

ou mais nucleótidos de uma sequência tracr do tipo selvagem), também pode formar parte de um complexo CRISPR, tal como por hibridação ao longo de pelo menos uma porção da sequência tracr à totalidade ou uma porção de uma sequência de emparelhamento tracr operativamente ligada à sequência guia. Em algumas formas de realização, a sequência tracr tem complementaridade suficiente com uma sequência de emparelhamento tracr para hibridar e participar na formação de um complexo CRISPR. Tal como acontece com a sequência alvo, acredita-se que a complementaridade total não é necessária, desde que seja suficiente para ser funcional. Em algumas formas de realização, a sequência tracr tem, pelo menos, 50%, 60%, 70%, 80%, 90%, 95% ou 99% de complementaridade de sequência ao longo do comprimento da sequência de emparelhamento tracr quando otimamente alinhadas. Em algumas formas de realização, são introduzidos um ou mais vetores de expressão que orientam a expressão de um ou mais elementos de um sistema CRISPR numa célula hospedeira, de modo que a expressão dos elementos do sistema CRISPR conduza à formação direta de um complexo CRISPR num ou mais locais alvo. Por exemplo, uma enzima Cas, uma sequência guia ligada a uma sequência de emparelhamento tracr, e uma sequência tracr poderiam ser individualmente ligadas operacionalmente para separar elementos reguladores em vetores separados. Como alternativa, dois ou mais dos elementos expressos a partir dos mesmos ou diferentes elementos reguladores podem ser combinados num único vetor, com um ou mais vetores adicionais fornecendo quaisquer componentes do sistema CRISPR não incluídos no primeiro vetor. Os elementos do sistema CRISPR que são combinados num único vetor pode ser dispostos em qualquer orientação adequada, tal como um elemento localizado a 5' em relação a

("a montante" de) ou 3' em relação a ("a jusante" de) um segundo elemento. A sequência codificante de um elemento pode estar localizada na mesma ou na cadeia oposta da sequência codificante de um segundo elemento, e orientada na mesma ou oposta direção. Em algumas formas de realização, um único promotor leva à expressão de um transcrito que codifica para uma enzima CRISPR e uma ou mais das sequências guia, sequência de emparelhamento tracr (opcionalmente ligada de forma operativa à sequência guia), e uma sequência tracr incorporada dentro de uma ou mais sequências de intrão (por exemplo, cada uma num intrão diferentes, duas ou mais em pelo menos um intrão, ou todas num único intrão). Em algumas formas de realização, a enzima CRISPR, a sequência guia, a sequência de emparelhamento tracr e a sequência tracr são ligadas de forma operativa e expressas a partir do mesmo promotor.

Em algumas formas de realização, um vetor compreende um ou mais locais de inserção, tais como uma sequência de reconhecimento da endonuclease de restrição (também referido como um "local de clonagem"). Em algumas formas de realização, um ou mais locais de inserção (por exemplo, aproximadamente ou mais do que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais locais de inserção) estão localizados a montante e/ou a jusante de um ou mais elementos da sequência de um ou mais vetores. Em algumas formas de realização, um vetor compreende um local de inserção a montante de uma sequência de emparelhamento tracr, e, opcionalmente, a jusante de um elemento regulador ligado operacionalmente à sequência de emparelhamento tracr, de tal modo que após a inserção de uma sequência guia no local de inserção e após a expressão da sequência guia dirige a ligação específica da sequência de um

complexo de CRISPR a uma sequência alvo numa célula eucariótica. Em algumas formas de realização, um vetor compreende dois ou mais locais de inserção, cada local de inserção está localizado entre duas sequências de emparelhamento tracr de modo a permitir a inserção de uma sequência guia em cada local. Em tal disposição, as duas ou mais sequências guia podem compreender duas ou mais cópias de uma única sequência guia, duas ou mais sequências guia diferentes, ou combinações das mesmas. Quando várias sequências guia diferentes são utilizadas, um único vetor (*construct*) de expressão pode ser utilizado para direcionar a atividade CRISPR para múltiplas sequências alvo correspondentes diferentes dentro de uma célula. Por exemplo, um único vetor pode compreender aproximadamente ou aproximadamente mais de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, ou mais sequências guia. Em algumas formas de realização, podem ser fornecidos aproximadamente ou aproximadamente mais do que 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais de tais vetores contendo sequência guia, e opcionalmente entregues a uma célula.

Em algumas formas de realização, um vetor compreende um elemento regulador ligado de forma operacional a uma sequência codificante da enzima que codifica uma enzima CRISPR, tal como uma proteína Cas. Os exemplos não limitativos de proteínas Cas incluem Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (também conhecido como Csn1 e Csx12), Cas10, 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, homólogos das mesmas, ou versões modificadas das mesmas. Estas enzimas são conhecidas; por exemplo, a

sequência de aminoácidos da proteína Cas9 de *S. pyogenes* pode ser encontrada na base de dados SwissProt com o número de acesso Q99ZW2. Em algumas formas de realização, a enzima CRISPR não modificada tem atividade de clivagem de DNA, tal como a Cas9. Em algumas formas de realização a enzima CRISPR é a Cas9, e pode ser Cas9 de *S. pyogenes* ou de *S. pneumoniae*. Em algumas formas de realização, a enzima CRISPR dirige a clivagem de uma ou ambas as cadeias no local de uma sequência alvo, tal como dentro da sequência alvo e/ou no interior do complemento da sequência alvo. Em algumas formas de realização, a enzima CRISPR dirige a clivagem de uma ou ambas as cadeias dentro de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, ou mais pares de bases a partir do primeiro ou do último nucleótido de uma sequência alvo. Em algumas formas de realização, um vetor codifica uma enzima CRISPR que é mutada em relação a uma enzima de tipo selvagem correspondente de tal modo que a enzima CRISPR mutada não tem a capacidade para clivar uma ou ambas as cadeias de um polinucleótido alvo contém uma sequência alvo. Por exemplo, uma substituição de aspartato por alanina (D10A) no domínio catalítico RuvC I da Cas9 de *S. pyogenes* converte a Cas9 de uma nuclease que cliva ambas as cadeias para uma *nickase* (cliva uma cadeia simples). Outros exemplos de mutações que tornam a Cas9 uma *nickase* incluem, sem limitação, H840A, N854A, e N863A. Em algumas formas de realização, uma Cas9 *nickase* pode ser usada em combinação com uma ou umas sequências guia, por exemplo, duas sequências de guia, que têm como alvo, respetivamente, as cadeias sense e anti-sense do DNA alvo. Esta combinação permite que ambas as cadeias sejam cortadas e utilizadas para induzir NHEJ. Os requerentes demonstraram (dados não apresentados) a eficácia de duas *nickase* alvos (ou seja,

sgRNAs direcionados para o mesmo local, mas com diferentes cadeias de DNA) na indução de NHEJ mutagénico. Uma única *nickase* (Cas9-D10A com um único sgRNA) é incapaz de induzir NHEJ e criar indels mas os Requerentes demonstraram que a dupla *nickase* (Cas9-D10A e dois sgRNAs direcionados para diferentes cadeias na mesma localização) pode fazê-lo em células estaminais embrionárias humanas (hESCs). A eficácia é aproximadamente de 50% da nuclease (por exemplo, Cas9 regular sem mutação D10) em hESCs.

Como mais um exemplo, dois ou mais domínios catalíticos de Cas9 (RuvC I, RuvC II, e RuvC III) podem ser mutados para produzir um mutante Cas9 que carece substancialmente de toda a atividade de clivagem de DNA. Em algumas formas de realização, uma mutação D10A é combinado com uma ou mais mutações H840A, N854A, ou N863A para produzir uma enzima Cas9 que carece substancialmente de toda a atividade de clivagem de DNA. Em algumas formas de realização, uma enzima CRISPR é considerada que carece substancialmente de toda a atividade de clivagem de DNA em que a atividade de clivagem do DNA da enzima mutada é menos do que aproximadamente 25%, 10%, 5%, 1%, 0,1%, 0,01%, ou inferior, com respeito à sua forma não mutada. Outras mutações podem ser úteis; quando a enzima Cas9 ou outra enzima CRISPR é de uma espécie diferente de *S. pyogenes*, mutações nos aminoácidos correspondentes podem ser feitas para conseguir efeitos similares.

Em algumas formas de realização, uma sequência codificante da enzima que codifica uma enzima CRISPR tem codões otimizados para a expressão em células específicas, tais como as células eucarióticas. As células eucarióticas podem ser aquelas de ou derivadas de um organismo específico, tal

como um mamífero, incluindo, mas não limitado a, humanos, ratinho, rato, coelho, cão, ou um primata não humano. Em geral, a otimização de codões refere-se a um processo de modificação de uma sequência de ácidos nucleicos para aumentar a expressão nas células hospedeiras de interesse através da substituição de pelo menos um codão (por exemplo, aproximadamente ou mais do que aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, ou mais codões) da sequência nativa com codões que são mais frequentes ou mais frequentemente utilizados nos genes dessa célula hospedeira enquanto mantendo a sequência de aminoácidos nativa. Várias espécies exibem preferências para determinados codões de um aminoácido particular. As preferências de codão (diferenças na utilização de codões entre organismos) frequentemente estão correlacionadas com a eficiência de tradução de RNA mensageiro (mRNA), que por sua vez que se acredita que depende, entre outras coisas, das propriedades dos codões a serem traduzidos e disponibilidade de determinadas moléculas de RNA de transferência (tRNA). O predomínio de tRNA selecionados numa célula é geralmente um reflexo dos codões utilizados mais frequentemente na síntese de péptidos. Consequentemente, os genes podem ser adaptados para uma expressão genética otimizada num dado organismo com base na otimização de codões. Tabelas de utilização de codões estão prontamente disponíveis, por exemplo, na "Codon Usage Database" (base de dados de uso de codões), e essas tabelas pode ser adaptadas num variado número de maneiras. Ver Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence database: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Também estão disponíveis algoritmos computacionais para a otimização de codão de uma sequência particular para a expressão numa célula hospedeira particular, tal como Gene Forge (Aptagen;

Jacobus, PA). Em algumas formas de realização, um ou mais codões (por exemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, ou mais, ou todos os codões) de uma sequência que codifica uma enzima CRISPR corresponde ao codão utilizado mais frequentemente para um aminoácido em particular.

Em algumas formas de realização, um vetor codifica uma enzima CRISPR que compreende uma ou mais sequências de localização nuclear (NLS), tal como aproximadamente ou mais que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais NLS. Em algumas formas de realização, a enzima CRISPR compreende aproximadamente ou mais que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais NLS no terminal amino ou perto deste, aproximadamente ou mais que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais NLS no terminal carboxílico ou perto deste, ou uma combinação dos mesmos (por exemplo, uma ou mais NLS no terminal amino e uma ou mais NLS no terminal carboxílico). Quando mais do que uma NLS está presente, cada uma pode ser selecionada independentemente das outras, de tal modo que uma única NLS pode estar presente em mais do que uma cópia e/ou em combinação com uma ou mais outras NLS presente numa ou mais cópias. Numa forma de realização preferida da invenção, a enzima CRISPR compreende no máximo 6 NLS. Em algumas formas de realização, uma NLS é considerada perto do N- ou C-terminal, quando o aminoácido mais próximo da NLS está dentro de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, ou mais aminoácidos ao longo da cadeia polipeptídica a partir do N- ou C-terminal. Tipicamente, uma NLS consiste em uma ou mais sequências curtas de lisinas ou argininas carregadas positivamente expostas na superfície da proteína, mas outros tipos de NLS são conhecidos. Exemplos não limitativos de NLS incluem uma

sequência NLS derivada: da NLS do antígeno T grande do vírus SV40, que tem a sequência de aminoácidos PKKKRKV; da NLS da nucleoplasmina (por exemplo, a NLS bipartida da nucleoplasmina com a sequência KRPAATKKAGQAKKKK); das NLS de c-myc que possui a sequência de aminoácido PAAKRVKLD ou RQRRNELKRSP; da NLS de hRNPA1 M9 tendo a sequência NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; da sequência RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV do domínio IBB da importina-alfa; das sequências VSRKRPRP e PPKKARED da proteína T de mioma; da sequência POPKKKPL de p53 humana; da sequência SALIKKKKKMAP de c-abl IV de rato; das sequências DRLRR e PKQKKRK do vírus NS1 da gripe; da sequência RKLKKKIKKL do antígeno delta do vírus da Hepatite; da sequência REKKKFLKRR da proteína Mx1 de rato; da sequência KRKGDEV DGVDEVAKKSKK da poli(ADP-ribose)-polimerase humana; e da sequência RKCLQAGMNLEARKTKK de glucocorticoides de recetores de hormonas esteroides (humano).

Em geral, a uma ou mais NLS são de uma resistência suficiente para levar à acumulação da enzima CRISPR em quantidade detetável no núcleo de uma célula eucariótica. Em geral, a força da atividade de localização nuclear pode derivar a partir do número de NLS na enzima CRISPR, a ou as NLS usadas em particular, ou uma combinação destes fatores. A deteção da acumulação no núcleo pode ser realizada por qualquer técnica adequada. Por exemplo, um marcador detetável pode ser fundido com a enzima CRISPR, de modo a que a localização dentro de uma célula possa ser visualizada, tal como em combinação com um meio para detetar a localização do núcleo (por exemplo, um corante específico para o núcleo, tal como DAPI). Exemplos de marcadores detetáveis incluem proteínas fluorescentes (tais

como as proteínas fluorescentes de cor Verde, ou GFP; RFP; CFP), e marcadores de epítipo (etiqueta (*tag*) HA, etiqueta flag, etiqueta SNAP). Os núcleos das células também podem ser isolados a partir de células, cujos conteúdos se podem então analisar por qualquer processo adequado para a detecção de proteína, tal como imunohistoquímica, Western blot, ou ensaios de atividade enzimática. A acumulação no núcleo pode também ser determinada indiretamente, tal como por um ensaio para o efeito de formação de complexos CRISPR (por exemplo, ensaio de clivagem de DNA ou mutação na sequência alvo, ou ensaio para a alteração da atividade de expressão do gene alterada afetada pela formação do complexo CRISPR e/ou atividade da enzima CRISPR), em comparação com um controlo não expostos à enzima ou complexo CRISPR, ou expostos a uma enzima CRISPR que tem em falta um ou mais NLS.

De um modo geral, uma sequência guia é qualquer sequência de polinucleótidos tendo complementaridade suficiente com uma sequência de polinucleótidos alvo para hibridar com a sequência alvo e orientar a ligação específica de sequências de um complexo CRISPR com a sequência alvo. Em algumas formas de realização, o grau de complementaridade entre a sequência guia e a sua sequência alvo correspondente, quando otimamente alinhadas utilizando um algoritmo de alinhamento adequado, é aproximadamente ou mais do que aproximadamente 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99%, ou mais. O alinhamento ótimo pode ser determinado com a utilização de qualquer algoritmo adequado para alinhar sequências, e exemplos não limitativos dos mesmos incluem o algoritmo de Smith-Waterman, o algoritmo de Needleman-Wunsch, algoritmos baseados na Transformação de Burrows-Wheeler (por exemplo, o Alinhador de Burrows

Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, CA), SOAP (disponível em soap.genomics.org.cn), e Maq (disponível em maq.sourceforge.net). Em alguns formas de realização, uma sequência guia é aproximadamente ou mais do que aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, ou mais nucleótidos de comprimento. Em algumas formas de realização, uma sequência guia é menor do que aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, ou menos nucleótidos de comprimento. A capacidade de uma sequência guia para orientar a ligação específica da sequência de um complexo CRISPR a uma sequência alvo pode ser avaliada por qualquer ensaio adequado. Por exemplo, os componentes de um sistema CRISPR suficientes para formar um complexo CRISPR, incluindo a sequência guia a ser testada, podem ser fornecidos a uma célula hospedeira tendo a correspondente sequência alvo, tal como por transfecção com vetores que codificam os componentes da sequência CRISPR, seguida por uma avaliação da clivagem preferencial no interior da sequência alvo, tal como pelo ensaio Surveyor descrito na presente divulgação. Da mesma forma, a clivagem de uma sequência de polinucleótidos alvo pode ser avaliada num tubo de ensaio fornecendo a sequência alvo, componentes de um complexo CRISPR, incluindo a sequência guia para ser testada e uma sequência guia de controlo diferente da sequência guia do teste, e comparando a ligação ou taxa de clivagem na sequência alvo entre reações com sequência guia em teste e controlo. Outros ensaios são possíveis, e irão ocorrer aos peritos na técnica.

Uma sequência guia pode ser selecionada para ser dirigida para qualquer sequência alvo. Em algumas formas de

realização, a sequência alvo é uma sequência dentro do genoma de uma célula. Sequências alvo exemplares incluem aqueles que são únicas no genoma alvo. Por exemplo, para a Cas9 de *S. pyogenes*, uma sequência alvo única num genoma pode incluir um local alvo Cas9 na forma MMMMMMMNNNNNNNNNNNNXGG onde NNNNNNNNNNNXGG (N é A, G, T, ou C; e X pode ser qualquer coisa) tem uma única ocorrência no genoma. Uma sequência alvo única num genoma pode incluir um local alvo Cas9 de *S. pyogenes* na forma MMMMMMMNNNNNNNNNNNNXGG onde NNNNNNNNNNNXGG (N é A, G, T, ou C; e X pode ser qualquer coisa) tem uma única ocorrência no genoma. Para a CRISPR1 Cas9 de *S. thermophilus*, uma sequência alvo única num genoma pode incluir um local alvo Cas9 na forma MMMMMMMNNNNNNNNNNNNXXAGAAW onde NNNNNNNNNNNXXAGAAW (N é A, G, T, ou C; X pode ser qualquer coisa; e W é A ou t) tem uma única ocorrência no genoma. Uma sequência alvo única num genoma pode incluir um local alvo CRISPR1 Cas9 de *S. thermophilus* na forma MMMMMMMNNNNNNNNNNNNXXAGAAW onde NNNNNNNNNNNXXAGAAW (N é A, G, T, ou C; X pode ser qualquer coisa; e W é A ou T) tem uma única ocorrência no genoma. Para a Cas9 de *S. pyogenes*, uma sequência alvo única num genoma pode incluir um local alvo Cas9 na forma MMMMMMMNNNNNNNNNNNNXGGXG onde NNNNNNNNNNNXGGXG (N é A, G, T, ou C; e X pode ser qualquer coisa) tem uma única ocorrência no genoma. Uma sequência alvo única num genoma pode incluir um local alvo Cas9 de *S. pyogenes* da forma MMMMMMMNNNNNNNNNNNNXGGXG onde NNNNNNNNNNNXGGXG (N é A, G, T, ou C; e X pode ser qualquer coisa) tem uma única ocorrência no genoma. Em cada uma destas sequências "M" pode ser A, G, T ou C, e não necessitam de ser considerados na identificação de uma sequência como única.

Em algumas formas de realização, uma sequência guia é selecionada para reduzir o grau de estrutura secundária no interior da sequência guia. A estrutura secundária pode ser determinada por qualquer algoritmo adequado de folding de polinucleótidos. Alguns programas são baseados no cálculo da energia livre de Gibbs mínima. Um exemplo de um tal algoritmo é mFold, tal como descrito por Zuker e Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Outro exemplo de algoritmo de folding é o servidor da web on-line RNAfold, desenvolvido no Instituto de Química Teórica na Universidade de Viena, usando o algoritmo de previsão de estrutura centróide (ver por exemplo A.R. Gruber *et al.*, 2008, *Cell* 106(1):23-24.; e P.A. Carr e GM Church, 2009, *Nature Biotechnology* 27(12):1151-1162).

De um modo geral, uma sequência de emparelhamento tracr inclui qualquer sequência que tem uma complementaridade suficiente com uma sequência tracr para promover um ou mais de: (1) excisão de uma sequência guia flanqueada por sequências de emparelhamento tracr numa célula que contém a sequência tracr correspondente; e (2) formação de um complexo CRISPR numa sequência alvo, em que o complexo CRISPR compreende a sequência de emparelhamento tracr hibridada com a sequência tracr. Em geral, o grau de complementaridade refere-se ao alinhamento ótimo da sequência de emparelhamento tracr e a sequência tracr, ao longo do comprimento da mais curta das duas sequências. O alinhamento ótimo pode ser determinado por qualquer algoritmo de alinhamento adequado, e pode ainda ter conta estruturas secundárias, como a auto complementaridade dentro da sequência tracr ou da sequência de emparelhamento tracr. Em algumas formas de realização, o grau de complementaridade entre a sequência tracr e a sequência de

emparelhamento tracr ao longo do comprimento do mais curto dos dois quando otimamente alinhado é de aproximadamente ou mais do que aproximadamente 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99%, ou superior. Nas Figuras 12B e 13B são fornecidas ilustrações exemplo de alinhamento ótimo entre uma sequência tracr e uma sequência de emparelhamento tracr. Em algumas formas de realização, a sequência tracr tem aproximadamente ou mais do que aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, ou mais nucleótidos de comprimento. Em algumas formas de realização, a sequência tracr e a sequência de emparelhamento tracr estão contidas dentro de um único transcrito, de modo a que a hibridação entre os dois produz um transcrito que tem uma estrutura secundária, tal como uma forquilha. As sequências preferidas que formam alças para uso em estruturas forquilha têm quatro nucleótidos de comprimento, e mais preferivelmente têm a sequência GAAA. No entanto, sequências alça mais curtas ou mais longas podem ser usadas, assim como se podem usar sequências alternativas. As sequências incluem, de preferência um triplete de nucleótidos (por exemplo, AAA), e um nucleótido adicional (por exemplo, C ou G). Exemplos de sequências que formam alças incluem CAAA e AAAG. Numa forma de realização da divulgação, a transcrição ou a sequência de polinucleótidos transcritos tem, pelo menos, duas ou mais forquilhas. Em outras formas de realização da divulgação, o transcrito tem duas, três, quatro ou cinco forquilhas. Numa outra forma de realização, o transcrito tem como máximo cinco forquilhas. Em algumas formas de realização, o transcrito individual inclui ainda uma sequência de finalização da transcrição; de preferência esta é uma sequência poli-T, por exemplo, seis nucleótidos T. Uma ilustração com exemplo de uma tal estrutura de forquilha é

fornecida na porção inferior da Figura 13B, em que a porção da sequência 5' do "N" final e a montante da alça corresponde à sequência de emparelhamento tracr, e a porção da sequência 3' da alça corresponde à sequência tracr. Outros exemplos não limitativos de polinucleótidos individuais que compreendem uma sequência guia, uma sequência de emparelhamento tracr, e uma sequência tracr são os seguintes (listados 5' para 3'), em que "N" representa uma base de uma sequência guia, o primeiro bloco de letras minúsculas representam a sequência de emparelhamento tracr, e o segundo bloco de letras minúsculas representam a sequência tracr, e a sequência final poli-T representa o terminador da transcrição:

- (1) NNNNNNNNNNNNNNNNNNNNNNNgittttgtactctcaagatttaGAAAtaaaacttgcagaagctacaaagataaggett
catgccgaaatcaacaccctgtcattttatggcaggggtgttttcgtatttaaTTTTTT;
- (2) NNNNNNNNNNNNNNNNNNNNNNNgittttgtactctcaGAAAtgcagaagctacaaagataaggettcatgccgaaatca
acaccctgtcattttatggcaggggtgttttcgtatttaaTTTTTT;
- (3) NNNNNNNNNNNNNNNNNNNNNNNgittttgtactctcaGAAAtgcagaagctacaaagataaggettcatgccgaaatca
acaccctgtcattttatggcaggggtgtTTTTT;
- (4) NNNNNNNNNNNNNNNNNNNNNNNgitttagagcttaGAAAtagcaagttaaaataaggctagtcctgtatcaacttgaaaa
agtggcaccgagtcgggtgcTTTTTT;
- (5) NNNNNNNNNNNNNNNNNNNNNNNgitttagagcttaGAAATAGcaagttaaaataaggctagtcctgtatcaacttgaa
aaagtgTTTTTT;
- (6) NNNNNNNNNNNNNNNNNNNNNNNgitttagagcttagAAATAGcaagttaaaataaggctagtcctgtatcaTTTTT
TTT.

Em algumas formas de realização, as sequências (1) a (3) são utilizadas em combinação com a Cas9 de CRISPR1 de *S. thermophilus*. Em algumas formas de realização, as sequências (4) a (6) são utilizadas em combinação com Cas9 de *S. pyogenes*. Em algumas formas de realização, a sequência tracr é um transcrito separado de uma transcrição que compreende a sequência de emparelhamento tracr (tal como se ilustra na parte superior da Figura 13B).

Em algumas formas de realização, um modelo de recombinação é também fornecido. Um modelo de recombinação pode ser um componente de um outro vetor, tal como descrito na presente divulgação, contido num vetor separado, ou fornecido como um polinucleótido separado. Em algumas formas de realização, um modelo de recombinação é concebido para servir como um molde no processo de recombinação homóloga, tal como no interior ou perto de uma sequência alvo nicked ou clivada por uma enzima CRISPR como parte de um complexo de CRISPR. Um polinucleótido molde pode ser de qualquer comprimento adequado, tal como aproximadamente ou mais do que aproximadamente 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, ou mais nucleótidos de comprimento. Em algumas formas de realização, o polinucleótido molde é complementar a uma porção de um polinucleótido que compreende a sequência alvo. Quando perfeitamente alinhado, um polinucleótido molde pode sobrepor-se com um ou mais nucleótidos de uma sequência alvo (por exemplo, aproximadamente ou mais de aproximadamente 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 ou mais nucleótidos). Em algumas formas de realização, quando uma sequência molde e um polinucleótido compreendendo uma sequência alvo estão otimamente alinhados, o nucleótido mais próximo do polinucleótido molde está dentro de aproximadamente 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000 ou mais nucleótidos a partir da sequência alvo.

Em algumas formas de realização, a enzima CRISPR é parte de uma proteína de fusão que compreende um ou mais domínios heterólogos de proteínas (por exemplo, aproximadamente ou mais do que aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou mais domínios, para além da enzima CRISPR). Uma proteína

de fusão de enzima CRISPR pode compreender qualquer sequência de proteínas adicional e, opcionalmente, uma sequência de ligação entre quaisquer dois domínios. Exemplos de domínios de proteínas que podem ser fundidos com uma enzima CRISPR incluem, sem limitação, marcadores de epítopo, sequências de gene repórter, e domínios de proteínas que possuam uma ou mais das seguintes atividades: atividade de metilase, atividade desmetilase, atividade de ativação da transcrição, atividade de repressão de transcrição, atividade do fator libertação da transcrição, atividade de modificação das histonas, atividade de clivagem de RNA e a atividade de ligação de ácido nucleico. Os exemplos não limitativos de marcadores de epítopo incluem etiquetas de histidina (His), etiqueta V5, etiquetas FLAG, hemaglutinina (HA) do vírus da gripe, etiqueta Myc, etiqueta VSV-G e etiqueta tiorredoxina (Trx). Exemplos de genes repórter incluem, mas não estão limitados a, glutathione-S-transferases (GST), peroxidase de rábano (HRP), cloranfenicol acetiltransferase (CAT), beta-galactosidase, beta-glucoronidase, luciferase, proteína fluorescente de cor verde (GFP), HcRed, DsRed, proteína fluorescente de cor ciano (PCP), proteína fluorescente de cor amarela (YFP), e proteínas autofluorescentes, incluindo proteína fluorescente de cor azul (BFP). Uma enzima CRISPR pode ser fundida a uma sequência de um gene que codifica uma proteína ou um fragmento de uma proteína que se liga a moléculas de DNA ou se liga a outras moléculas celulares, incluindo mas não se limitando a proteína de ligação à maltose (MBP), etiqueta-S, fusões do domínio de ligação ao DNA (DBD) Lex A, fusões do domínio ligação ao DNA GAL4, e fusões da proteína BP16 do vírus do herpes simplex (HSV). Domínios adicionais que podem formar parte de uma proteína de fusão que compreende uma enzima CRISPR são descritos em

US20110059502, aqui incorporada como referência. Em algumas formas de realização, uma enzima CRISPR com etiquetas é usado para identificar a localização de uma sequência alvo.

Em alguns aspetos, a presente divulgação prevê métodos que compreendem a entrega de um ou mais polinucleótidos, tais como um ou mais vetores tal como descrito nesta divulgação, um ou mais transcritos do mesmo, e/ou uma ou proteínas transcritas do mesmo, a uma célula hospedeira. Em alguns aspetos, a presente divulgação prevê ainda células produzidas por tais métodos, e organismos (tais como animais, plantas ou fungos) compreendendo, ou produzidos, a partir de tais células. Em algumas formas de realização, uma enzima CRISPR em combinação com (e, opcionalmente, em complexo com) uma sequência guia é entregue a uma célula. Podem ser usados métodos de transferência genética convencionais baseados em vírus ou não para introduzir ácidos nucleicos em células ou tecidos alvo de mamífero. Tais métodos podem ser utilizados para administrar ácidos nucleicos que codificam componentes de um sistema CRISPR em células em cultura, ou num organismo hospedeiro. Os sistemas de distribuição de vetores não virais incluem plasmídeos de DNA, RNA (por exemplo, a transcrição de um vetor descrita na presente divulgação), ácido nucleico nu e ácido nucleico complexado com um veículo de entrega, tal como um lipossoma. Os sistemas de distribuição de vetores virais incluem vírus de DNA e RNA, que têm que genomas epissômicos ou integrados após entrega à célula. Para uma revisão dos procedimentos de terapia genética, ver Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology

6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada *et al.*, em Current Topics in Microbiology and Immunology Doerfler e Böhm (eds) (1995); e Yu *et al.*, Gene Therapy 1:13-26 (1994).

Os métodos de administração não virais de ácidos nucleicos incluem lipofecção, nucleofecção, microinjeção, biolística, virossomas, lipossomas, imunolipossomas, policação ou conjugados de lípido:ácido nucleico, DNA nu, viriões artificiais e absorção do DNA aumentada por agente. A lipofecção é descrita em, por exemplo, Patente U.S. N°s 5,049,386, 4,946,787; e 4,897,355) e os reagentes de lipofecção são vendidos comercialmente (por exemplo, Transfectam™ e Lipofectin™). Os lípidos catiónicos e neutros adequados para a lipofecção eficiente de reconhecimento do recetor de polinucleótidos incluem os que são apresentados em Felgner, WO 91/17424; WO 91/16024. A entrega pode ser realizada em células (por exemplo, *in vitro* ou *ex vivo* da administração) ou tecidos alvo (por exemplo, a administração *in vivo*).

A preparação dos complexos lípido:ácido nucleico, incluindo lipossomas dirigidos, tais como complexos de imuno-lípidos, é bem conhecida de um perito na técnica (ver, por exemplo, Crystal, Science 270:404-410 (1995); Blaese *et al.*, Cancer Gene Ther. 2:291-297 (1995); Behr *et al.*, Bioconjugate Chem. 5:382-389 (1994); Remy *et al.*, Bioconjugate Chem. 5:647-654 (1994); Gao *et al.*, Gene Therapy 2:710-722 (1995); Ahmad *et al.*, Cancer Res. 52:4817-4820 (1992); Patente U.S. N°s 4,186,183, 4,217,344, 4,235,871,

4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, e 4,946,787).

A utilização de sistemas baseados em vírus de RNA ou de DNA para a entrega de ácidos nucleicos traz a vantagem de processos altamente desenvolvidos para direcionar vírus a células específicas no corpo e o tráfico de carga viral para o núcleo. Os vetores virais podem ser administrados diretamente aos pacientes (*in vivo*) ou podem ser utilizados para tratar células *in vitro*, e as células modificadas podem opcionalmente ser administradas a doentes (*ex vivo*). Os sistemas baseados em vírus convencionais podem incluir vetores retrovirais, de lentivírus, adenovirais, adeno-associados e de vírus herpes simplex para transferência genética. A integração no genoma do hospedeiro é possível com os métodos de transferência de genes de retrovírus, lentivírus, e de vírus adeno-associados, resultando frequentemente em expressão a longo prazo do transgene inserido. Além disso, foram observadas eficiências de transdução elevadas em muitos tipos diferentes de células e tecidos alvo.

O tropismo de um retrovírus pode ser alterado por incorporação de proteínas de envoltório externas, expandindo a população potencial alvo de células alvo. Os vetores lentivirais são vetores retrovirais que são capazes de transduzir ou infectar células que não estão em divisão e produzem tipicamente elevados títulos virais. A seleção de um sistema de transferência de genes retroviral depende portanto do tecido alvo. Os vetores retrovirais são constituídos por repetições terminais longas que atuam *in cis* com capacidade de empacotamento até 6-10 kb de sequência externa. As LTR que atuam *in cis* mínimas são

suficientes para replicação e empacotamento dos vetores, que são então utilizados para integrar o gene terapêutico na célula alvo levando à expressão permanente do transgene. Vetores retrovirais que são amplamente usados incluem aqueles com base em vírus da leucemia de murino (MuLV), vírus da leucemia de macaco gibão (GaLV), vírus da Imunodeficiência Símio (SIV), vírus da imunodeficiência humana (VIH) e suas combinações (ver, por exemplo, Buchscher *et al.*, J. Virol. 66:2731-2739 (1992); Johann *et al.*, J. Virol. 66:1635-1640 (1992); Sommnerfelt *et al.*, Virol. 176:58-59 (1990); Wilson *et al.*, J. Virol. 63:2374-2378 (1989); Miller *et al.*, J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

Nas aplicações em que se prefere a expressão transiente, podem ser utilizados sistemas baseados em adenovírus. Os vetores baseados em adenovírus são capazes de eficiência de transdução muito elevada em muitos tipos celulares e não requerem divisão celular. Com tais vetores podem ser obtidos títulos e níveis de expressão elevados. Este vetor pode ser produzido em grandes quantidades num sistema relativamente simples. Os vetores do vírus adeno-associados ("AAV") também podem ser utilizados para transduzir células com ácidos nucleicos alvo, por exemplo, na produção *in vitro* de ácidos nucleicos e péptidos, para procedimentos de terapia genética *in vivo* e *ex vivo* (ver, por exemplo, West *et al.*, Virology 160:38-47 (1987); Patente U.S. N°. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94.:1351 (1994) A construção de vetores AAV recombinantes estão descritos em várias publicações, incluindo Patente U.S. N° 5,173,414; Tratschin *et al.*, Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, *et al.*, Mol. Cell. Biol. 4:2072-2081 (1984);

Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); e Samulski et al., J. Virol. 63:03822-3828 (1989).

As células de empacotamento são tipicamente utilizadas para formar partículas de vírus que são capazes de infectar uma célula hospedeira. Tais células incluem células 293, que empacotam adenovírus, e ψ 2 células ou células PA317, que empacotam retrovírus. Os vetores virais utilizados em terapia genética são gerados normalmente através da produção de uma linha celular que empacota um vetor de ácido nucleico para numa partícula viral. Os vetores contêm tipicamente as sequências virais mínimas necessárias para o empacotamento e subsequente integração num hospedeiro, sendo substituídas outras sequências virais por uma cassete de expressão para o polinucleótido ou polinucleótidos a expressar. As funções virais em falta são geralmente fornecidas em trans através da linha celular de empacotamento. Por exemplo, os vetores AAV utilizados em terapia genética geralmente possuem apenas sequências ITR do genoma AAV que são necessárias para empacotamento e integração no genoma do hospedeiro. O DNA viral é empacotado numa linha celular que contém um plasmídeo auxiliar que codifica outros genes AAV, nomeadamente rep e cap, mas não possuindo sequências ITR. A linha celular também pode ser infectada com adenovírus tal como um auxiliar. O vírus auxiliar promove a replicação do vetor AAV e a expressão de genes AAV do plasmídeo auxiliador. O plasmídeo auxiliar não é empacotado em quantidade significativa devido à ausência das sequências ITR. A contaminação com adenovírus pode ser reduzida, por exemplo, mediante o tratamento térmico ao qual é mais sensível o adenovírus do que o AAV. Métodos adicionais para a entrega

de ácidos nucleicos a células são conhecidos dos peritos na técnica. Ver, por exemplo, US20030087817.

Em algumas formas de realização, uma célula hospedeira é transfectada transitoriamente ou não transitoriamente com um ou mais vetores descritos na presente divulgação. Em algumas formas de realização, uma célula é transfectada na forma como ocorre naturalmente num sujeito. Em algumas formas de realização, uma célula que é transfectada é retirada de um sujeito. Em algumas formas de realização, a célula tem origem em células retiradas de um sujeito, tal como uma linha celular. Uma ampla variedade de linhas celulares para cultura de tecidos é conhecida na técnica. Exemplos de linhas de células incluem, mas não estão limitados a células, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKA, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, RAW264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, epiteliais renais de macaco BS-C-1, fibroblastos de embriões de rato BALB/3T3, 3T3 Swiss, 3T3-L1, fibroblastos fetais humanos 132-d5; fibroblastos de rato 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, células JY, células K562, Ku812, KCL22, KG1,

KY01, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, linhas celulares OPCN/OPCT, Peer, PNT-1A/PNT2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SKBR3, T2, T-47D, T84, linha celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, Yar e variedades transgênicas das mesmas. As linhas celulares estão disponíveis a partir de diversas fontes conhecidas dos peritos na técnica (ver, por exemplo, a Coleção Americana de Culturas Tipo (ATCC, American Type Culture Collection) (Manassus, VA.)). Em algumas formas de realização, uma célula transfectada com um ou mais vetores descritos na presente divulgação é utilizada para estabelecer uma nova linha celular que compreende uma ou mais sequências derivadas de vetores. Em algumas formas de realização, uma célula transfectada de forma transitória com os componentes de um sistema CRISPR tal como descrito na presente divulgação (tal como, através de transfecção transiente de um ou mais vetores, ou transfecção com RNA), e modificada por meio da atividade de um complexo CRISPR, é usada para estabelecer uma nova linha celular compreendendo as células que contêm a modificação, mas sem qualquer outra sequência exógena. Em algumas formas de realização, as células transfectadas de forma transitória ou não transitória com um ou mais vetores descritos na presente divulgação, ou linhas celulares derivadas de tais células são usados na avaliação de um ou mais compostos de teste.

Em algumas formas de realização, um ou mais vetores descritos na presente divulgação são usados para produzir um animal transgênico não-humano ou planta transgênica. Em

algumas formas de realização, o animal transgénico é um mamífero, tal como um ratinho, rato, ou coelho. Em certas formas de realização, o organismo ou sujeito é uma planta. Em certas formas de realização, o organismo ou sujeito ou planta são algas. Os métodos para a produção de plantas e animais transgénicos são conhecidos na técnica, e geralmente começam com um método de transfecção celular, tal como descrito na presente divulgação. Animais transgénicos também são previstos, assim como as plantas transgénicas, especialmente culturas e algas. O animal ou planta transgénico pode ser útil em aplicações que vão para além de fornecer um modelo de doença. Estas podem incluir a produção de alimentos ou alimentos para animais, através da expressão de níveis mais elevados, por exemplo, de proteína, hidratos de carbono, nutrientes ou vitaminas do que seria normalmente observado no tipo selvagem. A este respeito, são preferidas as plantas transgénicas, especialmente os legumes e tubérculos, e animais, especialmente mamíferos, tais como gado (vacas, ovelhas, cabras e porcos), mas também aves domésticas e insetos comestíveis.

Algas ou outras plantas transgénicas, tais como a colza, podem ser particularmente úteis na produção de óleos vegetais ou de biocombustíveis, tais como álcoois (especialmente metanol e etanol), por exemplo. Estas podem ser manipuladas por engenharia para expressar ou sobre-expressar níveis elevados de óleo ou álcoois para utilização nas indústrias do petróleo ou de biocombustível.

Num aspeto, a presente divulgação prevê métodos para a modificação de um polinucleótido alvo numa célula eucariótica. Em algumas formas de realização, o método

compreende permitir que um complexo CRISPR se ligue ao polinucleótido alvo para efetuar a clivagem do referido polinucleótido alvo modificando assim o polinucleótido alvo, e em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com uma sequência guia hibridada com uma sequência alvo no interior do referido polinucleótido alvo, em que a referida sequência guia está ligada a uma sequência de emparelhamento tracr que por sua vez hibrida com uma sequência tracr.

Num aspeto, a presente divulgação prevê um método de modificar a expressão de um polinucleótido numa célula eucariótica. Em algumas formas de realização, o método compreende permitir a ligação de um complexo CRISPR ao polinucleótido tal que a referida ligação resulte dum aumento ou diminuição da expressão do referido polinucleótido; em que o complexo CRISPR compreende uma enzima CRISPR que forma um complexo com uma sequência guia hibridada com uma sequência alvo no interior do referido polinucleótido, em que a referida sequência guia é ligada a uma sequência de emparelhamento tracr que por sua vez hibrida com uma sequência tracr.

Com os avanços recentes em genómica de culturas, a capacidade para usar sistemas CRISPR-Cas para realizar a edição e manipulação de genes de modo eficiente e rentável permitirá a rápida seleção e comparação de manipulações genéticas individuais e multiplexadas para transformar esses genomas para melhoria da produção e aperfeiçoamento dos traços. A este respeito é feita referência a patentes US e publicações: Patente US N°. 6,603,061 - *Agrobacterium-Mediated Plant Transformation Method*; Patente US N°. 7,868,149 - *Plant Genome Sequences and Uses Thereof* e US

2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits. Na prática da presente divulgação, o conteúdo e divulgação de Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 29 Dezembro de 2011; 13(2):85-96 são também referidos. Numa forma de realização vantajosa da presente divulgação, o sistema CRISPR/Cas9 é utilizado para manipular microalgas por engenharia. Assim, pode-se aplicar na presente divulgação a referência a células animais, com as mudanças necessárias, a células de plantas a menos que seja evidente de outro modo.

Num aspeto, a presente divulgação prevê métodos para a modificação de um polinucleótido alvo numa célula eucariótica, que pode ser *in vivo*, *ex vivo* ou *in vitro*. Em algumas formas de realização, o método compreende a amostragem de uma célula ou população de células a partir de um animal humano ou não-humano ou planta (incluindo microalgas), e a modificação da célula ou células. A cultura pode ocorrer em qualquer fase *ex vivo*. A célula ou células podem mesmo voltar a ser introduzidas no animal não humano ou uma planta (incluindo microalgas).

Num aspeto, a presente divulgação prevê kits contendo qualquer um ou mais dos elementos descritos nos métodos e composições anteriores. Em algumas formas de realização, o kit compreende um sistema vetor e instruções para utilizar o kit. Em algumas formas de realização, o sistema de vetor compreende (a) um primeiro elemento regulador ligado operacionalmente a uma sequência de emparelhamento tracr e um ou mais locais de inserção para a inserção de uma ou mais sequências guia a montante da sequência de emparelhamento tracr, em que, quando expressa, a sequência guia direciona a ligação específica da sequência de um

complexo CRISPR a uma sequência alvo na célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR num complexo com (1) a sequência guia que hibrida com a sequência alvo, e (2) a sequência de emparelhamento tracr que hibrida com a sequência tracr; e/ou (b) um segundo elemento regulador, ligado operacionalmente a uma sequência que codifica enzimas que codificam a referida enzima CRISPR compreendendo uma sequência de localização nuclear. Os elementos podem ser fornecidos individualmente ou em combinações, e podem ser fornecidos em qualquer recipiente adequado, tal como um frasco, uma garrafa ou um tubo. Em algumas formas de realização, o kit inclui instruções numa ou mais línguas, por exemplo em mais do que uma língua.

Em algumas formas de realização, um kit compreende um ou mais reagentes para utilização num processo que utiliza um ou mais dos elementos descritos na presente divulgação. Os reagentes podem ser fornecidos em qualquer recipiente adequado. Por exemplo, um kit pode fornecer um ou mais tampões de reação ou de armazenamento. Os reagentes podem ser fornecidos numa forma que é utilizável num ensaio particular, ou numa forma que requer a adição de um ou mais outros componentes antes da sua utilização (por exemplo na forma concentrada ou liofilizada). Um tampão pode ser qualquer tampão, que inclui mas não se limita a um tampão de carbonato de sódio, tampão de bicarbonato de sódio, tampão borato, tampão Tris, tampão MOPS, tampão HEPES, e suas combinações. Em algumas formas de realização, o tampão é alcalino. Em algumas formas de realização, o tampão tem um pH de aproximadamente 7 a aproximadamente 10. Em algumas formas de realização, o kit compreende um ou mais oligonucleótidos correspondentes a uma sequência guia para inserção num vetor, de modo a ligar operacionalmente a

sequência guia e um elemento regulador. Em algumas formas de realização, o kit compreende um polinucleótido molde de recombinação homóloga.

Num aspeto, a divulgação prevê métodos para utilizar um ou mais elementos de um sistema CRISPR. O complexo CRISPR da divulgação fornece um meio efetivo para a modificação de um polinucleótido alvo. O complexo CRISPR da divulgação tem uma ampla variedade de utilizações incluindo a modificação (por exemplo, deleção, inserção, translocação, inativação, ativação) de um polinucleótido alvo numa multiplicidade de tipos de células. Como tal, o complexo CRISPR da divulgação tem um amplo espectro de aplicações em, por exemplo, terapia genética, rastreio de fármacos, diagnóstico de doenças e prognóstico. Um complexo CRISPR, a modo de exemplo, compreende uma enzima CRISPR que forma um complexo com uma sequência guia hibridada com uma sequência alvo dentro do polinucleótido alvo. A sequência guia liga-se a uma sequência de emparelhamento tracr, que por sua vez hibrida com uma sequência tracr.

O polinucleótido alvo de um complexo CRISPR pode ser qualquer polinucleótido endógeno ou exógeno para a célula eucariótica. Por exemplo, o polinucleótido alvo pode ser um polinucleótido que reside no núcleo da célula eucariótica. O polinucleótido alvo pode ser uma sequência que codifica um produto de um gene (por exemplo, uma proteína) ou uma sequência não codificadora (por exemplo, um polinucleótido regulador ou um DNA sem utilidade). Sem se pretender ficar limitado pela teoria, acredita-se que a sequência alvo deve ser associada a um PAM (motivo adjacente proto-espçador); isto é, uma sequência curta reconhecida pelo complexo CRISPR. Os requisitos específicos de sequência e

comprimento para o PAM diferem dependendo da enzima CRISPR utilizada, mas PAM são tipicamente sequências de 2-5 pares de bases adjacente ao proto-espacedor (isto é, a sequência alvo). Exemplos de sequências de PAM são dados na secção de exemplos abaixo, e o perito na técnica será capaz de identificar outras sequências de PAM para utilização com uma determinada enzima CRISPR.

O polinucleótido alvo de um complexo CRISPR pode incluir uma série de genes e polinucleótidos associados com uma doença, assim como genes e polinucleótidos associados a vias bioquímicas de sinalização.

Exemplos de polinucleótidos alvo incluem uma sequência associada com uma via bioquímica de sinalização, por exemplo, um gene ou polinucleótido associado a uma via bioquímica de sinalização. Exemplos de polinucleótidos alvo incluem um gene ou polinucleótido associado com doenças. Um gene ou polinucleótido "associado a doenças" refere-se a qualquer gene ou polinucleótido que está a produzir produtos de transcrição ou tradução a um nível anormal ou numa forma anormal em células derivadas de um tecido afetado por doença em comparação com os tecidos ou células de um controlo sem doença. Pode ser um gene que se expressa a um nível anormalmente elevado; pode ser um gene que se expressa a um nível anormalmente baixo, em que a expressão alterada se correlaciona com a ocorrência e/ou progressão da doença. Um gene associado à doença também se refere a um gene que possui mutação ou mutações ou variação genética que são diretamente responsáveis ou estão em desequilíbrio de ligação com um gene ou genes que são responsáveis pela etiologia de uma doença. Os produtos transcritos ou

traduzidos podem ser conhecidos ou desconhecidos, e podem estar presentes a um nível normal ou anormal.

Exemplos de genes e polinucleótidos associados à doença estão disponíveis a partir de Instituto McKusick-Nathans de Medicina Genética, da Universidade Johns Hopkins (Baltimore, Md.) e do Centro Nacional para Informação sobre Biotecnologia, Biblioteca Nacional de Medicina (Bethesda, Md.), disponível na World Wide Web.

Exemplos de genes e polinucleótidos associados à doença estão listados nos Tabelas A e B. A informação específica sobre a doença está disponível a partir do Instituto McKusick-Nathans de Medicina Genética, da Universidade Johns Hopkins (Baltimore, Md.) e do Centro Nacional para Informação sobre Biotecnologia, Biblioteca Nacional de Medicina (Bethesda, Md.), disponível na World Wide Web. Exemplos de genes e polinucleótidos associados a vias bioquímicas de sinalização estão listados na Tabela C.

As mutações nestes genes e vias podem resultar na produção de proteínas incorretas ou proteínas em quantidades inadequadas que afetam a função. Tais genes, proteínas e vias podem ser o polinucleótido alvo de um complexo CRISPR.

Tabela A

DOENÇAS/DISTÚRBIOS	GENE ou GENES
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR gama; WT1 (Tumor de Wilms); membros da Família do Recetor FGF (5 membros: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1;

	VHL; BRCA1; BRCA2; AR (Recetor dos Androgénios); TSG101; IGF; Recetor IGF; Igf1 (4 variantes); Igf2 (3 variantes); Recetor Igf 1; Recetor Igf 2; Bax; Bcl2; família das caspases (9 membros: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Degeneração Macular relacionada com a Idade	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsina D; Vldlr; Ccr2
Esquizofrenia	Neuregulina 1 (Nrg1); Erb4 (recetor para a Neuregulina); Complexina 1 (Cplx1); Tph1 Triptofano hidrolase 1; Tph2 Triptofano hidrolase 2; Neurexina 1; GSK3; GSK3a; GSK3b
Distúrbios	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Daol)
Distúrbios de Repetições de Trinucleótidos	HTT (Dx de Huntington); SBMA/SMAX1/AR (Dx de Kennedy); FXN/X25 (Ataxia de Friedrich); ATX3 (Dx de Machado-Joseph); ATXN1 e ATXN2 (ataxias espinocerebelares); DMPK (distrofia miotónica); Atrofina-1 e Atn1 (Dx DRPLA); CBP (Creb-BP- instabilidade global); VLDLR (Alzheimer); Atxn7; Atxn10
Síndrome do X Frágil	FMR2; FXR1; FXR2; mGLUR5
Distúrbios Relacionados com a Secretase	APH-1 (alfa e beta); Presenilina (Psen1); nicastrina (Ncstn); PEN-2
Outros	Nos1; Parp1; Nat1; Nat2
Distúrbios relacionados com Priões	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c)
Dependência de drogas	Prkce (álcool); Drd2; Drd4; ABAT (álcool); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Grial (álcool)
Autismo	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexina 1; X Frágil (FMR2 (AFF2); FXR1; FXR2; Mglur5)

Doença de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, Aquaporina 1); Uchl1; Uchl3; APP
Inflamação	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 para IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
Doença de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkin; PINK1

Tabela B:

Doenças e distúrbios do sangue e coagulação	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Síndrome do linfócito nu (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Distúrbios de hemorragia (TBXA2R, P2RX1, P2X1); Fator H e fator tipo H 1 (HF1, CFH, HUS); Fator V e fator VIII (MCFD2); deficiência do Fator VII (F7); deficiência do Fator X (F10); deficiência do Fator XI (F11); deficiência do Fator XII (F12, HAF); deficiência do Fator XIII A (F13A1, F13A); deficiência do Fator XIII B (F13B); anemia de Fanconi (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCI, PHF9, FANCL, FANCM, KIAA1596); Síndrome de linfocitose hemofagocítica (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB), Síndrome Hemorrágica (PI, ATT, F5); Deficiências e distúrbios por leucócitos (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Anemia falciforme (HBB); Talassemia (HBA2, HBB, HBD, LCRB, HBA1).
Doenças e distúrbios de	Linfoma não Hodgkin de células B (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2,

desregulação celular e oncologia	FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Doenças e Distúrbios relacionados com Inflamação e Sistema imune	SIDA (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Síndrome linfoproliferativa autoimune (TNFRSF6, APT1, FAS, CD95, ALPS1A); Imunodeficiência combinada (IL2RG, SCIDX1, SCIDX, IMD4); VIH-1 (CCL5, SCYA5, D17S136E, TCP228), Suscetibilidade ou infecção por VIH (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCKR5 (CCR5)); Imunodeficiências (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflamação (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), IL-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Imunodeficiências combinadas graves (SCIDs) (JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Doenças e distúrbios do metabolismo, do fígado, de rins e de proteínas	Neuropatia amiloide (TTR, PALB); Amiloidose (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrose (KRT18, KRT8, C1RH1A, NAIC, TEX292, K1AA1988); Fibrose cística (CFTR, ABCC7, CF, MRP7); Doenças de armazenamento do glicogénio (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3),

	Insuficiência hepática, início precoce, e distúrbios neurológicos (SCOD1, SC01), Deficiência da lipase hepática (LIPC), Hepatoblastoma, cancro e carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Doença quística renal medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Fenilcetonúria (PAH, PKU1, QDPR, DHPR, PTS); Doença do rim policístico e hepática (FCYT, PKHD1, ARPKD, PKD 1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Doenças e distúrbios musculares/esqueléticos	Distrofia muscular de Becker (DMD, BMD, MYF6), Distrofia muscular de Duchenne (DMD, BMD); Distrofia muscular de Emery-Dreifuss (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Distrofia muscular de facioescapuloumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrose (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Doenças e distúrbios neurológicos e neuronais	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Doença de Alzheimer (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLA2, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autismo (MECP2, BZRAP1, MDGA2, Sema5A, Neurexina 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Síndrome do X Frágil (FMR2, FXR1, FXR2, mGLUR5); Doença de Huntington e distúrbios

	semelhantes à doença (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Doença de Parkinson (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Síndrome de Rett (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Sinucleína, DJ-1); Esquizofrenia (Neuregulina1 (Nrg1), Erb4 (recetor para a Neuregulina), Complexina1 (Cplx1), Tph1 Triptofano hidrolase 1, Tph2, Triptofano hidrolase 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Distúrbios relacionados com Secretase (APH-1 (alfa e beta), Presenilina (Psen1), nicastrina, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Distúrbios de Repetições de Trinucleótidos (HTT (Dx de Huntington), SBMA/SMAx1/AR (Dx de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Dx de Machado-Joseph), ATXN1 e ATXN2 (ataxias espinocerebelares), DMPK (distrofia miotónica), Atrofina-1 e Atn1 (Dx DRPLA), CBP (Creb-BP- instabilidade global), VLDLR (Alzheimer), Atxn7, Atxn10).
Doenças e distúrbios oculares	Degeneração macular relacionada com a idade (Abcr, Ccl2, Cc2, cp (ceruloplasmina), Timp3, catepsina D, Vldlr, Ccr2); Cataratas (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Opacidade e distúrbios da córnea (APOA1, TGFB1, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Córnea plana congénita (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1,

	GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Amaurose congênita de Leber (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Distrofia Macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).
--	--

Tabela C:

FUNÇÃO CELULAR	GENES
Via de sinalização PI3K/AKT	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2; PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNA1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1
Via de sinalização ERK/MAPK	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKCA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4;

	PRKCA; SRF; STAT1; SGK
Via de sinalização do Recetor dos Glucocorticoides	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPKB; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Via de sinalização de Orientação Axonal	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA
Via de sinalização do Recetor da Efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC I; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1;

	FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK
Via de sinalização do Citoesqueleto de Actina	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
Via de sinalização da Doença de Huntington	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPKB; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Via de sinalização da Apoptose	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA;

	CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1
Via de sinalização do Recetor das Células B	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGRI; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Via de sinalização do Extravasamento de Leucócitos	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; TTGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Via de sinalização da Integrina	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
Via de sinalização da Resposta em Fase Aguda	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13;

	IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
Via de sinalização PTEN	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L1; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
Via de sinalização de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3
Via de sinalização do Recetor do Hidrocarboneto Aromático	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1
Via de sinalização do Metabolismo de Xenobióticos	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8;

	PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
Via de sinalização SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC 1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; R1PK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK
Via de sinalização PPAR/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB 1; TGFB1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
Via de sinalização de NF-KB	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Via de sinalização de Neuregulina	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1;

	PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
Via de sinalização de Wnt e Beta-catenina	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFB1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
Via de sinalização do Recetor da Insulina	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
Via de sinalização de IL-6	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Colestase Hepática	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2;

	NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
Via de sinalização de IGF-1	IGF 1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1
Resposta ao Stress Oxidativo mediada por mNRF2	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Fibrose Hepática/Ativação de Células Hepáticas Estreladas	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFB1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Via de sinalização do PPAR	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Via de sinalização de Fc Épsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1;

	PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA
Via de sinalização do Recetor Acoplado à Proteína-G	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFkB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFkB1; BRAF; ATF4; AKT3; PRKCA
Metabolismo do Fosfato Inositol	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
Via de sinalização de PDGF	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
Via de sinalização de VEGF	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA
Via de sinalização das Células Assassinas Naturais	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Ciclo celular: Regulação do Ponto de Controlo G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;

	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
Via de sinalização do Recetor das Células T	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3
Via de sinalização do Recetor <i>Death</i> (FAS)	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
Via de sinalização de FGF	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF
Via de sinalização de GM-CSF	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
Via de sinalização da Esclerose Lateral Amiotrófica	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
Via de sinalização JAK/Stat	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1
Metabolismo de	PRKCE; IRAK1; PRKAA2; EIF2AK2;

Nicotinato e Nicotinamida	GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Via de sinalização de Quimiocinas	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; FLAN; CCL2; PRKCA
Via de sinalização de IL-2	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Depressão Sináptica de Longo Prazo	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Via de sinalização do Recetor dos Estrogénios	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Via da Ubiquitinação de Proteínas	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
Via de sinalização de IL-10	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6
Ativação de VDR/RXR	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5;

	CEBPB; FOXO1; PRKCA
Via de sinalização de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFB1; SMAD4; JUN; SMAD5
Via de sinalização do Recetor Toll-like	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBK; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
Via de sinalização de p38 MAPK	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFB1; MYC; ATF4; IL1R1; SRF; STAT1
Via de sinalização Neurotrofina/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4
Ativação de FXR/RXR	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Potenciação Sináptica de Longo Prazo	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Via de sinalização do Cálcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
Via de sinalização de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Via de sinalização da Hipoxia no Sistema Cardiovascular	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1

Inibição da Função do RXR mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
Ativação de LXR/RXR	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
Processamento da Amiloide	PRKCE; CSNK1E; MAPK1; CAPNS 1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
Via de sinalização de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCKS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1
Ciclo celular: Regulação do Ponto de Controlo G2/M de Danos no DNA	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
Via de sinalização do Óxido Nítrico no Sistema Cardiovascular	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Metabolismo das Purinas	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
Via de sinalização mediada por cAMP	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunção Mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3
Via de sinalização Notch	HES1; JAG1; NUMB; NOTCH4; ADAM 17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Via de Stress do Retículo Endoplasmático	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3
Metabolismo da Pirimidina	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1;

	NME1
Via de sinalização na doença de Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
Via de sinalização Cardíaca e Beta Adrenérgica	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
Glicólise/Gliconeogénese	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Via de sinalização do Interferão	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Via de sinalização de Sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Metabolismo dos Glicerofosfolípidos	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Degradação de Fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo do Triptofano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradação da Lisina	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Via de Reparação por Excisão de Nucleótidos	ERCC5; ERCC4; XPA; XPC; ERCC1
Metabolismo do Amido e da Sacarose	UCHL1; HK2; GCK; GPI; HK1
Metabolismo de Aminoaçúcares	NQO1; HK2; GCK; HK1
Metabolismo do Ácido Araquidónico	PRDX6; GRN; YWHAZ; CYP1B1
Via de sinalização do Ritmo Circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de Coagulação	BDKRB1; F2R; SERPINE1; F3
Via de sinalização do Recetor da Dopamina	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Metabolismo da Glutathione	IDH2; GSTP1; ANPEP; IDH1
Metabolismo dos Glicerolípidos	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo do Ácido Linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo da Metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo do Piruvato	GLO1; ALDH1A1; PKM2; LDHA
Metabolismo da Arginina e da Prolina	ALDH1A1; NOS3; NOS2A
Via de sinalização dos Eicosanóides	PRDX6; GRN; YWHAZ