

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2013.12.12	(73) Titular(es): PRESIDENT AND FELLOWS OF HARVARD COLLEGE 17 QUINCY STREET CAMBRIDGE, MA 02138 US MASSACHUSETTS INSTITUTE OF TECHNOLOGY 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	(72) Inventor(es): FENG ZHANG US LE CONG US PATRICK HSU US FEI RAN US
(43) Data de publicação do pedido: 2014.09.03	(74) Mandatário: JOÃO LUÍS PEREIRA GARCIA RUA CASTILHO, 167 2º 1070-050 LISBOA PT
(45) Data e BPI da concessão: 2015.02.11 106/2015	

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

(57) Resumo:

A INVENÇÃO REFERE-SE A SISTEMAS, MÉTODOS E COMPOSIÇÕES PARA A MANIPULAÇÃO DE SEQUÊNCIAS E/OU ATIVIDADES DE SEQUÊNCIAS-ALVO. SÃO FORNECIDOS VETORES E SISTEMAS DE VETORES, ALGUNS DOS QUAIS CODIFICAM UM OU MAIS COMPONENTES DE UM COMPLEXO CRISPR, BEM COMO MÉTODOS PARA A CONCEÇÃO E UTILIZAÇÃO DESSES VETORES. IGUALMENTE, SÃO FORNECIDOS MÉTODOS DE ORIENTAÇÃO DA FORMAÇÃO DO COMPLEXO CRISPR NAS CÉLULAS EUCARIÓTICAS E MÉTODOS PARA SELECIONAR CÉLULAS ESPECÍFICAS ATRAVÉS DA INTRODUÇÃO DE MUTAÇÕES PRECISAS UTILIZANDO O SISTEMA CRISPR-CAS.

RESUMO

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

A invenção refere-se a sistemas, métodos e composições para a manipulação de sequências e/ou atividades de sequências-alvo. São fornecidos vetores e sistemas de vetores, alguns dos quais codificam um ou mais componentes de um complexo CRISPR, bem como métodos para a concepção e utilização desses vetores. Igualmente, são fornecidos métodos de orientação da formação do complexo CRISPR nas células eucarióticas e métodos para selecionar células específicas através da introdução de mutações precisas utilizando o sistema CRISPR-Cas.

DESCRIÇÃO

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

CAMPO DA INVENÇÃO

A presente divulgação refere-se geralmente a sistemas, métodos e composições utilizados para o controlo da expressão de genes envolvendo a escolha de sequências-alvo, tal como a perturbação de genomas ou a edição de genes, que possam utilizar sistemas de vetores relacionados com CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats - Repetições Palindrômicas Curtas Agrupadas e Regularmente Espaçadas) e componentes das mesmas.

RELATÓRIO DA INVESTIGAÇÃO FEDERAL

Esta invenção foi realizada com apoio governamental concedido pelos National Institutes of Health (Institutos Nacionais de Saúde), NIH Pioneer Award DP1MH100706. O governo detém certos direitos na invenção.

ANTECEDENTES DA INVENÇÃO

Os recentes avanços nas técnicas de sequenciação de genomas e nos métodos de análise aceleraram significativamente a capacidade para catalogar e mapear fatores genéticos associados a uma diversa gama de doenças e funções biológicas. As tecnologias precisas de escolha de genomas-alvo são necessárias para permitir a engenharia inversa sistemática de variações genéticas causais permitindo uma perturbação seletiva de elementos genéticos individuais, bem como para avançar as aplicações médicas, biotecnológicas e de biologia sintética. Embora as técnicas de edição de genomas, tais como dedos de zinco de concepção, efetores do tipo ativador de transcrição (TALEs), ou meganucleases de migração (*homing*) estejam disponíveis para a produção de perturbações de genomas-alvo, continuam a ser necessárias novas tecnologias de engenharia de genomas que

sejam acessíveis, fáceis de instalar, escaláveis e recetivas para a escolha de múltiplas posições-alvo no genoma eucariótico.

SUMÁRIO DA INVENÇÃO

Existe uma necessidade urgente de técnicas e sistemas alternativos e robustos para a escolha de sequências-alvo com uma vasta série de aplicações. Esta invenção aborda esta necessidade e fornece vantagens relacionadas. O CRISPR/Cas ou sistema CRISPR-Cas (ambos os termos são utilizados alternadamente em todo este pedido de patente) não necessita da geração de proteínas personalizadas para escolher sequências específicas como alvo, mas em vez disso é possível programar uma única enzima Cas através de uma molécula de ARN pequeno para reconhecer um alvo de ADN específico, por outras palavras, a enzima Cas pode ser recrutada para um alvo de ADN específico utilizando a referida molécula de ARN pequeno. A adição do sistema CRISPR-Cas ao repertório de técnicas de sequenciação de genomas e métodos de análise pode simplificar significativamente a metodologia e acelerar a capacidade para catalogar e mapear os fatores genéticos associados a uma gama diversa de doenças e funções biológicas. Para utilizar o sistema CRISPR-Cas eficazmente para a edição de genomas sem efeitos deletérios, é essencial compreender os aspetos de engenharia e otimização destas ferramentas de engenharia de genomas, que são aspetos da invenção reivindicada.

Num aspeto, a divulgação fornece um sistema de vetores compreendendo um ou mais vetores. Em algumas formas de realização, o sistema compreende: (a) um primeiro elemento regulador ligado de modo funcional a uma sequência *tracr-mate* e a um ou mais locais de inserção para inserir uma ou mais sequências-guia a montante da sequência *tracr-mate*, em que, quando expressada, a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para

uma sequência-alvo numa célula, p. ex., célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr; e (b) um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica a referida enzima CRISPR compreendendo uma sequência de localização nuclear; em que os componentes (a) e (b) se encontram situados nos mesmos ou em diferentes vetores do sistema. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência tracr-mate sob o controlo do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende ainda duas ou mais sequências-guia ligadas de modo funcional ao primeiro elemento regulador, em que, quando expressadas, cada uma das duas ou mais sequências-guia direciona a ligação específica da sequência de um complexo CRISPR para 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 da 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 tracr-mate quando se encontra idealmente alinhada. Em algumas formas de realização, o complexo CRISPR compreende uma ou mais sequências de localização nuclear de força suficiente para ativar a acumulação do referido complexo CRISPR numa quantidade detetável no núcleo de uma célula eucariótica. Sem ser pretendida qualquer limitação a uma teoria, parte-se do princípio de que não é necessária uma sequência de localização nuclear para a atividade do complexo CRISPR nos eucariotas, mas que a inclusão dessas sequências melhora a atividade do

sistema, especialmente no que respeita à escolha de moléculas de ácido nucleico como alvo 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 corresponde a Cas9 de *S. pneumoniae*, *S. pyogenes* ou *S. thermophilus*, e pode incluir Cas9 mutante derivada destes organismos. A enzima pode ser um homólogo ou ortólogo de 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 direciona a clivagem de um ou dois filamentos na localização da sequência-alvo. Em algumas formas de realização, a enzima CRISPR não tem atividade de clivagem de filamentos de ADN. Em algumas formas de realização, o primeiro elemento regulador é um promotor da polimerase III. Em algumas formas de realização, o segundo elemento regulador é um promotor da polimerase II. Em algumas formas de realização, a sequência-guia tem, pelo menos, 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10 e 30, ou entre 15 e 25, ou entre 15 e 20 nucleótidos de comprimento. Em geral, e em toda esta 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 se limitam a, moléculas de ácido nucleico que são de filamento único, de filamento duplo ou de filamento parcialmente duplo; moléculas de ácido nucleico que compreendem uma ou mais extremidades livres, ou nenhuma extremidade livre (p. ex. circulares); moléculas de ácido nucleico que compreendem ADN, ARN, ou ambos; e outras variedades de polinucleótidos conhecidos na técnica. Um tipo de vetor é um "plasmídeo", que se refere a uma estrutura de ADN de filamento duplo circular na qual podem ser inseridos segmentos de ADN adicionais, tal como através de técnicas

de clonagem molecular padrão. Outro tipo de vetor é um vetor viral, em que as sequências de ADN ou ARN obtidas de forma viral estão presentes no vetor para o empacotamento num vírus (p. ex. retrovírus, retrovírus de replicação defeituosa, adenovírus, adenovírus de replicação defeituosa e vírus adeno-associados). Os vetores virais incluem igualmente polinucleótidos transportados por um vírus para a transfeção numa célula hospedeira. Determinados vetores são capazes de replicação autónoma numa célula hospedeira na qual são introduzidos (p. ex. vetores bacterianos com uma origem bacteriana de replicação e vetores de mamífero episomais). Outros vetores (p. ex., vetores de mamífero não episomais) são integrados no genoma de uma célula hospedeira na introdução na célula hospedeira e, desse modo, são replicados juntamente com o genoma hospedeiro. Além disso, determinados vetores são capazes de direcionar a expressão dos genes aos quais se encontram ligados de modo eficaz. Esses vetores são aqui referidos como "vetores de expressão". Os vetores de expressão comuns de utilidade em técnicas de ADN recombinantes encontram-se muitas vezes na forma de plasmídeos.

Os vetores de expressão recombinantes podem compreender um ácido nucleico numa forma adequada para a expressão do ácido nucleico numa célula hospedeira, o que significa que os vetores de expressão recombinantes incluem um ou mais elementos reguladores que podem ser selecionados com base nas células hospedeiras a utilizar para a expressão, que é ligada de modo eficaz à sequência de ácidos nucleicos que deve ser expressada. Num vetor de expressão recombinante, o termo "ligado de modo funcional" significa que a sequência de nucleótidos de interesse se encontra ligada aos elementos reguladores de maneira a permitir a expressão da sequência de nucleótidos (p. ex. num sistema de transcrição/tradução *in vitro* ou numa célula hospedeira quando o vetor é introduzido na célula

hospedeira).

O termo "elemento regulador" pretende incluir promotores, intensificadores, locais internos de entrada do ribossoma (IRES), e outros elementos de controlo de expressão (p. ex. sinais de terminação de transcrição, tais como sinais de poliadenilação e sequências poli-U). Esses elementos reguladores são descritos, por exemplo em Goeddel, "GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY" 185, *Academic Press*, San Diego, Calif. (1990). Os elementos reguladores incluem os que direcionam a expressão constitutiva de uma sequência de nucleótidos em muitos tipos de célula hospedeira e os que direcionam a expressão da sequência de nucleótidos apenas em determinadas células hospedeiras (p. ex., sequências reguladoras específicas do tecido). Um promotor específico do tecido pode direcionar a expressão primeiramente num tecido desejado de interesse, tal como músculo, neurónio, osso, pele, sangue, órgãos específicos (p. ex. fígado, pâncreas), ou determinados tipos de célula (p. ex. linfócitos). Os elementos reguladores também podem direcionar a expressão de uma maneira dependente do tempo, tal como de uma maneira dependente do ciclo celular ou dependente do estágio do desenvolvimento, que pode ou não ser igualmente específica do tipo de célula ou do tecido. Em algumas formas de realização, um vetor compreende um ou mais promotores pol III (p. ex. 1, 2, 3, 4, 5 ou mais promotores pol III), um ou mais promotores pol II (p. ex. 1, 2, 3, 4, 5 ou mais promotores pol II), um ou mais promotores pol I (p. ex. 1, 2, 3, 4, 5 ou mais promotores pol I), ou combinações dos mesmos. Os exemplos de promotores pol III incluem, mas não se limitam a, promotores U6 e H1. Os exemplos de promotores pol II incluem, mas sem limitação, o promotor LTR do vírus sarcoma Rous (RSV) retroviral (opcionalmente com o intensificador RSV), o promotor do citomegalovírus (CMV) (opcionalmente

com o intensificador CMV) [consulte, p. ex., Boshart *et al.*, *Cell*, 41:521 a 530 (1985)], o promotor SV40, o promotor da diidrofolato redutase, o promotor β -actina, o promotor da fosfoglicerol quinase (PGK) e o promotor EF1 α . Igualmente abrangidos pelo termo "elemento regulador" são os elementos intensificadores, tais como WPRE; intensificadores CMV; o segmento R-U5' em LTR de HTLV-I (*Mol. Cell. Biol.*, Vol. 8(1), págs. 466 a 472, 1988); intensificador SV40; e a sequência de intrões entre os exões 2 e 3 de β -globina de coelho (*Proc. Natl. Acad. Sci. E.U.A.*, Vol. 78(3), págs. 1527 a 31, 1981). Será apreciado pelos peritos na técnica que a concepção do vetor de expressão possa depender de fatores como a escolha da célula hospedeira a transformar, o nível de expressão desejado, etc. Um vetor pode ser introduzido nas células hospedeiras para, desse modo, produzir transcrições, proteínas ou péptidos, incluindo péptidos ou proteínas de fusão, codificados por ácidos nucleicos conforme aqui descrito (p. ex., transcrições CRISPR (repetições palindrômicas curtas agrupadas e regularmente interpoladas), proteínas, enzimas, formas mutantes das mesmas, proteínas de fusão das mesmas, etc.).

Os vetores vantajosos incluem lentivírus e vírus adeno-associados, e é igualmente possível selecionar tipos desses vetores para a escolha de determinados tipos de células-alvo.

Num aspeto, a divulgação fornece um vetor que compreende um elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica uma enzima CRISPR compreendendo uma ou mais sequências de localização nuclear. Em algumas formas de realização, o referido elemento regulador ativa a transcrição da enzima CRISPR numa célula eucariótica, de modo a que a referida enzima CRISPR seja acumulada numa quantidade detetável no núcleo da célula eucariótica. Em algumas formas de

realização, o elemento regulador é um promotor da 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 corresponde a Cas9 de *S. pneumoniae*, *S. pyogenes* ou *S. thermophilus*, e pode incluir Cas9 mutante derivada destes 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 direciona a clivagem de um ou dois filamentos na localização da sequência-alvo. Em algumas formas de realização, a enzima CRISPR não tem atividade de clivagem de filamentos de ADN.

Num aspeto, a divulgação fornece uma enzima CRISPR que compreende uma ou mais sequências de localização nuclear de força suficiente para ativar 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 corresponde a Cas9 de *S. pneumoniae*, *S. pyogenes* ou *S. thermophilus*, e pode incluir Cas9 mutante derivada destes organismos. A enzima pode ser um homólogo ou ortólogo de Cas9. Em algumas formas de realização, a enzima CRISPR não tem a capacidade para clivar um ou mais filamentos de uma sequência-alvo à qual se liga.

Num aspeto, a divulgação fornece uma célula hospedeira eucariótica que compreende (a) um primeiro elemento regulador ligado de modo funcional a uma sequência tracr-mate e a um ou mais locais de inserção para inserir uma ou mais sequências-guia a montante da sequência tracr-mate, em que, quando expressada, a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para

uma sequência-alvo numa célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr; e/ou (b) um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica 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), o componente (b) ou os componentes (a) e (b) são integrados de forma estável num genoma da célula eucariótica hospedeira. Em algumas formas de realização, o componente (a) compreende ainda a sequência tracr a jusante da sequência tracr-mate sob o controlo do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende ainda duas ou mais sequências-guia ligadas de modo funcional ao primeiro elemento regulador, em que, quando expressadas, cada uma das duas ou mais sequências-guia direciona a ligação específica da sequência 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 da polimerase III, ligado de modo funcional 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 tracr-mate quando se encontra idealmente alinhada. Em algumas formas de realização, a enzima CRISPR compreende uma ou mais sequências de localização nuclear de força suficiente para ativar 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 corresponde a Cas9 de *S. pneumoniae*, *S. pyogenes* ou *S. thermophilus*, e pode incluir Cas9 mutante derivada destes organismos. A enzima pode ser um homólogo ou ortólogo de 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 direciona a clivagem de um ou dois filamentos na localização da sequência-alvo. Em algumas formas de realização, a enzima CRISPR não tem atividade de clivagem de filamentos de ADN. Em algumas formas de realização, o primeiro elemento regulador é um promotor da polimerase III. Em algumas formas de realização, o segundo elemento regulador é um promotor da polimerase II. Em algumas formas de realização, a sequência-guia tem, pelo menos, 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10 e 30, ou entre 15 e 25, ou entre 15 e 20 nucleótidos de comprimento. Num aspeto, a divulgação fornece um organismo eucariótico não humano; preferencialmente um organismo eucariótico multicelular, compreendendo uma célula hospedeira eucariótica de acordo com qualquer uma das formas de realização descritas. Noutros aspetos, a divulgação fornece um organismo eucariótico; preferencialmente um organismo eucariótico multicelular, compreendendo uma célula hospedeira eucariótica de acordo com qualquer uma das formas de realização descritas. O organismo em algumas formas de realização destes aspetos pode ser um animal; por exemplo um mamífero. Igualmente, 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 divulgação fornece um *kit* compreendendo um ou mais dos componentes aqui descritos. Em algumas formas de realização, o *kit* compreende um sistema de vetores e instruções de utilização do *kit*. Em algumas

formas de realização, o sistema de vetores compreende (a) um primeiro elemento regulador ligado de modo funcional a uma sequência *tracr-mate* e a um ou mais locais de inserção para inserir uma ou mais sequências-guia a montante da sequência *tracr-mate*, em que, quando expressada, a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para uma sequência-alvo numa célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência *tracr-mate* que é hibridizada para a sequência *tracr*; e/ou (b) um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica a referida enzima CRISPR compreendendo uma sequência de localização nuclear. Em algumas formas de realização, o *kit* compreende os componentes (a) e (b) situados nos mesmos ou em diferentes vetores do sistema. Em algumas formas de realização, o componente (a) compreende ainda a sequência *tracr* a jusante da sequência *tracr-mate* sob o controlo do primeiro elemento regulador. Em algumas formas de realização, o componente (a) compreende ainda duas ou mais sequências-guia ligadas de modo funcional ao primeiro elemento regulador, em que, quando expressadas, cada uma das duas ou mais sequências-guia direciona a ligação específica da sequência de um complexo CRISPR para uma sequência-alvo diferente numa célula eucariótica. Em algumas formas de realização, o sistema compreende ainda um terceiro elemento regulador, tal como um promotor da polimerase III, ligado de modo funcional 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 *tracr-mate* quando se encontra idealmente alinhada. Em algumas formas de realização, a enzima CRISPR compreende uma ou mais sequências de

localização nuclear de força suficiente para ativar 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 corresponde a Cas9 de *S. pneumoniae*, *S. pyogenes* ou *S. thermophilus*, e pode incluir Cas9 mutante derivada destes organismos. A enzima pode ser um homólogo ou ortólogo de 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 direciona a clivagem de um ou dois filamentos na localização da sequência-alvo. Em algumas formas de realização, a enzima CRISPR não tem atividade de clivagem de filamentos de ADN. Em algumas formas de realização, o primeiro elemento regulador é um promotor da polimerase III. Em algumas formas de realização, o segundo elemento regulador é um promotor da polimerase II. Em algumas formas de realização, a sequência-guia tem, pelo menos, 15, 16, 17, 18, 19, 20, 25 nucleótidos, ou entre 10 e 30, ou entre 15 e 25, ou entre 15 e 20 nucleótidos de comprimento.

Num aspeto, a divulgação fornece um método de modificação de um polinucleótido-alvo numa célula eucariótica. Em algumas formas de realização, o método compreende a permissão a um complexo CRISPR para o mesmo se ligar 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 complexada com uma sequência-guia hibridizada para uma sequência-alvo no referido polinucleótido-alvo, em que a referida sequência-guia se encontra ligada a uma sequência tracr-mate que, por sua vez, é hibridizada para uma sequência tracr. Em algumas

formas de realização, a referida clivagem compreende a clivagem de um ou dois filamentos na localização da sequência-alvo pela referida enzima CRISPR. Em algumas formas de realização, a referida clivagem resulta numa menor transcrição de um gene-alvo. Em algumas formas de realização, o método compreende ainda a reparação do referido polinucleótido-alvo clivado através da recombinação homóloga com um polinucleótido de modelo 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 em mais alterações de aminoácido numa proteína expressada a partir de um gene que compreende a sequência-alvo. Em algumas formas de realização, o método compreende ainda a distribuição de um ou mais vetores à referida célula eucariótica, em que um ou mais vetores ativam a expressão de uma ou mais entre: a enzima CRISPR, a sequência-guia ligada à sequência tracr-mate e a sequência tracr. Em algumas formas de realização, os referidos vetores são distribuídos à célula eucariótica num sujeito. Em algumas formas de realização, a referida modificação ocorre na referida célula eucariótica numa 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 a devolução da referida célula eucariótica e/ou células daí derivadas ao referido sujeito.

Num aspeto, a divulgação fornece um método de modificação de expressão de um polinucleótido numa célula eucariótica. Em algumas formas de realização, o método compreende a permissão a um complexo CRISPR para o mesmo se ligar ao polinucleótido, de modo a que a referida ligação resulte numa maior ou menor expressão do referido

polinucleótido, em que o complexo CRISPR compreende uma enzima CRISPR complexada com uma sequência-guia hibridizada para uma sequência-alvo no referido polinucleótido, em que a referida sequência-guia se encontra ligada a uma sequência tracr-mate que, por sua vez, é hibridizada para uma sequência tracr. Em algumas formas de realização, o método compreende ainda a distribuição de um ou mais vetores às referidas células eucarióticas, em que um ou mais vetores ativam a expressão de uma ou mais entre: a enzima CRISPR, a sequência-guia ligada à sequência tracr-mate e a sequência tracr.

Num aspeto, a divulgação fornece um método de geração de uma célula eucariótica modelo compreendendo um gene da doença mutante. Em algumas formas de realização, um gene da doença corresponde a qualquer gene associado a um aumento no risco de ter ou desenvolver uma doença. Em algumas formas de realização, o método compreende (a) a introdução de um ou mais vetores numa célula eucariótica, em que um ou mais vetores ativam a expressão de uma ou mais entre: uma enzima CRISPR, uma sequência-guia ligada a uma sequência tracr-mate e uma sequência tracr; e (b) a permissão a um complexo CRISPR para o mesmo se ligar a um polinucleótido-alvo para efetuar a clivagem do polinucleótido-alvo no referido gene da doença, em que o complexo CRISPR compreende a enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo no polinucleótido-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr, gerando assim uma célula eucariótica modelo compreendendo um gene da doença mutante. Em algumas formas de realização, a referida clivagem compreende a clivagem de um ou dois filamentos na localização da sequência-alvo pela referida enzima CRISPR. Em algumas formas de realização, a referida clivagem resulta numa menor transcrição de um gene-alvo. Em algumas formas de realização, o método compreende ainda a reparação

do referido polinucleótido-alvo clivado através da recombinação homóloga com um polinucleótido de modelo 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 em mais alterações de aminoácido numa expressão de proteínas a partir de um gene que compreende a sequência-alvo.

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

Num aspeto, a divulgação fornece um polinucleótido recombinante compreendendo uma sequência-guia a montante de uma sequência tracr-mate, em que a sequência-guia, quando expressada, direciona a ligação específica da sequência de um complexo CRISPR para 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 divulgação fornece um método de seleção de uma ou mais células procarióticas através da introdução

de uma ou mais mutações num gene numa ou em mais células procarióticas, o método compreendendo: a introdução de um ou mais vetores nas células procarióticas, em que um ou mais vetores ativam a expressão de um ou mais dos seguintes: uma enzima CRISPR, uma sequência-guia ligada a uma sequência tracr-mate, uma sequência tracr e um modelo de edição; em que o modelo de edição compreende uma ou mais mutações que anulam a clivagem da enzima CRISPR; a permissão da recombinação homóloga do modelo de edição com o polinucleótido-alvo na(s) célula(s) a selecionar; a permissão a um complexo CRISPR para o mesmo se ligar a um polinucleótido-alvo para efetuar a clivagem do polinucleótido-alvo no referido gene, em que o complexo CRISPR compreende a enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo no polinucleótido-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr, em que a ligação do complexo CRISPR ao polinucleótido-alvo induz a morte celular, permitindo assim uma ou mais células procarióticas nas quais foram introduzidas uma ou mais mutações para serem selecionadas. Numa forma de realização preferida, a enzima CRISPR é Cas9. Noutro aspeto da invenção, a célula a selecionar pode ser uma célula eucariótica. Os aspetos da invenção permitem a seleção de células específicas sem necessitar de um marcador de seleção ou um processo de duas etapas que possa incluir um sistema de seleção inversa.

Em alguns aspetos, a divulgação fornece uma composição produzida ou que ocorre de forma não natural compreendendo uma sequência de polinucleótidos de ARN quimérico (ARNqui) de sistema CRISPR-Cas, em que a sequência de polinucleótidos compreende (a) uma sequência-guia capaz de hibridização para uma sequência-alvo numa célula eucariótica, (b) uma sequência tracr-mate, e (c) uma sequência tracr, em que (a), (b) e (c) são dispostos numa orientação de 5' para 3', em que, quando transcrita, a

sequência tracr-mate é hibridizada para a sequência tracr, e a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para a sequência-alvo, em que o complexo CRISPR compreende uma enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr,

ou

um sistema de enzima CRISPR, em que o sistema é codificado por um sistema de vetores compreendendo um ou mais vetores compreendendo I. um primeiro elemento regulador ligado de modo funcional a uma sequência de polinucleótidos de ARN quimérico (ARNqui) de sistema CRISPR-Cas, em que a sequência de polinucleótidos compreende (a) uma ou mais sequências-guia capazes de hibridização para uma ou mais sequências-alvo numa célula eucariótica, (b) uma sequência tracr-mate, e (c) uma ou mais sequências tracr, e II. um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica uma enzima CRISPR compreendendo, pelo menos, uma ou mais sequências de localização nuclear, em que (a), (b) e (c) são dispostas numa orientação de 5' para 3', em que os componentes I e II se encontram situados nos mesmos ou em diferentes vetores do sistema, em que, quando transcrita, a sequência tracr-mate é hibridizada para a sequência tracr, e a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para a sequência-alvo, em que o complexo CRISPR compreende a enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr, ou um sistema de enzima CRISPR multiplexado, em que o sistema é codificado por um sistema de vetores compreendendo um ou mais vetores compreendendo I. um primeiro elemento regulador ligado de modo funcional a (a) uma ou mais sequências-guia capazes de hibridização

para uma sequência-alvo numa célula, e (b) pelo menos uma ou mais sequências tracr-mate, II. um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica uma enzima CRISPR, e III. um terceiro elemento regulador ligado de modo funcional a uma sequência tracr, em que os componentes I, II e III se encontram situados nos mesmos ou em diferentes vetores do sistema, em que, quando transcrita, a sequência tracr-mate é hibridizada para a sequência tracr, e a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para a sequência-alvo, em que o complexo CRISPR compreende a enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência tracr, e em que, no sistema multiplexado, são utilizadas múltiplas sequências-guia e uma única sequência tracr; e em que uma ou mais das sequências-guia, tracr e tracr-mate são modificadas para melhorar a estabilidade.

Em aspetos da invenção, a modificação compreende uma estrutura secundária produzida. Por exemplo, a modificação pode compreender uma redução numa região de hibridização entre a sequência tracr-mate e a sequência tracr. Por exemplo, a modificação também pode compreender a fusão da sequência tracr-mate e da sequência tracr através de uma estrutura artificial. A modificação pode compreender a sequência tracr com um comprimento entre 40 e 120 pb. Nas formas de realização, a sequência tracr tem entre 40 pb e o comprimento total de tracr. Em determinadas formas de realização, o comprimento de ARNtrac inclui, pelo menos, nucleótidos 1-67 e, em algumas formas de realização, pelo menos, nucleótidos 1-85 do ARNtrac de tipo selvagem. Em algumas formas de realização, podem ser utilizados pelo menos os nucleótidos correspondentes aos nucleótidos 1-67 ou 1-85 de ARNtrac Cas9 de *S. pyogenes* de tipo selvagem. Quando o sistema CRISPR utiliza enzimas diferentes de Cas9,

ou diferentes de SpCas9, podem estar presentes os nucleótidos correspondentes no ARNtrac de tipo selvagem relevante. Em algumas formas de realização, o comprimento de ARNtrac não inclui mais do que os nucleótidos 1-67 ou 1-85 do ARNtrac de tipo selvagem. A modificação pode compreender a otimização da sequência. Em determinados aspectos, a otimização da sequência pode compreender a redução da incidência de sequências poliT na sequência tracr e/ou tracr-mate. A otimização da sequência pode ser combinada com a redução na região de hibridização entre a sequência tracr-mate e a sequência tracr; por exemplo, uma sequência tracr de comprimento reduzido.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou o sistema de enzima CRISPR, em que a modificação compreende a redução nas sequências poliT na sequência tracr e/ou tracr-mate. Em alguns aspectos, um ou mais Ts presentes numa sequência poli-T da sequência de tipo selvagem relevante (ou seja, um alongamento superior a 3, 4, 5, 6 ou mais bases T contíguas; em algumas formas de realização, um alongamento não superior a 10, 9, 8, 7, 6 bases T contíguas) podem ser substituídos por um nucleótido não T, p. ex., um A, de modo a que a série seja decomposta em alongamentos mais pequenos de Ts com cada alongamento tendo 4, ou menos de 4 (por exemplo, 3 ou 2) Ts contíguos. É possível utilizar bases diferentes de A para a substituição, por exemplo C ou G, ou nucleótidos que ocorrem de forma não natural ou nucleótidos modificados. Se a série de Ts estiver envolvida na formação de um gancho (*hairpin*) (ou estrutura em pedúnculo (*stem loop*)), é vantajoso que a base complementar para a base não T seja alterada para complementar o nucleótido não T. Por exemplo, se a base não T for um A, o respetivo complemento pode ser alterado para um T, p. ex., para preservar ou ajudar na preservação da estrutura secundária. Por exemplo, 5'-TTTTT pode ser alterado para se tornar 5'-TTTAT e o 5'-AAAAA

complementar pode ser alterado para 5'-ATAAA.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende a adição de uma sequência de terminador poliT. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou o sistema de enzima CRISPR, em que a modificação compreende a adição de uma sequência de terminador poliT nas sequências tracr e/ou tracr-mate. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende a adição de uma sequência de terminador poliT na sequência-guia. A sequência de terminador poliT pode compreender 5 bases T contíguas, ou mais de 5.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende a alteração de estruturas (*loops*) e/ou ganchos (*hairpins*). Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende o fornecimento de um mínimo de dois ganchos na sequência-guia. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou o sistema de enzima CRISPR, em que a modificação compreende o fornecimento de um gancho formado mediante complementação entre a sequência tracr e tracr-mate (repetição direta). Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou o sistema de enzima CRISPR, em que a modificação compreende o fornecimento de um ou mais ganchos na extremidade 3', ou em direção à mesma, da sequência de ARNtracr. Por exemplo, é possível formar um gancho fornecendo sequências autocomplementares na sequência ARNtrac unidas por uma estrutura, de modo a que um gancho seja formado no próprio dobramento. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende o fornecimento de ganchos suplementares adicionados ao 3' da sequência-guia. Num aspeto, a divulgação fornece o sistema

CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende o aumento da extremidade 5' da sequência-guia. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende o fornecimento de um ou mais ganchos na extremidade 5' da sequência-guia. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende a junção da sequência (5'-AGGACGAAGTCCTAA) à extremidade 5' da sequência-guia. Outras sequências adequadas para formar ganchos serão conhecidas do perito na técnica, e podem ser utilizadas em determinados aspetos da invenção. Em alguns aspetos, são fornecidos, pelo menos, 2, 3, 4, 5 ou mais ganchos adicionais. Em alguns aspetos, não são fornecidos mais de 10, 9, 8, 7, 6 ganchos adicionais. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende dois ganchos. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende três ganchos. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende, no máximo, cinco ganchos.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende o fornecimento de reticulação, ou o fornecimento de um ou mais nucleótidos modificados na sequência de polinucleótidos. Os nucleótidos modificados e/ou a reticulação podem ser fornecidos em qualquer uma ou em todas as sequências tracr, tracr-mate e/ou guia, e/ou na sequência de codificação de enzimas e/ou nas sequências de vetores. As modificações podem incluir a inclusão de, pelo menos, um nucleótido que ocorre de forma não natural, ou um nucleótido modificado, ou análogos dos mesmos. Os nucleótidos modificados podem ser modificados na metade de ribose, fosfato e/ou base. Os nucleótidos modificados podem

incluir análogos de 2'-O-metil, análogos de 2'-desoxi ou análogos de 2'-fluoro. A estrutura principal do ácido nucleico pode ser modificada, por exemplo, pode ser utilizada uma estrutura principal de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA - Locked Nucleic Acids) ou ácidos nucleicos com ponte (BNA - Bridged Nucleic Acids) também pode ser possível. Outros exemplos de bases modificadas incluem, mas não se limitam a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanossina.

Será compreendido que qualquer uma ou todas as modificações acima podem ser fornecidas separadamente ou em combinação num determinado sistema CRISPR-Cas ou sistema de enzima CRISPR. Esse sistema pode incluir uma, duas, três, quatro, cinco ou mais das referidas modificações.

Num aspeto, a invenção fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a enzima CRISPR é uma enzima do sistema CRISPR de tipo II, p. ex., uma enzima Cas9. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a enzima CRISPR é constituída por menos de mil aminoácidos, ou menos de quatro mil aminoácidos. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a enzima Cas9 corresponde a StCas9 ou StlCas9, ou a enzima Cas9 corresponde a uma enzima Cas9 de um organismo selecionado a partir do grupo que consiste em *genus Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* ou *Corynebacterium*. Num aspeto, a invenção fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a enzima CRISPR é uma nuclease que direciona a clivagem de ambos os filamentos na localização da sequência-alvo.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas

ou sistema de enzima CRISPR, em que o primeiro elemento regulador é um promotor da polimerase III. Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que o segundo elemento regulador é um promotor da polimerase II.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a sequência-guia compreende, pelo menos, quinze nucleótidos.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a modificação compreende a sequência *tracr* otimizada e/ou o ARN de sequência-guia otimizada e/ou a estrutura de codobrimento de sequência *tracr* e/ou sequências *tracr*-mate e/ou estruturas secundárias de estabilização de sequência *tracr* e/ou sequência *tracr* com uma região reduzida de pares de bases e/ou elementos ARN fundidos de sequência *tracr*; e/ou no sistema multiplexado existem dois ARNs compreendendo um traçador e compreendendo uma diversidade de guias ou um ARN compreendendo uma diversidade de quiméricos.

Em certos aspetos, a arquitetura de ARN quimérico é ainda otimizada de acordo com os resultados de estudos de mutagénese. No ARN quimérico com dois ou mais ganchos (*hairpins*), as mutações na repetição direta proximal para estabilizar o gancho podem resultar na ablação da atividade do complexo CRISPR. As mutações na repetição direta distal para encurtar ou estabilizar o gancho podem não ter nenhum efeito na atividade do complexo CRISPR. A seleção aleatória da sequência na região de protuberância entre as repetições proximal e distal pode reduzir significativamente a atividade do complexo CRISPR. As alterações no único par de bases ou a seleção aleatória da sequência na região de ligante entre ganchos podem resultar na perda completa da atividade do complexo CRISPR. A estabilização do gancho dos ganchos distais que seguem o primeiro gancho após a sequência-guia pode resultar na manutenção ou no

melhoramento da atividade do complexo CRISPR. Em conformidade, nas formas de realização preferidas, a arquitetura de ARN quimérico pode ser ainda otimizada através da geração de um ARN quimérico mais pequeno que possa ser benéfico para as opções de administração terapêutica e outras utilizações, e isto pode ser alcançado alterando a repetição direta distal de modo a encurtar ou estabilizar o gancho. Noutras formas de realização preferidas, a arquitetura de ARN quimérico pode ser ainda otimizada estabilizando um ou mais dos ganchos distais. A estabilização dos ganchos pode incluir a modificação de sequências adequadas para a formação de ganchos. Em alguns aspetos, são fornecidos, pelo menos, 2, 3, 4, 5 ou mais ganchos adicionais. Em alguns aspetos, não são fornecidos mais de 10, 9, 8, 7, 6 ganchos adicionais. Em alguns aspetos, a estabilização pode ser a reticulação e outras modificações. As modificações podem incluir a inclusão de, pelo menos, um nucleótido que ocorre de forma não natural, ou um nucleótido modificado, ou análogos dos mesmos. Os nucleótidos modificados podem ser modificados na metade de ribose, fosfato e/ou base. Os nucleótidos modificados podem incluir análogos de 2'-O-metil, análogos de 2'-desoxi ou análogos de 2'-fluoro. A estrutura principal do ácido nucleico pode ser modificada, por exemplo, pode ser utilizada uma estrutura principal de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos com ponte (BNA) também pode ser possível. Outros exemplos de bases modificadas incluem, mas não se limitam a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina.

Num aspeto, a divulgação fornece o sistema CRISPR-Cas ou sistema de enzima CRISPR, em que a enzima CRISPR tem codões otimizados para a expressão numa célula eucariótica.

Em conformidade, em alguns aspetos, o comprimento do ARN_{trac} necessário numa construção da invenção, p. ex., uma

construção quimérica, não necessita necessariamente de ser fixo, e em alguns aspetos pode ter entre 40 e 120 pb, e em alguns aspetos até ao comprimento total do tracr, p. ex., em alguns aspetos até à extremidade 3' de tracr conforme realçado pelo sinal de terminação de transcrição no genoma bacteriano. Em determinadas formas de realização, o comprimento de ARNtrac inclui, pelo menos, nucleótidos 1-67 e, em algumas formas de realização, pelo menos, nucleótidos 1-85 do ARNtrac de tipo selvagem. Em algumas formas de realização, podem ser utilizados, pelo menos, os nucleótidos correspondentes aos nucleótidos 1-67 ou 1-85 de ARNtracr Cas9 de *S. pyogenes* de tipo selvagem. Quando o sistema CRISPR utiliza enzimas diferentes de Cas9, ou diferentes de SpCas9, podem estar presentes os nucleótidos correspondentes no ARNtrac de tipo selvagem relevante. Em algumas formas de realização, o comprimento de ARNtrac não inclui mais do que os nucleótidos 1-67 ou 1-85 do ARNtrac de tipo selvagem. Em relação à otimização da sequência (p. ex., redução nas sequências poliT), p. ex., relativamente às séries de Ts internos na tracr-mate (repetição direta) ou ARNtracr, em alguns aspetos, um ou mais Ts presentes numa sequência poli-T da sequência de tipo selvagem relevante (ou seja, um alongamento superior a 3, 4, 5, 6 ou mais bases T contíguas; em algumas formas de realização, um alongamento não superior a 10, 9, 8, 7, 6 bases T contíguas) podem ser substituídos por um nucleótido não T, p. ex., um A, de modo a que a série seja decomposta em alongamentos mais pequenos de Ts com cada alongamento tendo 4, ou menos de 4 (por exemplo, 3 ou 2) Ts contíguas. Se a série de Ts estiver envolvida na formação de um gancho (*hairpin*) (ou estrutura em pedúnculo (*stem loop*)), é vantajoso que a base complementar para a base não T seja alterada para complementar o nucleótido não T. Por exemplo, se a base não T for um A, o respetivo complemento pode ser alterado para um T, p. ex., para preservar ou ajudar na

preservação da estrutura secundária. Por exemplo, 5'-TTTTT pode ser alterado para se tornar 5'-TTTAT e o 5'-AAAAA complementar pode ser alterado para 5'-ATAAA. Quanto à presença de sequências de terminador poliT na transcrição tracr + tracr-mate, p. ex., um terminador poliT (TTTTT ou mais), em alguns aspectos é vantajoso que tal seja adicionado no final da transcrição, independentemente de se encontrar na forma de dois ARN (tracr e tracr-mate) ou de ARN-guia único. Relativamente às estruturas e aos ganchos nas transcrições tracr e tracr-mate, em alguns aspectos é vantajoso que um mínimo de dois ganchos esteja presente no ARN-guia quimérico. Um primeiro gancho pode ser o gancho formado mediante complementação entre a sequência tracr e tracr-mate (repetição direta). Um segundo gancho pode encontrar-se na extremidade 3' da sequência de ARNtracr, e isto pode fornecer uma estrutura secundária para a interação com Cas9. Os ganchos adicionais podem ser adicionados à 3' do ARN-guia, p. ex., em alguns aspectos da invenção para aumentar a estabilidade do ARN-guia. Adicionalmente, a extremidade 5' do ARN-guia, em alguns aspectos, pode ser aumentada. Em alguns aspectos, é possível considerar 20 pb na extremidade 5' como uma sequência-guia. A parte 5' pode ser aumentada. Um ou mais ganchos podem ser fornecidos na parte 5', p. ex., em alguns aspectos, isto também pode melhorar a estabilidade do ARN-guia. Em alguns aspectos, o gancho específico pode ser fornecido acrescentando a sequência (5'-AGGACGAAGTCCTAA) à extremidade 5' da sequência-guia e, em alguns aspectos, isto pode ajudar a melhorar a estabilidade. Outras sequências adequadas para formar ganchos serão conhecidas do perito na técnica, e podem ser utilizadas em determinados aspectos. Em alguns aspectos, são fornecidos, pelo menos, 2, 3, 4, 5 ou mais ganchos adicionais. Em alguns aspectos, não são fornecidos mais de 10, 9, 8, 7, 6 ganchos adicionais. O apresentado anteriormente também fornece aspectos que

envolvem a estrutura secundária nas sequências-guia. Em alguns aspetos podem existir a reticulação e outras modificações, p. ex., para melhorar a estabilidade. As modificações podem incluir a inclusão de, pelo menos, um nucleótido que ocorre de forma não natural, ou um nucleótido modificado, ou análogos dos mesmos. Os nucleótidos modificados podem ser modificados na metade de ribose, fosfato e/ou base. Os nucleótidos modificados podem incluir análogos de 2'-O-metil, análogos de 2'-desoxi ou análogos de 2'-fluoro. A estrutura principal do ácido nucleico pode ser modificada, por exemplo, pode ser utilizada uma estrutura principal de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos com ponte (BNA) também pode ser possível. Outros exemplos de bases modificadas incluem, mas não se limitam a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina. Essas modificações ou reticulação podem estar presentes na sequência-guia ou noutras sequências adjacentes à sequência-guia.

BREVE DESCRIÇÃO DAS FIGURAS

Uma melhor compreensão das funcionalidades e vantagens da presente invenção será obtida através da consulta da seguinte descrição detalhada que expõe formas de realização ilustrativas, nas quais os princípios da invenção são utilizados, e das figuras em anexo apresentadas em seguida.

A **Figura 1** ilustra um modelo esquemático do sistema CRISPR. A Cas9 nuclease de *Streptococcus pyogenes* (amarela) é escolhida como alvo para o ADN genómico através de um ARN-guia sintético (ARNgs) que consiste numa sequência-guia 20-nt (azul) e numa armação (vermelha). Os pares de bases da sequência-guia com o alvo de ADN (azul), diretamente a montante de um *protospacer* (protoseparador)-motivo adjacente (PAM - Protospacer Adjacent Motif; magenta) 5'-NGG requerido, e Cas9 medeiam uma quebra de filamento duplo (DSB - Double-

Stranded Break) ~3 pb a montante do PAM (triângulo vermelho).

As **Figuras 2A a F** ilustram um sistema CRISPR exemplar, um mecanismo possível de ação, uma adaptação de exemplo para a expressão nas células eucarióticas, e resultados de testes que avaliam a localização nuclear e a atividade CRISPR.

As **Figuras 3A a C** ilustram uma cassete de expressão exemplar para a expressão de elementos do sistema CRISPR nas células eucarióticas, estruturas previstas de sequências-guia de exemplo, e a atividade do sistema CRISPR conforme medido nas células eucarióticas e procarióticas.

As **Figuras 4A a D** ilustram resultados de uma avaliação de especificidade SpCas9 para um alvo de exemplo.

As **Figuras 5A a G** ilustram um sistema de vetores exemplar e os resultados para a respetiva utilização no direcionamento da recombinação homóloga nas células eucarióticas.

As **Figuras 6A a C** ilustram uma comparação de transcrições ARNtracr diferentes para escolha de genes-alvo mediados por Cas9.

As **Figuras 7A a D** ilustram um sistema CRISPR exemplar, uma adaptação de exemplo para a expressão nas células eucarióticas, e os resultados de testes que avaliam a atividade CRISPR.

As **Figuras 8A a C** ilustram manipulações exemplares de um sistema CRISPR para a escolha dos *loci* genómicos-alvo nas células de mamífero.

As **Figuras 9A e B** ilustram os resultados de uma análise por *Northern blot* de processamento ARNcr nas células de mamífero.

As **Figuras 10A a C** ilustram uma representação esquemática de ARNs quiméricos e os resultados de ensaios SURVEYOR relativamente à atividade do sistema CRISPR nas células

eucarióticas.

As **Figuras 11A e B** ilustram uma representação gráfica dos resultados de ensaios SURVEYOR relativamente à atividade do sistema CRISPR nas células eucarióticas.

A **Figura 12** ilustra estruturas secundárias previstas para ARNs quiméricos exemplares compreendendo uma sequência-guia, sequência tracr-mate e sequência tracr.

A **Figura 13** é uma árvore filogenética dos genes Cas.

As **Figuras 14A a F** ilustram a análise filogenética que revela cinco famílias de Cas9s, incluindo três grupos de Cas9s grandes (~1400 aminoácidos) e dois de Cas9s pequenas (~1100 aminoácidos).

A **Figura 15** ilustra um gráfico que representa a função de diferentes ARNs-guia otimizados.

A **Figura 16** ilustra a sequência e a estrutura de diferentes ARNs quiméricos de guia.

A **Figura 17** ilustra a estrutura de codobrimento da repetição direta e ARNtracr.

As **Figuras 18A e B** ilustram dados da otimização de ARN-guia quimérico de StlCas9 *in vitro*.

As **Figuras 19A e B** ilustram a clivagem de alvos metilados ou não metilados através do lisado de célula SpCas9.

As **Figuras 20A a G** ilustram a otimização da arquitetura ARN-guia para a edição do genoma de mamífero mediado por SpCas9. (a) Esquema do vetor de expressão bicistrónico (PX330) para ARN-guia único (ARNgu) ativado pelo promotor A6, e Cas9 humana de *Streptococcus pyogenes* com codões otimizados ativada pelo promotor CBh utilizada para todas as experiências subsequentes. O ARNgu consiste numa sequência-guia 20-nt (azul) e numa armação (vermelha), truncadas em várias posições conforme indicado. (b) O ensaio SURVEYOR para indels (inserções/deleções) mediadas por SpCas9 nos loci de *PVALB* e *EMX1* humanos. As setas indicam os fragmentos SURVEYOR esperados ($n = 3$). (c) A análise por *Northern blot* relativamente às quatro

arquiteturas de truncamento ARN_{gu}, com U1 como controle de carregamento. (d) Tanto o mutante de tipo selvagem (wt) como de entalhadura (*nickase*) (D10A) de SpCas9 provocaram a inserção de um local *HindIII* no gene *EMX1* humano. Os oligonucleótidos de filamento único (ssODNs), orientados na direção *sense* ou *antisense* relativa à sequência de genoma, foram utilizados como modelos de recombinação homólogos. (e) Esquema do *locus* de *SERPINB5* humano. Os ARN_{sgu} e PAMs são indicados por barras coloridas por cima da sequência; a metilcitosina (Me) é realçada (cor-de-rosa) e numerada em relação ao local de início de transcrição (TSS, +1). (f) Estado de metilação de *SERPINB5* ensaiado por sequenciação de bissulfato de 16 clones. Círculos preenchidos, CpG metilado; círculos abertos, CpG não metilado. (g) Eficiência de modificação por três ARN_{sgu} que escolhem a região metilada-alvo de *SERPINB5*, ensaiada através de sequenciação profunda ($n = 2$). As barras de erro indicam intervalos de Wilson (Métodos Online).

As **Figuras 21A e B** ilustram a outra otimização da arquitetura ARN_{gu} CRISPR-Cas. (a) Esquema de quatro arquiteturas ARN_{gu} adicionais, I-IV. Cada uma consiste numa sequência-guia 20-nt (azul) unida à repetição direta (DR, cinzento), que é hibridizada para o ARN_{tracr}, (vermelho). O híbrido DR-ARN_{tracr} é truncado em +12 ou +22, conforme indicado, com uma estrutura em pedúnculo (*stem loop*) GAAA artificial. As posições de truncamento ARN_{tracr} são numeradas de acordo com o local de início de transcrição anteriormente apresentado para ARN_{tracr}. As arquiteturas ARN_{gu} II e IV transportam mutações nos respectivos tratos poli-U, que podem servir como terminadores de transcrição prematuros. (b) Ensaio SURVEYOR para as indels mediadas por SpCas9 no *locus* de *EMX1* humano para locais-alvo 1-3. As setas indicam os fragmentos SURVEYOR esperados ($n = 3$).

A **Figura 22** ilustra a visualização de alguns locais-alvo no genoma humano.

As **Figuras 23A e B** ilustram (A) um esquema do ARNgu e (B) a análise SURVEYOR de cinco variantes ARNgu para SaCas9 relativamente a uma arquitetura truncada ideal com uma eficiência de clivagem mais elevada.

Neste caso, as figuras destinam-se apenas a fins ilustrativos e não se encontram necessariamente desenhadas à escala.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Os termos "polinucleótido", "nucleótido", "sequência de nucleótidos", "ácido nucleico" e "oligonucleótido" são utilizados alternadamente. Os mesmos referem-se a uma forma polimérica de nucleótidos de qualquer comprimento, quer desoxirribonucleótidos ou ribonucleótidos ou análogos dos mesmos. Os polinucleótidos podem ter qualquer estrutura tridimensional, e podem efetuar qualquer função, conhecida ou desconhecida. Em seguida, são apresentados exemplos não limitativos de polinucleótidos: regiões de codificação ou não codificação de um gene ou fragmento de gene, *loci* (*locus*) definidos a partir da análise de ligação, exões, intrões, ARN mensageiro (ARNm), ARN de transferência, ARN do ribossoma, pequeno ARN de interferência (pARNi), ARN de gancho pequeno (ARNgp), micro-ARN (miARN), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plasmídeos, vetores, ADN isolado de qualquer sequência, ARN 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 nucleótido. Se presentes, as modificações na estrutura de nucleótidos podem ser transmitidas antes ou depois da montagem do polímero. A sequência de nucleótidos pode ser interrompida por componentes que não são nucleótidos. Um polinucleótido pode ainda ser modificado após a polimerização, tal como mediante conjugação com um

componente de rotulação.

Nos aspetos da invenção, os termos "ARN quimérico", "ARN-guia quimérico", "ARN-guia" "ARN-guia único" e "ARN-guia sintético" são utilizados alternadamente e referem-se à sequência de polinucleótidos que compreende a sequência-guia, a sequência tracr e a sequência tracr-mate. O termo "sequência-guia" refere-se à sequência de cerca de 20 pb no ARN-guia que especifica o local-alvo e pode ser utilizado alternadamente com os termos "guia" ou "separador". O termo "sequência tracr-mate" também pode ser utilizado alternadamente com o termo "repetição direta".

Conforme aqui utilizado, o termo "tipo selvagem" é um termo da técnica compreendido pelos peritos na técnica e significa a forma típica de um organismo, estirpe, gene ou característica como ocorre na natureza distinguindo-se das formas mutantes ou variantes.

Conforme aqui utilizado, o significado do termo "variante" deve ser considerado como a apresentação de qualidades que têm um padrão que se desvia do que ocorre na natureza.

Os termos "que ocorre de forma não natural" ou "produzido" são utilizados alternadamente e indicam o envolvimento da mão do homem. Os termos, quando se referem a moléculas de ácido nucleico ou polipéptidos, significam que a molécula de ácido nucleico ou o polipéptido se encontra, pelo menos, substancialmente livre de, pelo menos, um outro componente ao qual está naturalmente associado na natureza e conforme encontrado na natureza.

"Complementaridade" refere-se à capacidade de um ácido nucleico formar ligações de hidrogénio a outra sequência de ácidos nucleicos através dos pares de bases Watson-Crick tradicionais ou outros tipos não tradicionais. Uma percentagem de complementaridade indica que a percentagem de resíduos numa molécula de ácido nucleico que pode formar ligações de hidrogénio (p. ex., pares de bases Watson-

Crick) a uma segunda sequência de ácidos nucleicos (p. ex., 5, 6, 7, 8, 9, 10 de 10 sendo 50%, 60%, 70%, 80%, 90% e 100% complementares). "Perfeitamente complementar" significa que todos os resíduos contíguos de uma sequência de ácidos nucleicos irão ligar-se por hidrogénio ao mesmo número de resíduos contíguos numa segunda sequência de ácidos nucleicos. "Substancialmente complementar" conforme aqui utilizado refere-se a um grau de complementaridade que é, pelo menos, de 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% ou 100% numa 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 são hibridizados em condições rigorosas.

Conforme aqui utilizado, a expressão "condições rigorosas" para a hibridização refere-se às condições nas quais um ácido nucleico com complementaridade com uma sequência-alvo é hibridizado predominantemente com a sequência-alvo, e não é substancialmente hibridizado para sequências sem ser alvo. As condições rigorosas são geralmente dependentes da sequência, e variam dependendo de uma quantidade de fatores. Em geral, quanto maior for a sequência, mais alta será a temperatura na qual a sequência é hibridizada especificamente para a respetiva sequência-alvo. Os 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*, Segundo Capítulo "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, Nova Iorque.

"Hibridização" refere-se a uma reação na qual um ou mais polinucleótidos reagem para formar um complexo que é estabilizado através de ligação de hidrogénio entre as bases dos resíduos de nucleótido. A ligação de hidrogénio pode ocorrer através de pares de bases Watson-Crick,

ligação Hoogstein, ou de qualquer outra maneira específica da sequência. O complexo pode compreender dois filamentos que formam uma estrutura dupla, três ou mais filamentos que formam um complexo de múltiplos filamentos, um filamento único de auto-hibridização, ou qualquer combinação dos mesmos. Uma reação de hibridização pode constituir uma etapa num processo mais extenso, tal como a iniciação de PCR, ou a clivagem de um polinucleótido por uma enzima. Uma sequência capaz de hibridizar com uma determinada sequência é referida como o "complemento" da sequência específica.

Conforme aqui utilizado, "estabilização" ou "estabilidade crescente" relativamente aos componentes do sistema CRISPR referem-se à fixação ou estabilização da estrutura da molécula. Isto pode ser realizado através da introdução de uma ou mais mutações, incluindo uma única ou múltiplas alterações de pares de bases, o aumento do número de ganchos, a reticulação, a quebra de alongamentos específicos de nucleótidos e outras modificações. As modificações podem incluir a inclusão de, pelo menos, um nucleótido que ocorre de forma não natural, ou um nucleótido modificado, ou análogos dos mesmos. Os nucleótidos modificados podem ser modificados na metade de ribose, fosfato e/ou base. Os nucleótidos modificados podem incluir análogos de 2'-O-metil, análogos de 2'-desoxi ou análogos de 2'-fluoro. A estrutura principal do ácido nucleico pode ser modificada, por exemplo, pode ser utilizada uma estrutura principal de fosforotioato. A utilização de ácidos nucleicos bloqueados (LNA) ou ácidos nucleicos com ponte (BNA) também pode ser possível. Outros exemplos de bases modificadas incluem, mas não se limitam a, 2-aminopurina, 5-bromo-uridina, pseudouridina, inosina, 7-metilguanosina. Estas modificações podem aplicar-se a qualquer componente do sistema CRISPR. Numa forma de realização preferida, estas modificações são efetuadas nos componentes ARN, p. ex. o ARN-guia ou a sequência de

polinucleótidos quiméricos.

Conforme aqui utilizado, "expressão" refere-se ao processo através do qual um polinucleótido é transcrito a partir de um modelo de ADN (tal como num ARNm ou outra transcrição ARN) e/ou ao processo através do qual um ARNm transcrito é subsequentemente traduzido em péptidos, polipéptidos ou proteínas. As transcrições e os polipéptidos codificados podem ser coletivamente referidos como "produto de gene". Se o polinucleótido for obtido a partir de ADN genómico, a expressão pode incluir o *splicing* do ARNm numa célula eucariótica.

Os termos "polipéptido", "péptido" e "proteína" são aqui utilizados alternadamente para se referirem a polímeros de aminoácidos de qualquer comprimento. O polímero pode ser linear ou ramificado, pode compreender aminoácidos modificados, e pode ser interrompido por não-aminoácidos. Os termos abrangem igualmente um polímero de aminoácido que foi modificado; por exemplo, formação de ligações dissulfureto, glicosilação, lipidação, acetilação, fosforilação, ou qualquer outra manipulação, tal como a conjugação com um componente de rotulação. Conforme aqui utilizado, o termo "aminoácido" inclui aminoácidos naturais e/ou não naturais ou sintéticos, incluindo glicina e os isómeros óticos D ou L, e análogos de aminoácido e peptidomiméticos.

Os termos "sujeito", "indivíduo" e "paciente" são aqui utilizados alternadamente para se referirem a um vertebrado, preferencialmente um mamífero, mais preferencialmente, um ser humano. Os mamíferos incluem, mas não se limitam a, ratinhos, símios, seres humanos, animais de criação, animais de desporto e animais de estimação. Os tecidos, as células e a respetiva progenitura de uma entidade biológica obtida *in vivo* ou cultivada *in vitro* são igualmente abrangidos. Em algumas formas de realização, um sujeito pode ser um animal invertebrado, por exemplo, um

inseto ou um nematode; enquanto noutras, um sujeito pode ser uma planta ou um fungo.

Os termos "agente terapêutico", "agente terapêutico capaz" ou "agente de tratamento" são utilizados alternadamente e referem-se a uma molécula ou um composto que confere algum efeito benéfico na administração num sujeito. O efeito benéfico inclui a permissão de determinações de diagnóstico; o melhoramento de uma doença, sintoma, distúrbio ou condição patológica; reduzindo ou prevenindo o início de uma doença, sintoma, distúrbio ou condição; e neutralizando geralmente uma doença, um sintoma, um distúrbio ou uma condição patológica.

Conforme aqui utilizado, "tratamento" ou "tratar", ou "atenuar" ou "melhorar" são utilizados alternadamente. Estes termos referem-se a uma abordagem para obter resultados benéficos ou desejados incluindo, mas não se limitando a, um benefício terapêutico e/ou um benefício profilático. A expressão "benefício terapêutico" significa qualquer efeito ou melhoramento terapeuticamente relevante numa ou em mais doenças, condições ou sintomas sob tratamento. Para o benefício profilático, as composições podem ser administradas num sujeito em risco de desenvolver uma determinada doença, condição ou sintoma, ou num sujeito que apresenta um ou mais dos sintomas fisiológicos de uma doença, mesmo que a doença, a condição ou o sintoma ainda não se tenha manifestado.

O termo "quantidade eficaz" ou "quantidade terapeuticamente eficaz" refere-se à quantidade de um agente que é suficiente para efetuar resultados benéficos ou desejados. A quantidade terapeuticamente eficaz pode variar dependendo de um ou mais dos fatores apresentados em seguida: o sujeito e a condição da doença que estão a ser tratados, o peso e a idade do sujeito, a gravidade da condição da doença, o modo de administração e afins, que podem ser prontamente determinados por um perito na

técnica. O termo aplica-se igualmente a uma dose que irá fornecer uma imagem para detecção através de qualquer um dos métodos de formação de imagens aqui descritos. A dose específica pode variar dependendo de um ou mais dos fatores apresentados em seguida: o agente específico escolhido, o regime de dosagem a ser seguido, se é administrado em conjunto com outros compostos, as horas de administração, o tecido do qual criar uma imagem, e o sistema de distribuição física no qual é transportado.

A prática da presente invenção utiliza, salvo indicação em contrário, técnicas convencionais de imunologia, bioquímica, química, biologia molecular, microbiologia, biologia celular, genómica e ADN recombinante, que fazem parte 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, and ANIMAL CELL CULTURE" (R.I. Freshney, ed. (1987)).

Diversos aspetos da invenção referem-se a sistemas de vetores compreendendo um ou mais vetores, ou vetores como tal. Os vetores podem ser concebidos para a expressão de transcrições CRISPR (p. ex. transcrições de ácido nucleico, proteínas, ou enzimas) nas células procarióticas ou eucarióticas. Por exemplo, as transcrições CRISPR podem ser expressadas nas células bacterianas, tais como *Escherichia coli*, células de inseto (utilizando vetores de expressão de baculovírus), células de levedura ou células de mamífero. As células hospedeiras adequadas são mais descritas em Goeddel, "GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY" 185, Academic Press, San Diego, Calif. (1990). Em alternativa, o vetor de expressão recombinante pode ser

transcrito e traduzido *in vitro*, por exemplo utilizando sequências reguladoras de promotor T7 e polimerase T7.

Os vetores podem ser introduzidos e propagados num procariota. Em algumas formas de realização, um procariota é utilizado para ampliar cópias de um vetor que deve ser introduzido numa célula eucariótica ou como um vetor intermediário na produção de um vetor que deve ser introduzido numa célula eucariótica (p. ex. ampliando um plasmídeo como parte de um sistema de empacotamento de vetores virais). Em algumas formas de realização, um procariota é utilizado para ampliar cópias de um vetor e expressar um ou mais ácidos nucleicos, de modo a fornecer uma origem de uma ou mais proteínas para a distribuição numa célula hospedeira ou num organismo hospedeiro. A expressão de proteínas nos procariotas é, na maioria das vezes, realizada em *Escherichia coli* com vetores que contêm promotores constitutivos ou induzidos que direcionam a expressão de proteínas de fusão ou sem ser de fusão. Os vetores de fusão adicionam um número de aminoácidos a uma proteína aí codificada, tal como ao amino-terminal da proteína recombinante. Esses vetores de fusão podem servir uma ou mais finalidades, tais como: (i) para aumentar a expressão de proteína recombinante; (ii) para aumentar a solubilidade da proteína recombinante; e (iii) para ajudar na purificação da proteína recombinante ao funcionar como um ligando na purificação de afinidade. Muitas vezes, nos vetores de expressão de fusão, é introduzido um local de clivagem proteolítica na junção da metade de fusão e da proteína recombinante para permitir a separação da proteína recombinante da metade de fusão subsequente à purificação da proteína de fusão. Essas enzimas, e respectivas sequências de reconhecimento cognato, incluem Fator Xa, trombina e enteroquinase. Os vetores de expressão de fusão de exemplo incluem pGEX (Pharmacia Biotech Inc; Smith e Johnson, 1988. *Gene* 67: 31 a 40), pMAL (New England

Biolabs, Beverly, Mass.) e pRIT5 (Pharmacia, Piscataway, N.J.) que fundem glutathione-S-transferase (GST), proteína de ligação E à maltose, ou proteína A, respectivamente, com a proteína recombinante-alvo.

Os exemplos de vetores de expressão *E. coli* de não fusão induzidos adequados incluem pTrc (Amrann *et al.*, (1988) *Gene* 69:301 a 315) e pET 11d (Studier *et al.*, "GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY" 185, *Academic Press*, San Diego, Calif. (1990) 60 a 89).

Em algumas formas de realização, um vetor é um vetor de expressão de levedura. Os exemplos de vetores para a expressão em *Saccharomyces cerevisiae* de levedura incluem pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229 a 234), pMFa (Kuijan e Herskowitz, 1982. *Cell* 30: 933 a 943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113 a 123), pYES2 (Invitrogen Corporation, San Diego, Calif.), e picZ (Invitrogen Corp, San Diego, Calif.).

Em algumas formas de realização, um vetor ativa a expressão de proteínas nas células de inseto utilizando vetores de expressão de baculovírus. Os vetores de baculovírus disponíveis para a expressão de proteínas nas células de inseto cultivadas (p. ex., células SF9) incluem a série pAc (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156 a 2165) e a série pVL (Lucklow e Summers, 1989. *Virology* 170: 31 a 39).

Em algumas formas de realização, um vetor é capaz de ativar a expressão de uma ou mais sequências nas células de mamífero utilizando um vetor de expressão de mamífero. Os exemplos de vetores de expressão de mamífero incluem pCDM8 (Seed, 1987. *Nature* 329: 840) e pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). Quando utilizadas nas células de mamífero, as funções de controlo do vetor de expressão são habitualmente fornecidas por um ou mais elementos reguladores. Por exemplo, os promotores geralmente utilizados são obtidos a partir de políoma, adenovírus 2,

citomegalovírus, vírus simiano 40, e outros aqui divulgados e conhecidos na técnica. Relativamente a outros sistemas de expressão adequados tanto para as células procarióticas como eucarióticas, consulte, p. ex., 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, Nova Iorque, 1989.

Em algumas formas de realização, o vetor de expressão de mamífero recombinante é capaz de direcionar a expressão do ácido nucleico preferencialmente num determinado tipo de célula (p. ex., são utilizados elementos reguladores específicos do tecido para expressar o ácido nucleico). Os elementos reguladores específicos do tecido são conhecidos na técnica. Os exemplos não limitativos de promotores adequados específicos do tecido incluem o promotor de albumina (específico do fígado; Pinkert, et al., 1987. *Genes Dev.* 1: 268 a 277), promotores linfoides específicos (Calame e Eaton, 1988. *Adv. Immunol.* 43: 235 a 275), em particular promotores de recetores de células T (Winoto e Baltimore, 1989. *EMBD J.* 8: 729 a 733) e imunoglobulinas (Baneiji, et al., 1983. *Cell* 33: 729 a 740; Queen e Baltimore, 1983. *Cell* 33: 741 a 748), promotores específicos do neurónio (p. ex., o promotor de neurofilamento; Byrne e Ruddie, 1989. *Proc. Natl. Acad. Sci. E.U.A.* 86: 5473 a 5477), promotores específicos do pâncreas (Edlund, et al., 1985. *Science* 230: 912 a 916), e promotores específicos da glândula mamária (p. ex., promotor de soro de leite; Patente U.S. n.º 4.873.316 e Publicação de Pedido de Patente europeia n.º 264.166). Os promotores regulados pelo desenvolvimento são igualmente abrangidos, p. ex., os promotores hox murinos (Kessel e Gruss, 1990. *Science* 249: 374 a 379) e o promotor α -fetoproteína (Campes e Tilghman, 1989. *Genes Dev.* 3:537 a 546).

Em algumas formas de realização, um elemento regulador

é ligado de modo funcional a um ou mais elementos de um sistema CRISPR de modo a ativar a expressão de um ou mais elementos do sistema CRISPR. Em geral, as CRISPRs (Repetições Palindrômicas Curtas Agrupadas e Regularmente Espaçadas), igualmente conhecidas como SPIDRs (SPacer Interspersed Direct Repeats - Repetições Diretas e Interpoladas de Separador), constituem uma família de *loci* de ADN que são habitualmente específicos de uma determinada espécie bacteriana. O *locus* de CRISPR compreende uma classe distinta de SSRs (Interspersed Short Sequence Repeats - Repetições de Sequência Curtas e Interpoladas) que foram reconhecidas em *E. coli* (Ishino *et al.*, *J. Bacteriol.*, 169:5429 a 5433 [1987]; e Nakata *et al.*, *J. Bacteriol.*, 171:3553 a 3556 [1989]), e genes associados. As SSRs interpoladas semelhantes foram identificadas em *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena* e *Mycobacterium tuberculosis* (Consulte Groenen *et al.*, *Mol. Microbiol.*, 10:1057 a 1065 [1993]; Hoe *et al.*, *Emerg. Infect. Dis.*, 5:254 a 263 [1999]; Masepohl *et al.*, *Biochim. Biophys. Acta* 1307:26 a 30 [1996]; e Mojica *et al.*, *Mol. Microbiol.*, 17:85 a 93 [1995]). Os *loci* de CRISPR diferem habitualmente das outras SSRs na estrutura das repetições, que foram designadas por SRSRs (Short Regularly Spaced Repeats - Repetições Regularmente Espaçadas e Curtas) (Janssen *et al.*, *OMICS J. Integ. Biol.*, 6:23 a 33 [2002]; e Mojica *et al.*, *Mol. Microbiol.*, 36:244 a 246 [2000]). Em geral, as repetições são elementos pequenos que ocorrem em agrupamentos que são regularmente espaçados por sequências intervenientes exclusivas com um comprimento substancialmente constante (Mojica *et al.*, [2000], *supra*). Embora as sequências de repetição sejam altamente conservadas entre estirpes, o número de repetições interpoladas e as sequências das regiões de separador diferem habitualmente de estirpe para estirpe (van Embden *et al.*, *J. Bacteriol.*, 182:2393 a 2401 [2000]). Os *loci* de

CRISPR foram identificados em mais de 40 procariotas (Consulte, p. ex., Jansen et al., *Mol. Microbiol.*, 43:1565 a 1575 [2002]; e Mojica et al., [2005]) incluindo, mas não se limitando a, *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosa-rcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlombium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxo-coccus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photo-bacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* e *Thermotoga*.

Em geral, "sistema CRISPR" refere-se coletivamente a transcrições e outros elementos envolvidos na expressão, ou no direcionamento, da atividade de genes associados a CRISPR ("Cas"), incluindo sequências que codificam um gene Cas, uma sequência tracr (CRISPR de ativação de trans) (p. ex. ARNtracr ou um ARNtracr parcial ativo), uma sequência tracr-mate (abrangendo uma "repetição direta" e uma repetição direta parcial processada por ARNtracr no contexto de um sistema CRISPR endógeno), uma sequência-guia (igualmente referida como um "separador" no contexto de um sistema CRISPR endógeno), ou outras sequências e transcrições de um locus de CRISPR. Em algumas formas de realização, são obtidos um ou mais elementos de um sistema CRISPR a partir de um sistema CRISPR de tipo I, tipo II ou tipo III. Em algumas formas de realização, são obtidos um ou mais elementos de um sistema CRISPR a partir de um determinado organismo que compreende um sistema CRISPR endógeno, tal como *Streptococcus pyogenes*. Em geral, um sistema CRISPR é caracterizado por elementos que provocam a formação de um complexo CRISPR no local de uma sequência-

alvo (igualmente referida como um *protospacer* no contexto de um sistema CRISPR endógeno). No contexto de formação de um complexo CRISPR, "sequência-alvo" refere-se a uma sequência para a qual é concebida uma sequência-guia com complementaridade, em que a hibridização entre a sequência-alvo e a sequência-guia provoca a formação de um complexo CRISPR. A complementaridade total não é necessariamente requerida, desde que exista uma complementaridade suficiente para causar a hibridização e provocar a formação de um complexo CRISPR. Uma sequência-alvo pode compreender qualquer polinucleótido, tais como polinucleótidos ADN e ARN. Em algumas formas de realização, uma sequência-alvo encontra-se no núcleo ou citoplasma de uma célula. Em algumas formas de realização, a sequência-alvo pode encontrar-se num organelo de uma célula eucariótica, por exemplo, mitocôndria ou cloroplasto. Uma sequência ou um modelo que pode ser utilizado para a recombinação no *locus* escolhido como alvo compreendendo as sequências-alvo é referido como um "modelo de edição" ou "polinucleótido de edição" ou "sequência de edição". Nos aspetos da invenção, um polinucleótido de modelo exógeno pode ser referido como um modelo de edição. Num aspeto da invenção, a recombinação é uma recombinação homóloga.

Habitualmente, no contexto de um sistema CRISPR endógeno, a formação de um complexo CRISPR (compreendendo uma sequência-guia hibridizada para uma sequência-alvo e complexada com uma ou mais proteínas Cas) resulta na clivagem de um ou ambos os filamentos na, ou junto da, (p. ex. em 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, ou mais pares de base da) sequência-alvo. Sem ser pretendida qualquer limitação a uma teoria, a sequência *tracr*, que pode compreender ou consistir na totalidade ou numa parte de uma sequência *tracr* de tipo selvagem (p. ex. cerca de, ou mais de cerca de, 20, 26, 32, 45, 48, 54, 63, 67, 85 ou mais nucleótidos de uma sequência *tracr* de tipo selvagem),

também pode fazer parte de um complexo CRISPR, tal como por hibridização ao longo de, pelo menos, uma parte da sequência tracr para a totalidade ou uma parte de uma sequência tracr-mate que é ligada de modo funcional à sequência-guia. Em algumas formas de realização, a sequência tracr tem complementaridade suficiente para uma sequência tracr-mate ser hibridizada e participar na formação de um complexo CRISPR. Tal como com a sequência-alvo, parte-se do princípio de que a complementaridade completa 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 tracr-mate quando se encontra idealmente alinhada. Em algumas formas de realização, um ou mais vetores que ativam a expressão de um ou mais elementos de um sistema CRISPR são introduzidos numa célula hospedeira, de modo a que a expressão dos elementos do sistema CRISPR direcione a formação de um complexo CRISPR num ou em mais locais-alvo. Por exemplo, uma enzima Cas, uma sequência-guia ligada a uma sequência tracr-mate e uma sequência tracr podem, cada uma delas, ser ligadas de modo funcional para separar elementos reguladores em vetores separados. Em alternativa, dois ou mais dos elementos expressados a partir dos mesmos ou diferentes elementos reguladores podem ser combinados num único vetor, em que um ou mais vetores adicionais fornecem quaisquer componentes do sistema CRISPR não incluído no primeiro vetor. Os elementos do sistema CRISPR que são combinados num único vetor podem ser dispostos em qualquer orientação adequada, tal como um elemento situado em 5' em relação a ("a montante" de) ou em 3' em relação a ("a jusante" de) um segundo elemento. A sequência de codificação de um elemento pode estar situada no mesmo, ou no oposto, filamento da sequência de codificação de um segundo elemento, e orientada na mesma,

ou na oposta, direção. Em algumas formas de realização, um único promotor ativa a expressão de uma transcrição que codifica uma enzima CRISPR e uma ou mais entre a sequência-guia, sequência tracr-mate (opcionalmente ligada de modo funcional à sequência-guia) e uma sequência tracr incorporada numa ou em mais sequências de intrões (p. ex. cada uma num intrão diferente, duas ou mais pelo menos num intrão, ou todas num único intrão). Em algumas formas de realização, a enzima CRISPR, a sequência-guia, a sequência tracr-mate e a sequência tracr são ligadas de modo funcional ao mesmo promotor e expressadas a partir do mesmo.

Em algumas formas de realização, um vetor compreende um ou mais locais de inserção, tal como uma sequência de reconhecimento de endonuclease de restrição (igualmente referida como um "local de clonagem"). Em algumas formas de realização, um ou mais locais de inserção (p. ex. cerca de, ou mais de cerca de, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais locais de inserção) encontram-se situados a montante e/ou a jusante de um ou mais elementos de sequência de um ou mais vetores. Em algumas formas de realização, um vetor compreende um local de inserção a montante de uma sequência tracr-mate, e opcionalmente a jusante de um elemento regulador ligado de modo funcional à sequência tracr-mate, de modo a que, a seguir à inserção de uma sequência-guia no local de inserção e na expressão, a sequência-guia direcione a ligação específica da sequência de um complexo CRISPR para uma sequência-alvo numa célula eucariótica. Em algumas formas de realização, um vetor compreende dois ou mais locais de inserção, cada local de inserção situando-se entre duas sequências tracr-mate para permitir a inserção de uma sequência-guia em cada local. Numa disposição assim, 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 destas. Quando

são utilizadas múltiplas sequências-guia diferentes, é possível utilizar uma construção de expressão única para escolher a atividade CRISPR-alvo para múltiplas e diferentes sequências-alvo correspondentes numa célula. Por exemplo, um único vetor pode compreender cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 ou mais sequências-guia. Em algumas formas de realização, cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais desses vetores que contêm sequências-guia podem ser fornecidos e opcionalmente distribuídos a uma célula.

Em algumas formas de realização, um vetor compreende um elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas 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 (igualmente conhecidos 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 dos mesmos, ou versões modificadas dos mesmos. 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 ADN, tal como Cas9. Em algumas formas de realização, a enzima CRISPR é Cas9, e pode ser Cas9 de *S. pyogenes* ou *S. pneumoniae*. Em algumas formas de realização, a enzima CRISPR direciona a clivagem de um ou ambos os filamentos na localização de uma sequência-alvo, tal como na sequência-alvo e/ou no complemento da sequência-alvo. Em algumas formas de realização, a enzima CRISPR direciona a clivagem de um ou ambos os filamentos em cerca de 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 último nucleótido de uma sequência-alvo. Em algumas formas de realização, um vetor codifica uma enzima CRISPR que sofre uma mutação em relação a uma enzima de tipo selvagem correspondente, de modo a que a enzima CRISPR mutante não tenha a capacidade para clivar um ou ambos os filamentos de um polinucleótido-alvo que contém uma sequência-alvo. Por exemplo, uma substituição de aspartato por alanina (D10A) no domínio catalítico RuvC I de Cas9 de *S. pyogenes* converte Cas9 de uma nuclease que cliva ambos os filamentos numa *nickase* (enzima de entalhadura) (cliva um único filamento). Outros exemplos de mutações que tornam Cas9 numa *nickase* incluem, sem limitação, H840A, N854A e N863A. Em algumas formas de realização, uma *nickase* Cas9 pode ser utilizada em conjunto com sequências-guia, p. ex., duas sequências-guia, que escolhem como alvo respetivamente filamentos *sense* e *antisense* do alvo de ADN. Esta combinação permite que ambos os filamentos sejam entalhados (*nicked*) e utilizados para induzir NHEJ. Os requerentes demonstraram (dados não ilustrados) a eficácia de dois alvos de *nickase* (isto é, ARNsgu escolhidos como alvo na mesma localização, mas para filamentos diferentes de ADN) na indução de NHEJ mutagénico. Uma única *nickase* (Cas9-D10A com um único ARNsgu) não é capaz de induzir NHEJ e criar indels, mas os Requerentes mostraram que a *nickase* dupla (Cas9-D10A e dois ARNsgu escolhidos como alvo para filamentos diferentes na mesma localização) pode fazê-lo nas células estaminais embrionárias humanas (hESCs). A eficiência é de cerca de 50% de nuclease (isto é, Cas9 regular sem mutação D10) em hESCs.

Como um outro exemplo, dois ou mais domínios catalíticos de Cas9 (RuvC I, RuvC II e RuvC III) podem sofrer mutações para produzir uma Cas9 mutante que não tem substancialmente toda a atividade de clivagem de ADN. Em algumas formas de realização, uma mutação D10A é combinada

com uma ou mais das mutações H840A, N854A ou N863A para produzir uma enzima Cas9 que não tem substancialmente toda a atividade de clivagem de ADN. Em algumas formas de realização, uma enzima CRISPR é considerada como não tendo substancialmente toda a atividade de clivagem de ADN quando a atividade de clivagem de ADN da enzima mutante é inferior a cerca de 25%, 10%, 5%, 1%, 0,1%, 0,01% ou menos relativamente à respetiva forma não mutante. Outras mutações podem ser úteis; quando a Cas9 ou outra enzima CRISPR é de uma espécie diferente de *S. pyogenes*, é possível efetuar mutações nos aminoácidos correspondentes para alcançar efeitos semelhantes.

Em algumas formas de realização, uma sequência de codificação de enzimas que codifica uma enzima CRISPR tem codões otimizados para a expressão em determinadas células, tais como células eucarióticas. As células eucarióticas podem ser as de, ou as derivadas a partir de, um determinado organismo, tal com um mamífero, incluindo, mas não se limitando a, ser humano, ratinho, rato, coelho, cão ou primata não humano. Em geral, a otimização de codões refere-se a um processo de modificação de uma sequência de ácido nucleicos para uma expressão melhorada nas células hospedeiras de interesse substituindo, pelo menos, um codão (p. ex. cerca de, ou mais de cerca de, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 ou mais codões) da sequência nativa por codões que são mais frequentemente, ou os mais frequentemente, utilizados nos genes dessa célula hospedeira enquanto a sequência de aminoácidos nativa é mantida. Várias espécies apresentam uma determinada predisposição para certos codões de um aminoácido específico. A predisposição dos codões (diferenças na utilização de codões entre organismos) é muitas vezes correlacionada com a eficiência de tradução do ARN mensageiro (ARNm) que, por sua vez, se parte do princípio de que é dependente, entre outras coisas, das propriedades

dos codões que estão a ser traduzidos e da disponibilidade de determinadas moléculas de ARN de transferência (ARNt). O predomínio de ARNst selecionados numa célula é geralmente um reflexo dos codões utilizados com mais frequência na síntese de péptidos. Em conformidade, os genes podem ser ajustados para a expressão de genes ideal num determinado organismo com base na otimização de codões. As tabelas de utilização de codões estão facilmente disponíveis, for exemplo, na "Base de dados de Utilização de Codões" (Codon Usage Database), e estas tabelas podem ser adaptadas de várias maneiras. Consulte Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). Os algoritmos de computador para a otimização de codões de uma determinada sequência para a expressão numa determinada célula hospedeira estão igualmente disponíveis, tal como Gene Forge (Aptagen; Jacobus, PA). Em algumas formas de realização, um ou mais codões (p. ex. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 ou mais, ou todos os codões) numa sequência que codifica uma enzima CRISPR correspondem ao codão utilizado com mais frequência para um determinado aminoácido.

Em algumas formas de realização, um vetor codifica uma enzima CRISPR que compreende uma ou mais sequências de localização nuclear (NLSs), tal como cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais NLSs. Em algumas formas de realização, a enzima CRISPR compreende cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais NLSs no, ou junto ao, amino-terminal, cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais NLSs no, ou junto ao, carboxi-terminal, ou uma combinação destes (p. ex. uma ou mais NLS no amino-terminal e uma ou mais NLS no carboxi-terminal). Quando estão presentes mais de uma NLS, cada uma pode ser selecionada independentemente das outras, de modo a que uma única NLS possa estar presente em mais de uma

cópia e/ou em conjunto com uma ou mais outras NLSs presentes numa ou em mais cópias. Numa forma de realização preferida da invenção, a enzima CRISPR compreende, no máximo, 6 NLSs. Em algumas formas de realização, uma NLS é considerada como estando junto ao N-terminal ou C-terminal quando o aminoácido mais próximo da NLS se encontra em cerca de 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 ou mais aminoácidos ao longo da cadeia de polipéptidos a partir do N-terminal ou C-terminal. Habitualmente, uma NLS consiste em uma ou mais sequências curtas de lisinas ou argininas positivamente carregadas expostas na superfície da proteína, mas são conhecidos outros tipos de NLS. Os exemplos não limitativos de NLSs incluem uma sequência NLS obtida a partir: da NLS do T-antígeno grande de vírus SV40, com a sequência de aminoácidos PKKKRKV; da NLS da nucleoplasmina (p. ex. a NLS bipartida de nucleoplasmina com a sequência KRPAATKKAGQAKKKK); da NLS c-myc com a sequência de aminoácidos PAAKRVKLD ou RQRRNELKRSP; da NLS hRNPAl M9 com a sequência NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRN-QGGY; da sequência RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDFQILKRRNV do domínio IBB a partir de importin-alfa; das sequências VSRKRPRP e PPKKARED da proteína T de mioma; da sequência POPKKKPL de p53 humano; da sequência SALIKKKKKMAP de c-abl IV de ratinho; das sequências DRLRR e PKQKKRK do vírus da influenza NS1; da sequência RKLKKKIKKL do antígeno delta do vírus da Hepatite; da sequência REKKKFLKRR da proteína Mx1 de ratinho; da sequência KRKGDEV DGVDEVAKKKSKK da poli(ADP-ribose) polimerase humana; e da sequência RKCLQAG-MNLEARKTKK do glicocorticoide (humano) dos recetores de hormona esteroide.

Em geral, uma ou mais NLSs têm força suficiente para ativar a acumulação da enzima CRISPR numa quantidade detetável no núcleo de uma célula eucariótica. Em geral, a força de atividade de localização nuclear pode ser obtida a

partir do número de NLSS na enzima CRISPR, das NLSS específicas utilizadas ou de uma combinação destes fatores. A detecção de acumulação no núcleo pode ser efetuada através de qualquer técnica adequada. Por exemplo, um marcador detetável pode ser fundido na enzima CRISPR, de modo a que a localização numa célula possa ser visualizada, tal como em conjunto com um meio de detecção da localização do núcleo (p. ex. uma coloração específica para o núcleo, tal como DAPI). Os exemplos de marcadores detetáveis incluem proteínas fluorescentes (tais como proteínas de cor verde fluorescente, ou GFP; RFP; CFP), e etiquetas de epítipo (etiqueta HA, etiqueta FLAG, etiqueta SNAP). Os núcleos celulares também podem ser isolados das células, cujos conteúdos podem ser depois analisados através de qualquer processo adequado de detecção de proteína, tal como ensaio de atividade enzimática, *Western blot* ou imunohistoquímica. A acumulação no núcleo também pode ser determinada indiretamente, tal como através de um ensaio para o efeito da formação de complexos CRISPR (p. ex. ensaio para a mutação ou clivagem de ADN na sequência-alvo, ou ensaio para a atividade de expressão de genes alterada afetada pela formação de complexos CRISPR e/ou atividade de enzima CRISPR), em comparação com um controlo não exposto ao complexo ou à enzima CRISPR, ou exposto a uma enzima CRISPR que não tem um ou mais NLSS.

Em geral, uma sequência-guia corresponde a qualquer sequência de polinucleótidos com uma complementaridade suficiente com uma sequência-alvo de polinucleótidos para hibridizar com a sequência-alvo e direcionar a ligação específica da sequência de um complexo CRISPR para a sequência-alvo. Em algumas formas de realização, o grau de complementaridade entre uma sequência-guia e a respetiva sequência-alvo correspondente, quando alinhada de forma ideal utilizando um algoritmo de alinhamento adequado, é de cerca ou mais de cerca de 50%, 60%, 75%, 80%, 85%, 90%,

95%, 97,5%, 99% ou mais. O alinhamento ideal pode ser determinado com a utilização de qualquer algoritmo adequado para alinhar sequências, cujo exemplo não limitativo inclui o algoritmo de Smith-Waterman, o algoritmo de Needleman-Wunsch, os algoritmos baseados na Transformada de Burrows-Wheeler (p. ex. o Alinhamento 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 algumas formas de realização, uma sequência-guia tem um comprimento de cerca ou mais de cerca de 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 de nucleótidos. Em algumas formas de realização, uma sequência-guia tem um comprimento inferior a cerca de 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 ou menos nucleótidos. A capacidade de uma sequência-guia para direcionar a ligação específica da sequência de um complexo CRISPR para uma sequência-alvo pode ser avaliada através de qualquer ensaio adequado. Por exemplo, os componentes de um sistema CRISPR suficientes para formar um complexo CRISPR, incluindo a sequência-guia a testar, podem ser fornecidos a uma célula hospedeira com a sequência-alvo correspondente, tal como mediante transfeção com vetores que codificam os componentes da sequência CRISPR, seguida de uma avaliação da clivagem preferencial na sequência-alvo, tal como através de ensaio Surveyor conforme aqui descrito. De forma semelhante, a clivagem de uma sequência-alvo de polinucleótidos pode ser avaliada num tubo de ensaio fornecendo à sequência-alvo componentes de um complexo CRISPR, incluindo a sequência-guia a testar e uma sequência-guia de controlo diferente da sequência-guia de teste, e comparando a ligação ou taxa de clivagem na sequência-alvo entre as reações da sequência-guia de controlo e de teste. São possíveis outros ensaios, o que

será compreendido pelos peritos na técnica.

Uma sequência-guia pode ser selecionada para escolher como alvo qualquer sequência-alvo. Em algumas formas de realização, a sequência-alvo é uma sequência num genoma de uma célula. As sequências-alvo exemplares incluem as que são exclusivas no genoma-alvo. Por exemplo, para a Cas9 de *S. pyogenes*, uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 da forma MMMMMMMMNNNNNNNNNNNNNNXGG, em que NNNNNNNNNNNNNXGG (N corresponde a A, G, T ou C; e X pode corresponder a qualquer coisa) tem uma única ocorrência no genoma. Uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 de *S. pyogenes* da forma MMMMMMMMNNNNNNNNNNNNNNXGG, em que NNNNNNNNNNNNNXGG (N corresponde a A, G, T ou C; e X pode corresponder a qualquer coisa) tem uma única ocorrência no genoma. Para a Cas9 CRISPR1 de *S. thermophilus*, uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 da forma MMMMMMMMNNNNNNNNNNNNNNXXAGAA W, em que NNNNNNNNNNNNNXXAGAAW (N corresponde a A, G, T ou C; X pode corresponder a qualquer coisa; e W corresponde a A ou T) tem uma única ocorrência no genoma. Uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 CRISPR1 de *S. thermophilus* da forma MMMMMMMMNNNNNNNNNNNNNNMXXAGAAW, em que NNNNNNNNNNNNNXXAGAAW (N corresponde a A, G, T ou C; X pode corresponder a qualquer coisa; e W corresponde a A ou T) tem uma única ocorrência no genoma. Para a Cas9 de *S. pyogenes*, uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 da forma MMMMMMMMNNNNNNNNNNNNNNXGGXG, em que NNNNNNNNNNNNNXGGXG (N corresponde a A, G, T ou C; e X pode corresponder a qualquer coisa) tem uma única ocorrência no genoma. Uma sequência-alvo exclusiva num genoma pode incluir um local-alvo de Cas9 de *S. pyogenes* da forma MMMMMMMMNNNNNNNNNNNNNNXGGXG, em que NNNNNNNNNNNNNXGGXG (N

corresponde a A, G, T ou C; e X pode corresponder a qualquer coisa) tem uma única ocorrência no genoma. Em cada uma destas sequências, "M" pode corresponder a A, G, T ou C, e não necessita de ser considerado na identificação de uma sequência como exclusivo.

Em algumas formas de realização, uma sequência-guia é selecionada para reduzir o grau de estrutura secundária na sequência-guia. A estrutura secundária pode ser determinada através de qualquer algoritmo de dobramento de polinucleótidos adequado. Alguns programas baseiam-se no cálculo da energia livre de Gibbs mínima. Um exemplo de um algoritmo desses é mFold (Dobram), conforme descrito por Zuker e Stiegler (*Nucleic Acids Res.* 9 (1981), 133 a 148). Outro algoritmo de dobramento de exemplo é RNAfold (dobraARN) do servidor da Web online, desenvolvido no Institute for Theoretical Chemistry at the University of Vienna, utilizando o algoritmo de previsão de estrutura central (consulte, p. ex. A.R. Gruber *et al.*, 2008, *Cell* 106(1): 23 e 24; e PA Carr e GM Church, 2009, *Nature Biotechnology* 27(12): 1151 a 62).

Em geral, uma sequência tracr-mate inclui qualquer sequência que tenha complementaridade suficiente com uma sequência tracr para provocar um ou mais entre o apresentado em seguida: (1) excisão de uma sequência-guia flanqueada pelas sequências tracr-mate 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 tracr-mate hibridizada para a sequência tracr. Em geral, o grau de complementaridade refere-se ao alinhamento ideal da sequência tracr-mate e sequência tracr, ao longo do comprimento da mais curta das duas sequências. O alinhamento ideal pode ser determinado por qualquer algoritmo de alinhamento adequado, e pode ainda justificar as estruturas secundárias, tal como a autocomplementaridade na sequência tracr ou sequência

tracr-mate. Em algumas formas de realização, o grau de complementaridade entre a sequência tracr e a sequência tracr-mate ao longo do comprimento da mais curta das duas quando alinhadas de forma ideal é de cerca, ou mais de cerca, de 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99% ou mais. As ilustrações de exemplo de alinhamento ideal entre uma sequência tracr e uma sequência tracr-mate são fornecidas nas Figuras 12B e 13B. Em algumas formas de realização, a sequência tracr tem um comprimento de cerca ou mais de cerca de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50 ou mais nucleótidos. Em algumas formas de realização, a sequência tracr e a sequência tracr-mate estão contidas numa única transcrição, de modo a que a hibridização entre as duas produza uma transcrição com uma estrutura secundária, tal como um gancho (*hairpin*). As sequências de formação de estruturas preferidas para a utilização nas estruturas em gancho têm um comprimento de quatro nucleótidos e, mais preferencialmente, a sequência GAAA. Contudo, podem ser utilizadas sequências de estrutura mais longas ou mais curtas, tais como sequências alternativas. As sequências incluem preferencialmente um triplete de nucleótidos (por exemplo, AAA), e um nucleótido adicional (por exemplo C ou G). Os exemplos de sequências de formação de estruturas incluem CAAA e AAAG. Numa forma de realização, a transcrição ou a sequência de polinucleótidos transcrita tem, pelo menos, dois ou mais ganchos. Nas formas de realização preferidas, a transcrição tem dois, três, quatro ou cinco ganchos. Numa outra forma de realização, a transcrição tem, no máximo, cinco ganchos. Em algumas formas de realização, a única transcrição inclui ainda uma sequência de terminação de transcrição; preferencialmente trata-se de uma sequência poliT, por exemplo seis nucleótidos T. Uma ilustração de exemplo de uma estrutura em gancho assim é fornecida na parte inferior da Figura

13B, em que a parte da sequência 5' do "N" final e a montante da estrutura corresponde à sequência tracr-mate, e a parte da sequência 3' da estrutura corresponde à sequência tracr. Outros exemplos não limitativos de polinucleótidos únicos compreendendo uma sequência-guia, uma sequência tracr-mate e uma sequência tracr são conforme apresentados em seguida (indicados de 5' a 3'), em que "N" representa uma base de uma sequência-guia, o primeiro bloco de letras minúsculas representa a sequência tracr-mate, e o segundo bloco de letras minúsculas representa a sequência tracr, e a sequência poli-T final representa o terminador de transcrição:

(1)
NNNNNNNNNNNNNNNNNNNNNNggttttgtactctcaagatttaGAAAtaatettgcaga
agctacaaagataaggcttt catgccgaaat-

caacaccctgtcattttatggcaggggtgttttcgttatttaaTTTTT; (2)

NNNNNNNNNNNNNNNNNNNNNNggttttgtactctcaGAAAtgcagaagct-
acaaagataaggcttcatgccgaaatca
acaccctgtcattttatggcaggggtgttttcgttatttaaTTTTT; (3)

NNNNNNNNNNNNNNNNNNNNNN-
gttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca
acaccctgtcattttatggcaggggtgtTTTTT; (4)

NNNNNNNNNNNNNNNNNNNNNNggttttagagctaGAAAtagcaaggttaaataaggctag
tccgttatcaacttgaaaa agtggcaccgagtcggt-gcTTTTT; (5)

NNNNNNNNNNNNNNNNNNNNNNggttttagagctaGAAATAGcaaggttaaataaggctagt
ccgttatcaacttgaa aaagt-gTTTTT; e (6)

NNNNNNNNNNNNNNNNNNNNNNggttttagagctagAAATAGcaagftamataaggctagtc
cgttatcaTTTTT TTT. Em algumas formas de realização, as sequências (1) a (3) são utilizadas em conjunto com Cas9 de *S. thermophilus* CRISPR1. Em algumas formas de realização, as sequências (4) a (6) são utilizadas em conjunto com Cas9 de *S. pyogenes*. Em algumas formas de realização, a sequência tracr é uma transcrição separada de uma transcrição que compreende a sequência tracr-mate (tal conforme ilustrado na parte superior da Figura 13B).

Em algumas formas de realização, é igualmente

fornecido um modelo de recombinação. Um modelo de recombinação pode ser um componente de outro vetor conforme aqui descrito, 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 modelo na recombinação homóloga, tal como numa, ou junto a uma, sequência-alvo entalhada (*nicked*) ou clivada por uma enzima CRISPR como uma parte de um complexo CRISPR. Um polinucleótido modelo pode ter qualquer comprimento adequado, tal como um comprimento de cerca, ou mais de cerca, de 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000 ou mais nucleótidos. Em algumas formas de realização, o polinucleótido modelo é complementar a uma porção de um polinucleótido que compreende a sequência-alvo. Quando alinhado de forma ideal, um polinucleótido modelo poderá ser sobreposto por um ou mais nucleótidos de uma sequência-alvo (p. ex. cerca ou mais de cerca de 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 de modelo e um polinucleótido compreendendo uma sequência-alvo se encontram idealmente alinhados, o nucleótido mais próximo do polinucleótido modelo encontra-se a cerca de 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000 ou mais nucleótidos da sequência-alvo.

Em algumas formas de realização, a enzima CRISPR é uma parte de uma proteína de fusão que compreende um ou mais domínios de proteína heteróloga (p. ex. cerca ou mais de cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais domínios além da enzima CRISPR). Uma proteína de fusão de enzima CRISPR pode compreender qualquer sequência de proteínas adicional e, opcionalmente, uma sequência de ligante entre quaisquer dois domínios. Os exemplos de domínios de proteína que podem ser fundidos numa enzima CRISPR incluem, sem limitação, etiquetas de epítopo, sequências de genes

repórteres e domínios de proteína com uma ou mais das seguintes atividades: atividade de metilase, atividade de demetilase, atividade de ativação de transcrição, atividade de repressão de transcrição, atividade de fator de libertação de transcrição, atividade de modificação de histona, atividade de clivagem de ARN e atividade de ligação de ácido nucleico. Os exemplos não limitativos de etiquetas de epítipo incluem etiquetas de histidina (His), etiquetas V5, etiquetas FLAG, etiquetas de influenza hemaglutinina (HA), etiquetas Myc, etiquetas VSV-G e etiquetas de tioredoxina (Trx). Os exemplos de genes repórteres incluem, mas não se limitam a, glutathione-S-transferase (GST), peroxidase de rábano-silvestre (HRP), cloranfenicol acetiltransferase (CAT), beta-galactosidase, beta-glucuronidase, luciferase, proteína verde fluorescente (GFP), HcRed, DsRed, proteína ciano fluorescente (CFP), proteína amarela fluorescente (YFP) e proteínas autofluorescentes incluindo proteína azul fluorescente (BFP). Uma enzima CRISPR pode ser fundida numa sequência de genes que codifica uma proteína ou um fragmento de uma proteína que liga moléculas de ADN ou liga outras moléculas celulares, incluindo, mas não se limitando a, proteína de ligação à maltose (MBP), etiqueta S, fusões de domínio e ligação ao ADN Lex A (DBD), fusões de domínio de ligação ao ADN GAL4, e fusões de proteína BP16 do vírus herpes Simplex (HSV). Os domínios adicionais que podem fazer parte de uma proteína de fusão compreendendo uma enzima CRISPR são descritos em US20110059502. Em algumas formas de realização, uma enzima CRISPR com etiqueta é utilizada para identificar a localização de uma sequência-alvo.

Em alguns aspetos, a divulgação fornece métodos que compreendem a distribuição de um ou mais polinucleótidos, tais como um ou mais vetores conforme aqui descrito, uma ou mais transcrições dos mesmos, e/ou uma ou mais proteínas transcritas dos mesmos, a uma célula hospedeira. Em alguns

aspectos, a divulgação fornece ainda células produzidas por esses métodos, e organismos (tais como animais, plantas ou fungos) que compreendem ou são produzidos a partir dessas células. Em algumas formas de realização, é fornecida a uma célula uma enzima CRISPR em conjunto com (e opcionalmente complexada com) uma sequência-guia. Os métodos de transferência de genes com bases virais e não virais convencionais podem ser utilizados para introduzir ácidos nucleicos nas células de mamífero ou tecidos-alvo. Esses métodos podem ser utilizados para administrar ácidos nucleicos que codificam componentes de um sistema CRISPR para células na cultura, ou num organismo hospedeiro. Os sistemas de distribuição de vetores não virais incluem plasmídeos de ADN, ARN (p. ex. uma transcrição de um vetor aqui descrito), ácido nucleico nu, e ácido nucleico complexado com um veículo de distribuição, tal como um lipossoma. Os sistemas de distribuição de vetores virais incluem vírus de ADN e ARN, que têm genomas episomais ou integrados após o fornecimento à célula. Para obter uma revisão dos procedimentos de terapia genética, consulte Anderson, *Science* 256:808 a 813 (1992); Nabel & Felgner, *TIBTECH* 11:211 a 217 (1993); Mitani & Caskey, *TIBTECH* 11:162 a 166 (1993); Dillon, *TIBTECH* 11:167 a 175 (1993); Miller, *Nature* 357:455 a 460 (1992); Van Brunt, *Biotechnology* 6(10):1149 a 1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35 a 36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31 a 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 a 26 (1994).

Os métodos de distribuição não viral de ácidos nucleicos incluem lipofecção, nucleofecção, microinjeção, biolística, virossomas, lipossomas, imunolipossomas, policação ou conjugados de lípido:ácido nucleico, ADN nu, viriões artificiais, e absorção de ADN melhorada pelo

agente. A lipofecção é descrita, p. ex., nas Patentes U.S. n.º 5.449.386, 4.946.787; e 4.897.355, e os reagentes de lipofecção são comercializados (p. ex., Transfectam™ e Lipofectin™). Os lípidos catiónicos e neutros que são adequados para uma lipofecção de polinucleótidos eficaz de reconhecimento de recetores incluem os de Felgner, WO 91/17424; WO 91/16024. O fornecimento pode ser a células (p. ex. administração *in vitro* ou *ex vivo*) ou tecidos-alvo (p. ex. administração *in vivo*).

A preparação de complexos de lípido:ácido nucleico, incluindo lipossomas escolhidos como alvo, tais como complexos de imunolípido, é bem conhecida de um perito na técnica (consulte, p. ex., Crystal, *Science* 270:404 a 410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291 a 297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382 a 389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647 a 654 (1994); Gao *et al.*, *Gene Therapy* 2:710 a 722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817 a 4820 (1992); Patentes U.S. n.ºs 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 à base de ARN ou ADN viral para o fornecimento de ácidos nucleicos tira partido de processos altamente evoluídos para a escolha de um vírus como alvo para células específicas no corpo e o tráfego da carga viral até ao núcleo. Os vetores virais podem ser administrados diretamente nos pacientes (*in vivo*) ou podem ser utilizados para tratar células *in vitro*, e as células modificadas podem opcionalmente ser administradas nos pacientes (*ex vivo*). Os sistemas com bases virais convencionais podem incluir vetores retrovirais, de lentivírus, adenovirais, adeno-associados e de vírus herpes simplex para a transferência de genes. A integração no genoma hospedeiro é possível com os métodos de transferência de genes de retrovírus, lentivírus e vírus adeno-associado, resultando muitas vezes na expressão de

longo prazo do transgene inserido. Adicionalmente, foram observadas eficiências de alta transdução em muitos tipos de célula e tecidos-alvo diferentes.

O tropismo de um retrovírus pode ser alterado incorporando proteínas envelope estranhas, expandindo a população-alvo potencial de células-alvo. Os vetores lentivirais são vetores retrovirais que são capazes de transformar ou infectar células sem divisão e produzir habitualmente titulações virais elevadas. Por conseguinte, a seleção de um sistema de transferência de genes retrovirais dependeria do tecido-alvo. Os vetores retrovirais são constituídos por repetições terminais longas que atuam em cis com uma capacidade de empacotamento de, no máximo, 6 a 10 kb de sequência estranha. As LTRs que atuam em cis mínimas são suficientes para a replicação e o empacotamento dos vetores, que são depois utilizados para integrar o gene terapêutico na célula-alvo para fornecer a expressão de transgenes permanente. Os vetores retrovirais muito utilizados incluem os que se baseiam no vírus da leucemia murina (MuLV), vírus da leucemia do macaco gibão (GaLV), vírus da Imunodeficiência Símia (VIS), vírus da imunodeficiência humana (VIH), e combinações dos mesmos (consulte, p. ex., Buchscher *et al.*, *J. Virol.* 66:2731 a 2739 (1992); Johann *et al.*, *J. Virol.* 66:1635 a 1640 (1992); Sommnerfelt *et al.*, *Virol.* 176:58 a 59 (1990); Wilson *et al.*, *J. Virol.* 63:2374 a 2378 (1989); Miller *et al.*, *J. Virol.* 65:2220 a 2224 (1991); PCT/US94/05700).

Nas aplicações em que a expressão transitória é preferida, podem ser utilizados sistemas baseados em adenovírus. Os vetores baseados em adenovírus são capazes de uma eficiência de transdução muito elevada em muitos tipos de célula e não necessitam de divisão celular. Com esses vetores, foram obtidos níveis de expressão e titulações elevados. Este vetor pode ser produzido em grandes quantidades num sistema relativamente simples. Os

vetores de vírus adeno-associado ("AAV") também podem ser utilizados para transformar células com ácidos nucleicos-alvo, p. ex., na produção *in vitro* de ácidos nucleicos e péptidos, e relativamente a procedimentos de terapia genética *in vivo* e *ex vivo* (consulte, p. ex., West et al., *Virology* 160:38 a 47 (1987); Patente U.S. n.º 4.797.368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793 a 801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). A construção de vetores AAV recombinantes é descrita numa quantidade de publicações, incluindo a Patente U.S. n.º 5.173.414; Tratschin et al., *Mol. Cell. Biol.* 5:3251 a 3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072 a 2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466 a 6470 (1984); e Samulski et al., *J. Virol.* 63:03822 a 3828 (1989).

As células de empacotamento são habitualmente utilizadas para formar partículas de vírus que sejam capazes de injetar uma célula hospedeira. Essas células incluem células 293, que empacotam adenovírus, e células ψ 2 ou células PA317, que empacotam retrovírus. Os vetores virais utilizados na terapia genética são habitualmente gerados através da produção de uma linha celular que empacota um vetor de ácido nucleico numa partícula viral. Geralmente, os vetores contêm as sequências virais mínimas necessárias para o empacotamento e subsequente integração num hospedeiro, em que outras sequências virais são substituídas por uma cassete de expressão para que os polinucleótidos sejam expressados. As funções virais em falta são normalmente fornecidas em trans pela linha celular de empacotamento. Por exemplo, habitualmente, os vetores AAV utilizados na terapia genética apenas têm sequências ITR do genoma AAV que sejam necessárias para o empacotamento e a integração no genoma hospedeiro. O ADN viral é empacotado numa linha celular, que contém um plasmídeo auxiliar que codifica os outros genes AAV, nomeadamente rep e cap, mas não tem sequências ITR. A linha

celular também pode ser infectada com adenovírus como um auxiliar. O vírus auxiliar provoca a replicação do vetor AAV e a expressão de genes AAV a partir do plasmídeo auxiliar. O plasmídeo auxiliar não é empacotado em quantidades significativas devido a uma falta de sequências ITR. A contaminação com adenovírus pode ser reduzida através, p. ex., de tratamento térmico ao qual o adenovírus é mais sensível do que o AAV. Os métodos adicionais para a distribuição de ácidos nucleicos a células são conhecidos dos peritos na técnica. Consulte, por exemplo, US20030087817.

Em algumas formas de realização, uma célula hospedeira é transfetada de forma transitória ou não transitória com um ou mais vetores aqui descritos. Em algumas formas de realização, uma célula é transfetada à medida que ocorre naturalmente num sujeito. Em algumas formas de realização, uma célula que é transfetada é retirada de um sujeito. Em algumas formas de realização, a célula é obtida a partir de células retiradas de um sujeito, tal como uma linha celular. Uma vasta variedade de linhas celulares para a cultura de tecidos é conhecida na técnica. Os exemplos de linhas celulares incluem, mas não se limitam a, 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, células epiteliais de rim de macaco BS-C-1, fibroblasto de embrião de ratinho BALB/3T3, 3T3 Swiss, 3T3-L1, fibroblastos fetais humanos 132-d5; fibroblastos de ratinho 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, CHORE, 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, Hepalclc7, 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/PNT 2, 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 uma variedade de origens conhecidas dos peritos na técnica (consulte, p. ex., a Coleção Americana de Culturas Celulares (ATTC - American Type Culture Collection) (Manassus, Va.)). Em algumas formas de realização, é utilizada uma célula transfetada com um ou mais vetores aqui descritos para estabelecer uma nova linha celular compreendendo uma ou mais sequências derivadas de vetores. Em algumas formas de realização, uma célula transfetada transitoriamente com os componentes de um sistema CRISPR conforme aqui descrito (tal como por transfeção transitória de um ou mais vetores, ou transfeção com ARN), e modificada através da atividade de um complexo CRISPR, é utilizada para estabelecer uma nova linha celular compreendendo células que contêm a modificação, mas não têm qualquer outra sequência exógena. Em algumas formas de realização, as células transfetadas de forma transitória ou não transitória com um ou mais vetores aqui descritos, ou linhas celulares derivadas dessas células, são utilizadas na avaliação de um ou mais compostos de teste.

Em algumas formas de realização, são utilizados um ou mais vetores aqui descritos para produzir uma planta

transgénica ou animal transgénico não humano. Em algumas formas de realização, o animal transgénico é um mamífero, tal como um ratinho, um rato ou um coelho. Em determinadas formas de realização, o organismo ou sujeito é uma planta. Em determinadas formas de realização, o organismo ou o sujeito ou a planta são algas. Os métodos de produção de plantas e animais transgénicos são conhecidos na técnica, e começam geralmente com um método de transfeção de células, tal como aqui descrito. Os animais transgénicos são igualmente fornecidos, tal como as plantas transgénicas, especialmente colheitas e algas. A planta ou o animal transgénico pode ser útil em aplicações fora do fornecimento de um modelo de doença. Estas podem incluir alimentos ou produção de alimentos através da expressão de, por exemplo, níveis mais elevados de proteínas, hidratos de carbono, nutrientes ou vitaminas do que os normalmente observados no tipo selvagem. Relativamente a isto, são preferíveis as plantas transgénicas, especialmente legumes e tubérculos, e os animais, especialmente mamíferos, tais como gado (vacas, ovelhas, cabras e porcos), mas também aves domésticas e insetos comestíveis.

As algas transgénicas ou outras plantas, tais como colza, podem ser particularmente úteis na produção de óleos vegetais ou biocombustíveis, tais como álcoois (especialmente metanol e etanol), por exemplo. Estes podem ser produzidos para expressar ou sobre-expressar níveis elevados de óleo ou álcoois para a utilização nas indústrias de óleo ou biocombustível.

Num aspeto, a divulgação fornece métodos de modificação de um polinucleótido-alvo numa célula eucariótica. Em algumas formas de realização, o método compreende a permissão a um complexo CRISPR para o mesmo se ligar 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 complexada com uma sequência-guia hibridizada para uma sequência-alvo no referido polinucleótido-alvo, em que a referida sequência-guia se encontra ligada a uma sequência tracr-mate que, por sua vez, é hibridizada para uma sequência tracr.

Num aspeto, a divulgação fornece um método de modificação de expressão de um polinucleótido numa célula eucariótica. Em algumas formas de realização, o método compreende a permissão a um complexo CRISPR para o mesmo se ligar ao polinucleótido, de modo a que a referida ligação resulte numa maior ou menor expressão do referido polinucleótido, em que o complexo CRISPR compreende uma enzima CRISPR complexada com uma sequência-guia hibridizada para uma sequência-alvo no referido polinucleótido, em que a referida sequência-guia se encontra ligada a uma sequência tracr-mate que, por sua vez, é hibridizada para uma sequência tracr.

Com os recentes avanços na genómica das colheitas, a capacidade de utilizar sistemas CRISPR-Cas para efetuar uma manipulação e uma edição de genes eficazes e economicamente viáveis irá permitir a comparação e seleção rápidas de manipulações genéticas únicas e multiplexadas para transformar esses genomas para uma melhor produção e características melhoradas. Relativamente a isto, é feita referência a publicações e patentes US: patente US n.º 6.603.061 - "Agrobacterium-Mediated Plant Transformation Method" (Método de Transformação de Plantas Mediado por Agrobactérias); patente US n.º 7.868.149 - "Plant Genome Sequences and Uses Thereof" (Sequências de Genomas de Plantas e Respetivas Utilizações) e US 2009/0100536 - "Transgenic Plants with Enhanced Agronomic Traits" (Plantas Transgénicas com Características Agronómicas Melhoradas), Morrell et al. "Crop genomics: advances and applications" *Nat Rev Genet.* 29 de dezembro de 2011; 13(2):85 a 96. Numa forma de realização vantajosa da invenção, o sistema

CRISPR/Cas9 é utilizado para produzir microalgas (Exemplo 14). Em conformidade, neste caso, a referência a células de animais também pode aplicar-se, *mutatis mutandis*, às células de plantas, salvo evidência em contrário.

Num aspeto, a divulgação fornece métodos de 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 de uma planta (incluindo microalgas) ou animal humano ou não humano, e a modificação da célula ou das células. A cultura pode ocorrer em qualquer estágio *ex vivo*. A célula ou as células até podem ser reintroduzidas na planta (incluindo microalga) ou no animal não humano.

Num aspeto, a divulgação fornece *kits* que contêm qualquer um ou mais dos elementos divulgados nos métodos e nas composições acima. Em algumas formas de realização, o *kit* compreende um sistema de vetores e instruções de utilização do *kit*. Em algumas formas de realização, o sistema de vetores compreende (a) um primeiro elemento regulador ligado de modo funcional a uma sequência *tracr-mate* e a um ou mais locais de inserção para inserir uma sequência-guia a montante da sequência *tracr-mate*, em que, quando expressada, a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para uma sequência-alvo numa célula eucariótica, em que o complexo CRISPR compreende uma enzima CRISPR complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência *tracr-mate* que é hibridizada para a sequência *tracr*; e/ou (b) um segundo elemento regulador ligado de modo funcional a uma sequência de codificação de enzimas que codifica 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 num ou em mais idiomas, por exemplo em mais de um idioma.

Em algumas formas de realização, um *kit* compreende um ou mais reagentes para a utilização num processo que utiliza um ou mais dos elementos aqui descritos. Os reagentes podem ser fornecidos em qualquer recipiente adequado. Por exemplo, um *kit* pode fornecer um ou mais tampões de armazenamento ou reação. Os reagentes podem ser fornecidos numa forma que seja utilizável num determinado ensaio, ou numa forma que necessite da adição de um ou mais outros componentes antes da utilização (p. ex. na forma concentrada ou liofilizada). Um tampão pode ser qualquer tampão, incluindo, mas não se limitando a, um tampão de carbonato de sódio, um tampão de bicarbonato de sódio, um tampão de borato, um tampão Tris, um tampão MOPS, um tampão HEPES e combinações dos mesmos. Em algumas formas de realização, o tampão é alcalina. Em algumas formas de realização, o tampão tem um pH de cerca de 7 a cerca de 10. Em algumas formas de realização, o *kit* compreende um ou mais oligonucleótidos correspondentes a uma sequência-guia para a inserção num vetor, de modo a ligar de modo funcional a sequência-guia e um elemento regulador. Em algumas formas de realização, o *kit* compreende um polinucleótido modelo de recombinação homóloga.

Num aspeto, a divulgação fornece métodos para pedir um ou mais elementos de um sistema CRISPR. O complexo CRISPR da invenção fornece um meio eficaz para modificar um polinucleótido-alvo. O complexo CRISPR da invenção tem uma vasta variedade de utilidades incluindo a modificação (p. ex., deleção, inserção, translocação, inativação, ativação) de um polinucleótido-alvo numa multiplicidade de tipos de célula. Como tal, o complexo CRISPR da invenção tem um largo espectro de aplicações, p. ex., na terapia genética, rastreio de fármacos, diagnóstico de doenças e prognóstico.

Um complexo CRISPR exemplar compreende uma enzima CRISPR complexada com uma sequência-guia hibridizada para uma sequência-alvo no polinucleótido-alvo. A sequência-guia é ligada a uma sequência tracr-mate que, por sua vez, é hibridizada para uma sequência tracr.

O polinucleótido-alvo de um complexo CRISPR pode ser qualquer polinucleótido endógeno ou exógeno em relação à 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 gene (p. ex., uma proteína) ou uma sequência de não-codificação (p. ex., um polinucleótido regulador ou uma ADN lixo). Sem ser pretendida qualquer limitação a uma teoria, parte-se do princípio de que a sequência-alvo deve ser associada a um PAM (*protospacer*-motivo adjacente); ou seja, uma sequência curta reconhecida pelo complexo CRISPR. Os requisitos precisos de sequência e comprimento para o PAM diferem consoante a enzima CRISPR utilizada, mas os PAMs têm habitualmente 2 a 5 sequências de pares de bases adjacentes ao *protospacer* (ou seja, a sequência-alvo). Os exemplos de sequências PAM são fornecidos na secção de exemplos abaixo, e o perito na técnica poderá identificar mais sequências PAM para a utilização com uma determinada enzima CRISPR.

O polinucleótido-alvo de um complexo CRISPR pode incluir um número de polinucleóticos e genes associados à doença, bem como polinucleótidos e genes associados à via bioquímica de sinalização.

Os exemplos de polinucleótidos-alvo incluem uma sequência associada a uma via bioquímica de sinalização, p. ex., um polinucleótido ou gene associado à via bioquímica de sinalização. Os exemplos de polinucleótidos-alvo incluem um polinucleótido ou gene associado à doença. Um polinucleótido ou gene "associado à doença" refere-se a qualquer polinucleótido ou gene que produza produtos de

transcrição ou tradução num nível anormal ou numa forma anormal nas células derivadas de tecidos afetados por uma doença em comparação com tecidos ou células de um controlo sem doença. Pode ser um gene que é expressado num nível anormalmente elevado; pode ser um gene que é expressado num 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 uma mutação ou mutações que têm genes ou uma variação genética que é diretamente responsável ou se encontra em desequilíbrio de ligação com um gene ou genes responsáveis pela etiologia de uma doença. Os produtos transcritos ou traduzidos podem ser conhecidos ou desconhecidos, e podem encontrar-se num nível normal ou anormal.

Os exemplos de polinucleótidos e genes associados à doença estão disponíveis a partir de McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) e National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), disponível na World Wide Web.

Os exemplos de polinucleótidos e genes associados à doença são indicados nas Tabelas A e B. As informações específicas da doença estão disponíveis a partir de McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) e National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), disponíveis na World Wide Web. Os exemplos de polinucleótidos e genes associados à via bioquímica de sinalização são indicados na Tabela C.

As mutações nestes genes e vias podem resultar na produção de proteínas impróprias ou proteínas em quantidades impróprias que afetam a função. Esses genes, proteínas e vias podem ser o polinucleótido-alvo de um complexo CRISPR.

Tabela A

DOENÇA/DISTÚRBIOS	GENE(S)
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 de Recetores FGF
	(5 membros: 1, 2, 3, 4, 5); CDKN2a; APC;
	RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2;
	AR (Recetor de 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	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsinaD;
relacionada com a Idade	Vldlr; Ccr2
Esquizofrenia	Neuregulina 1 (Nrg1); Erb4 (recetor para Neuregulina);
	Complexinal (Cplx1); Triptofano hidroxilase Tph1; Tph2
	Triptofano hidroxilase 2; Neurexina 1; GSK3; GSK3a;
	GSK3b
Distúrbios	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Distúrbios de	HTT (Diagnóstico de Huntington); SBMA/SMAX1/AR (Diagnóstico de Kennedy);
Repetição de Trinucleótidos	FXN/X25 (Ataxia de Friedreich); ATX3 (Diagnóstico de Machado-
	Joseph); ATXN1 e ATXN2 (ataxias

	espinocerebelares); DMPK (distrofia miotônica); Atrofina-1 e Atn1
	(Diagnóstico de DRPLA); CBP (Creb-BP - instabilidade global); VLDLR
	(Alzheimer); Atxn7; Atxn10
Síndrome do X Frágil	FMR2; FXR1; FXR2; mGLUR5
Distúrbios Relacionados com Secretase	APH-1 (alfa e beta); Presenilina (Psen1); nicastrina (Ncstn); PEN-2
Outros	Nosl; Parp1; Nat1; Nat2
Distúrbios relacionados com prião	Prp
ELA	SOD1; ELA2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-c)
Toxicodependência	Prkce (álcool); Drd2; Drd4; ABAT (álcool); GRIA2; Grm5; Grin1; Htrlb; 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; Ubal; 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 for IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3cl1

Doença de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkina; PINK1
---------------------	---

Tabela B

Doenças e distúrbios da coagulação sanguínea	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 hemorrágicos (TBXA2R, P2RX1, P2X1); Fator H e fator tipo H 1 (HF1, CFH, HUS); Fator V e fator VIII (MCFD2); Deficiência de fator VII (F7); Deficiência de fator X (F10); Deficiência de fator XI (F11); Deficiência de fator XII (F12, HAF); Deficiência de fator XIII A (F13A1, F13A); Deficiência de 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, BATCH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Distúrbios de linfocitose hemofagocítica (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB), Distúrbios hemorrágicos (PI, ATT, F5); Distúrbios e deficiências de leucócitos (ITGB2, CD 18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Anemia de células falciformes (HBB); Talassemia (HBA2, HBB, HBD, LCRB, HBA1).
Doenças e distúrbios oncológicos e desregulação	Linfoma não-Hodgkin de células B (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10,

celular	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 imunes	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), infecção ou suscetibilidade ao VIH (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCR5 (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 para IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Imunodeficiências combinadas graves (SGIDs) (JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).

Doenças e distúrbios metabólicos, do fígado, do rim e das proteínas	Neuropatia amiloide (TTR, PALB); Amiloidose (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrose (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Fibrose cística (CFTR, ABCC7, CF, MRP7); Doenças de armazenamento de glicogénio (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3), Insuficiência hepática, começo prematuro e distúrbio neurológico (SCOD1, SCO1) 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 renal cística medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Fenilcetonúria (PAH, PKU1, QDPR, DHPR, PTS); Doença renal e hepática policística (FCYT, PKHD 1, ARPKD, PKD 1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Doenças e distúrbios musculares/do esqueleto	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 facioescapuloumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDCID, 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, PLED1, PLTN, EBS1); Osteopetrose (LRP5,

	BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ELA8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SHARD1).
Doenças e distúrbios neurológicos e neuronais	ELA (SOD1, ELA2, 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, PILAU, 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 tipo 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 (Neuregulinal (Nrg1), Erb4 (receptor para Neuregulina), Complexinal (Cplx1), Triptofano Hidroxilase Tph1, Tph2, Triptofano hidroxilase 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drdla), 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 Repetição de Trinucleótidos (HTT (Diagnóstico de Huntington), SBMA/SMAX1/AR (Diagnóstico de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Diagnóstico de Machado-Joseph), ATXN1 e ATXN2 (ataxias espinocerebelares), DMPK (distrofia miotónica), Atrofin-1 e Atn1 (DRPLA Dx), 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, catepsinaD, Vldlr, Ccr2); Catarata (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); Distrofia e opacidade da córnea (APOA1, TGFBI, CSD2, CDGG1, CSD, BIG3, 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 congénita de Leber (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH 12, LCA3); Distrofia

	Macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).
--	---

Tabela C

FUNÇÃO CELULAR	GENES
Sinalização de PIK/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; .1TGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;

	HSP90AA1; RPS6KB1
Sinalização de ERK/MAPK	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	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; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; POM1; PIK3C2A; ITGB7; YWHAZ;
	PPF1CC; 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
Sinalização de Recetores	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
de Glucocorticoides	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; 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; TGFB1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
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

Sinalização de Recetores de Efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; GDK8; 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
Sinalização de	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
Citosqueleto de Actina	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CD8;
	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
Sinalização de	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
Doença de Huntington	(MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; JGF1R; 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
Sinalização de Apoptose	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPONS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CLASP8; 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
Sinalização de Recetores de 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;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Sinalização de	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
Extravasamento de Leucócitos	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;

	MAPK4; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Sinalização de Integrinas	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
Sinalização de	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
Respostas de Fase Aguda	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
Sinalização de PTEN	ITGAM; ITGA5; RAC1; PATEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; TTGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
Sinalização de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	FMAIP1; 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
Sinalização de	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;

Recetores de Hidrocarboneto de Arilo	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
Sinalização de	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
Metabolismo Xenobiótico	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRTP1; 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
Sinalização de SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;

	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	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
Sinalização de PPAR/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABC1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	AD1POQ
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; AKT 1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
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
Sinalização de	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
Wnt e Beta catenina	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; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Sinalização de Recetores de 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
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

Sinalização de IGF-1	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; GTGF; RPS6KB1
Resposta ao Stress	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
Oxidativo mediado por NRF2	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 das	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
Células Estreladas Hepáticas	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Sinalização de 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
Sinalização de Fc Épsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; FTPN11;
	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
Sinalização de Recetores	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
Acoplados à Proteína G	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 de	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;

Fosfato de Inositol	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
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
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
Sinalização de 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	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Ponto de Verificação G1/S	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
Sinalização de Recetores de Células	RAC1; ELK1; MAPK1; IKKBK; 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
Sinalização de Recetores de Morte	CRADD; HSPB1; BID; BIRC4; TBK1; IKKBK; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
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
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
Sinalização de Esclerose	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
Lateral Amiotrófica	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
Sinalização de 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; GRK6; MAPK1;
Nicotinato e	

Nicotinamida	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
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; JUN; CCL2; PRKCA
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	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
Sináptica de Longa Duração	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Sinalização de	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
Recetores de Estrogénio	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;

	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Via de	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
Ubiquitinação de Proteínas	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
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
Sinalização de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Sinalização de Recetores do tipo Toll	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;

	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
Sinalização de MAPK p38	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRF; STAT1
Sinalização de 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; MTPP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Potenciação	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
Sináptica de Longa Duração	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Sinalização de Cálcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;

	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
Sinalização de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Sinalização de Hipoxia no	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
Sistema Cardiovascular	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
Inibição Mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABC1;
da Função RXR	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 Amiloide	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
Sinalização de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1

Ciclo Celular: Regulação do	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
Ponto de Verificação G2/M	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
de Danos no ADN	PRKDC; ATM; SFN; CDKN2A
Sinalização de Óxido de Azoto no	KDR; FLT1; PGF; AKT2; P1K3CA; PIK3CB; PIK3C3;
Sistema Cardiovascular	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Metabolismo de Purinas	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
Sinalização mediada por cAMP	RAP1A; MAPK1; GNAS; CREB1; GAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunção Mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Sinalização de Notch	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Via do Stress de	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
Retículo Endoplasmático	EIF2AK3; CASP3
Metabolismo de Pirimidinas	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Sinalização de Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3

Sinalização de	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
Beta Adrenérgica e Cardíaca	PPP2R5C
Glicólise/Gliconeogénese	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Sinalização de Interferões	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sinalização de Sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Metabolismo de	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Glicerofosfolípido	
Degradação de Fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo de Triptofano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradação de Lisinas	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Via de Reparação de	ERCC5; ERCC4; XPA; XPC; ERCC1
Excisão de Nucleótidos	
Metabolismo de	UCHL1; HK2; GCK; GPI; HK1
Amido e Sacarose	
Metabolismo de aminaçúcares	NQO1; HK2; GCK; HK1
Metabolismo de	PRDX6; GRN; YWHAZ; CYP1B1
Ácido Araquidónico	
Sinalização de Ritmo Circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de Coagulação	BDKRB1; F2R; SERPINE1; F3
Sinalização de	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Recetores de Dopamina	
Metabolismo de Glutathione	IDH2; GSTP1; ANPEP; IDH1
Metabolismo de Glicerolípido	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo de Ácido Linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo de Metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo de Piruvato	GLO1; ALDH1A1; PKM2; LDHA
Metabolismo de	ALDH1A1; :NOS3; NOS2A
Arginina e Prolina	
Sinalização de Eicosanoides	PRDX6; GRN; YWHAZ

Metabolismo de	HK2; GCK; HK1
Frutose e Manose	
Metabolismo de Galactose	HK2; GCK; HK1
Biossíntese de Estilbeno,	PRDX6; PRDX1; TYR
Cumarina e Lignina	
Via de	CALR; B2M
Apresentação de Antígenos	
Biossíntese de Esteroides	NQO1; DHCR7
Metabolismo do Butanoato	ALDH1A1; NLGN1
Ciclo de Citrato	IDH2; IDH1
Metabolismo de Ácidos Gordos	ALDH1A1; CYP1B1
Metabolismo de	PRDX6; CHKA
Glicerofosfolípido	
Metabolismo da Histidina	PRMT5; ALDH1A1
Metabolismo do Inositol	ERO1L; APEX1
Metabolismo dos Xenobióticos	GSTP1; CYP1B1
por Citocromo p450	
Metabolismo do Metano	PRDX6; PRDX1
Metabolismo da Fenilalanina	PRDX6; PRDX1
Metabolismo do Propanoato	ALDH1A1; LDHA
Metabolismo dos	PRMT5; AHCY
Selenoaminoácidos	
Metabolismo de Esfingolípidos	SPHK1; SPHK2
Metabolismo de	PRMT5
Aminofosfonato	
Metabolismo de	PRMT5
Androgénio e Estrogénio	
Metabolismo de	ALDH1A1
Ascorbato e Aldarato	
Biossíntese de Ácido Biliar	ALDH1A1
Metabolismo da Cisteína	LDHA
Biossíntese de Ácidos Gordos	FASN
Sinalização de	GNB2L1

Recetores de Glutamato	
Resposta ao Stress	PRDX1
Oxidativo mediado por NRF2	
Via da	GPI
Pentose Fosfato	
Interconversões de	UCHL1
Pentose e Glicuronato	
Metabolismo do Retinol	ALDH1A1
Metabolismo da Riboflavina	TYR
Metabolismo da Tirosina	PRMT5, TYR
Biossíntese de Ubiquinona	PRMT5
Degradação da Valina,	ALDH1A1
Leucina e Isoleucina	
Metabolismo da Glicina,	CHKA
Serina e Treonina	
Degradação de Lisinas	ALDH1A1
Dor/Paladar	TRPM5; TRPA1
Dor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Função Mitocondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurologia do 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; Rein

As formas de realização referem-se igualmente a métodos e composições relacionados com a inativação (*knocking out*) de genes, a amplificação de genes e a reparação de determinadas mutações associadas a distúrbios neurológicos e instabilidade de repetição de ADN (Robert D. Wells, Tetsuo Ashizawa, "Genetic Instabilities and Neurological Diseases", Segunda Edição, *Academic Press*, 13 de outubro de 2011 - Medical). Aspectos específicos de sequências de repetição em *tandem* foram considerados 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 e outubro de 2010;7(5):551 a 8). O sistema CRISPR-Cas pode ser aproveitado para corrigir estes defeitos de instabilidade genómica.

Um outro aspeto da divulgação refere-se à utilização do sistema CRISPR-Cas para corrigir defeitos nos genes EMP2A e EMP2B que foram identificados como estando associados à doença de Lafora. A doença de Lafora é uma condição autossómica recessiva que é caracterizada pela epilepsia mioclónica progressiva que pode começar como ataques epiléticos na adolescência. Alguns casos da doença podem ser causados por mutações nos genes ainda por identificar. A doença causa ataques, espasmos musculares, dificuldade em andar, demência e, eventualmente, morte. Atualmente, não existe nenhuma terapia que tenha provado ser eficaz contra a progressão da doença. Outras anormalidades genéticas associadas à epilepsia também podem ser escolhidas como alvo pelo sistema CRISPR-Cas e a genética subjacente é ainda descrita em "Genetics of Epilepsy and Genetic Epilepsies", editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

Noutro aspeto ainda, o sistema CRISPR-Cas pode ser utilizado para corrigir defeitos oculares que surgem de

diversas mutações genéticas mais descritas em "Genetic Diseases of the Eye", Segunda Edição, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Diversos outros aspetos referem-se à correção de defeitos associados a uma vasta gama de doenças genéticas que são mais descritas no *site* da Internet dos National Institutes of Health na subsecção de tópico "Genetic Disorders" (Distúrbios Genéticos). As doenças do cérebro genéticas podem incluir, mas não se limitam 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 outros Distúrbios de Repetição de Tripletos, Doença de Leigh, Síndrome de Lesch-Nyhan, Doença de Menkes, Miopatias Mitocondriais e colpocefalia NINDS. Estas doenças são ainda descritas no *site* da Internet dos National Institutes of Health na subsecção "Genetic Brain Disorders" (Distúrbios Genéticos do Cérebro).

Em algumas formas de realização, a condição pode ser neoplasia. Em algumas formas de realização, em que a condição é neoplasia, os genes a serem escolhidos como alvo são quaisquer uns dos indicados 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 um Distúrbio Esquizofrénico. Em algumas formas de realização, a condição pode ser um Distúrbio de Repetição de Trinucleótidos. Em algumas formas de realização, a condição pode ser Síndrome do X Frágil. Em algumas formas de realização, a condição pode ser um Distúrbio Relacionado com Secretase. Em algumas formas de realização, a condição pode ser um Distúrbio relacionado com prião. Em algumas formas de realização, a condição pode ser ELA. Em algumas formas de realização, a condição pode ser uma

toxicodependência. Em algumas formas de realização, a condição pode ser Autismo. Em algumas formas de realização, a condição pode ser Doença de Alzheimer. Em algumas formas de realização, a condição pode ser inflamação. Em algumas formas de realização, a condição pode ser Doença de Parkinson.

Os exemplos de proteínas associados à doença de Parkinson incluem, mas não se limitam a, α -sinucleína, DJ-1, LRRK2, PINK1, Parkina, UCHL1, Sinfilina-1 e NURR1.

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

Os exemplos de proteínas relacionadas com inflamação podem incluir a proteína quimiotática de monócitos-1 (MCP1) codificada pelo gene Ccr2, o tipo de recetor de quimiocinas C-C 5 (CCR5) codificado pelo gene Ccr5, o recetor de IgG IIB (FCGR2b, igualmente designado por CD32) codificado pelo gene Fcgr2b, ou a proteína Fc épsilon Rlg (FCER1g) codificada pelo gene Fcerlg, por exemplo.

Os exemplos de proteínas associadas a doenças cardiovasculares podem incluir IL1B (interleucina 1, beta), XDH (xantina desidrogenase), TP53 (proteína de tumor p53), PTGIS (prostaglandina I2 (prostaciclina) sintase), 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.

Os exemplos de proteínas associadas à doença de Alzheimer podem incluir a proteína de recetor de lipoproteínas de densidade muito baixa (VLDLR) codificada pelo gene VLDLR, a enzima de ativação de modificador tipo ubiquitina 1 (UBA1) codificada pelo gene UBA1, ou a proteína de subunidade catalítica E1 de enzima de ativação de NEDD8 (UBE1C) codificada pelo gene UBA3, por exemplo.

Os exemplos de proteínas associadas ao Distúrbio do Espectro do Autismo podem incluir a proteína 1 associada ao recetor de benzodiazepinas (periférico) (BZRAP1) codificada

pelo gene BZRAP1, a proteína de membro 2 de família AF4/FMR2 (AFF2) codificada pelo gene AFF2 (igualmente designada por MFR2), a proteína de homólogo 2 autossômico de retardação mental do X frágil (FXR1) codificada pelo gene FXR1, ou a proteína de homólogo 2 autossômico de retardação mental do X frágil (FXR2) codificada pelo gene FXR2, por exemplo.

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

Os exemplos de proteínas associadas à Esquizofrenia podem incluir NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B e combinações dos mesmos.

Os exemplos de proteínas envolvidas na supressão tumoral podem incluir ATM (ataxia-telangiectasia mutante), ATR (ataxia-telangiectasia e relacionada com Rad3), EGFR (receptor de 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), Notch 1, Notch2, Notch 3 ou Notch 4, por exemplo.

Os exemplos de proteínas associadas a um distúrbio de secretase podem incluir PSENEN (homólogo 2 de intensificador de presenilina (*C. elegans*)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (proteína precursora amiloide beta (A4)), APH1B (homólogo B 1 de faringe anterior defeituosa (*C. elegans*)), PSEN2 (presenilina 2 (Doença de Alzheimer 4)), ou BACE1 (enzima 1 de clivagem APP de beta-local), por exemplo.

Os exemplos de proteínas associadas à Esclerose

Lateral Amiotrófica podem incluir SOD1 (superóxido dismutase 1), ELA2 (esclerose lateral amiotrófica 2), FUS (fundido no sarcoma), TARDBP (proteína de ligação ao ADN TAR), VAGFA (fator de crescimento endotelial vascular A), VAGFB (fator de crescimento endotelial vascular B) e VAGFC (fator de crescimento endotelial vascular C), e qualquer combinação dos mesmos.

Os exemplos de proteínas associadas a doenças relacionadas com prião podem incluir SOD1 (superóxido dismutase 1), ELA2 (esclerose lateral amiotrófica 2), FUS (fundido no sarcoma), TARDBP (proteína de ligação ao ADN TAR), VAGFA (fator de crescimento endotelial vascular A), VAGFB (fator de crescimento endotelial vascular B) e VAGFC (fator de crescimento endotelial vascular C), e qualquer combinação dos mesmos.

Os exemplos de proteínas relacionadas com condições neurodegenerativas nos distúrbios de prião podem incluir A2M (Alfa-2-Macroglobulina), AATF (Fator de transcrição antagonista de apoptose), ACPP (Próstata de fosfatase ácida), ACTA2 (aorta do músculo liso de actina alfa 2), ADAM22 (domínio de metalopeptidase ADAM), ADORA3 (recetor A3 de adenosinas), ou ADRA1D (recetor adrenérgico Alfa-1D para adrenorecetor Alfa-1D), por exemplo.

Os exemplos de proteínas associadas a 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 ATP, subfamília A (ABC1), membro 3]; por exemplo.

Os exemplos de proteínas associadas a Distúrbios de Repetição de Trinucleótidos incluem AR (recetor de androgénios), FMR1 (retardação mental 1 do X frágil), HTT (huntingtina), ou DMPK (distrofia miotónica-proteína quinase), FXN (frataxina), ATXN2 (ataxina 2), por exemplo.

Os exemplos de proteínas associadas a Distúrbios de Neurotransmissão incluem SST (somatostatina), NOS1 (sintase de óxido de azoto 1 (neuronal)), ADRA2A (recetor adrenérgico, alfa-2A), ADRA2C (recetor adrenérgico, alfa-2C), TACR1 (recetor de taquicininas 1), ou HTR2c (recetor 2C de 5-hidroxitriptaminas (serotonina)), por exemplo.

Os exemplos de sequências associadas a neurodesenvolvimento incluem A2BP1 [proteína 1 de ligação a ataxina 2], AADAT [aminoadipato 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 ATP, subfamília A (ABC1), membro 13], por exemplo.

Outros exemplos de condições preferidas que podem ser tratadas com o presente sistema podem ser selecionados a partir de: Síndrome de Aicardi-Goutières; Doença de Alexander; Síndrome de Allan-Herndon-Dudley; Distúrbios Relacionados com POLG; Alfa-Manosidose (Tipo II e III); Síndrome de Alström; Síndrome de Angelman; Ataxia-Telangiectasia; Lipofuscinoses Ceroides Neurais; Beta-Talassemia; Atrofia Ótica Bilateral e Atrofia Ótica (Infantil) Tipo 1; Retinoblastoma (bilateral); Doença de Canavan; Síndrome Cerebro-óculo-facio-esquelético 1 [COFS1]; Xantomatose Cerebrotendinosa; Síndrome de Cornelia de Lange; Distúrbios Relacionados com MAPT; Doenças Genéticas de Prião; Síndrome de Dravet; Doença de Alzheimer Familiar de Início Precoce; Ataxia de Friedreich [FRDA]; Síndrome de Fryns; Fucosidose; Distrofia Muscular Congénita de Fukuyama; Galactosialidose; Doença de Gaucher; Acidemias Orgânicas; Linfocitose Hemofagocítica; Síndrome de Hutchinson-Gilford ou Progeria; Mucopolidose II; Doença do Armazenamento de Ácido Siálico Livre Infantil; Neurodegeneração Associada a PLA2G6; Síndrome de Jervell e Lange-Nielsen; Epidermólise Bolhosa Juncional; Doença de

Huntington; Doença de Krabbe (Infantil); Síndrome de Leigh Associado a ADN Mitocondrial e NARP; Síndrome de Lesch-Nyhan; Lisencefalia Associada a LIS1; Síndrome de Lowe; Doença da Urina em Xarope de Ácer; Síndrome de Duplicação MECP2; Distúrbios de Transporte de Cobre Relacionado com ATP7A; Distrofia Muscular Relacionada com LAMA2; Deficiência de Arilsulfatase A; Mucopolissacaridose Tipos I, II ou III; Distúrbios da Biogénese de Peroxissomas, Espetro da Síndrome de Zellweger; Neurodegeneração com Distúrbios da Acumulação de Ferro no Cérebro; Deficiência de Esfingomielinase Ácida; Doença de Niemann-Pick Tipo C; Encefalopatia por Glicina; Distúrbios Relacionados ARX; Distúrbios do Ciclo de Ureia; Osteogénese Imperfeita Relacionada com COL1A1/2; Síndromes de Deleção de ADN Mitocondrial; Distúrbios Relacionados com PLP1; Síndrome de Perry; Síndrome de Phelan-McDermid; Doença de Armazenamento de Glicogénio 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 Deaminase; Síndrome de Smith-Lemli-Opitz; Atrofia Muscular Espinal; Ataxia Espinocerebelar de Início Infantil; Deficiência de Hexosaminidase A; Displasia Tanatofórica Tipo 1; Distúrbios Relacionados com Colagénio Tipo VI; Síndrome de Usher Tipo 1; Distrofia Muscular Congénita; Síndrome de Wolf-Hirschhorn; Deficiência de Lipase Ácida Lisossomal; e Xeroderma Pigmentoso.

A administração crónica de terapêutica de proteínas pode desencadear respostas imunitárias inaceitáveis à proteína específica. A imunogenicidade de fármacos de proteína pode ser atribuída a alguns epítomos de linfócitos T auxiliares (HTL) imunodominantes. A redução da afinidade de ligação a MHC destes epítomos HTL existentes nestas proteínas pode gerar fármacos com uma imunogenicidade mais

baixa (Tangri S, et al. ("Rationally engineered therapeutic proteins with reduced immunogenicity" *J. Immunol.* 15 de março de 2005;174(6):3187 a 96.) Em particular, a imunogenicidade da enzima CRISPR pode ser reduzida a seguir à abordagem apresentada em primeiro lugar em Tangri et al. relativamente a eritropoietina e subsequentemente desenvolvida. Em conformidade, a evolução direcionada ou a concepção racional podem ser utilizadas para reduzir a imunogenicidade do sistema CRISPR (por exemplo, uma Cas9) na espécie hospedeira (espécie humana ou outra).

Nas plantas, os agentes patogénicos são muitas vezes específicos do hospedeiro. Por exemplo, *Fusarium oxysporum* f. sp. *lycopersici* causa o emurchecimento do tomate, mas apenas ataca o tomate, e *F. oxysporum* f. *dianthii* *Puccinia graminis* f. sp. *tritici* apenas ataca o trigo. As plantas têm defesas existentes e induzidas para resistir à maioria dos agentes patogénicos. As mutações e os eventos de recombinação nas gerações de plantas originam uma variabilidade genética que provoca suscetibilidade, especialmente porque os agentes patogénicos se reproduzem com mais frequência do que as plantas. Nas plantas, pode existir uma resistência de não-hospedeiro, p. ex., o hospedeiro e o agente patogénico são incompatíveis. Igualmente, pode existir uma Resistência Horizontal, p. ex., resistência parcial contra todos os géneros de agente de patogénico, habitualmente controlada por muitos genes, e Resistência Vertical, p. ex., resistência completa a alguns géneros de um agente patogénico, mas não a outros géneros, habitualmente controlada por alguns genes. Num nível Gene-para-Gene, as plantas e os agentes patogénicos evoluem em conjunto, e as alterações genéticas num equilibram as alterações no outro. Em conformidade, utilizando a Variabilidade Natural, os criadores combinam os genes mais úteis para Quantidade Produzida, Qualidade, Uniformidade, Vigor, Resistência. As origens de genes de resistência

incluem Variedades nativas ou estranhas, Variedades Heirloom, Familiares de Plantas Selvagens e Mutações Induzidas, p. ex., através de tratamento do material de plantas com agentes mutagénicos. Mediante a utilização da presente divulgação, é fornecida aos criadores de plantas uma nova ferramenta para induzir mutações. Em conformidade, um perito na técnica pode analisar o genoma de origens de genes de resistência e, nas Variedades que têm as características ou particularidades desejadas, utilizar a presente invenção para induzir o aumento de genes de resistência com mais precisão dos que os anteriores agentes mutagénicos e, por isso, acelerar e melhorar os programas de criação de plantas.

Tal como será evidente, prevê-se que o presente sistema possa ser utilizado para escolher como alvo qualquer sequência de polinucleótidos de interesse. Alguns exemplos de condições ou doenças que talvez sejam utilmente tratadas utilizando o presente sistema são incluídos nas Tabelas acima, e os exemplos de genes atualmente associados a essas condições são igualmente aqui fornecidos. Contudo, os genes exemplificados não são exaustivos.

EXEMPLOS

Os exemplos seguintes são fornecidos para efeitos de ilustração de várias formas de realização da invenção e não pretendem limitar de modo algum a presente invenção. Os presentes exemplos, juntamente com os métodos aqui descritos, são atualmente representativos de formas de realização preferidas, são exemplares e não se destinam a ser limitações no âmbito da invenção.

Exemplo 1: Atividade do Complexo CRISPR no Núcleo de uma Célula Eucariótica

Um sistema CRISPR de tipo II de exemplo corresponde ao locus de CRISPR de tipo II de *Streptococcus pyogenes* SF370, que contém um agrupamento (*cluster*) de quatro genes Cas9, Cas1, Cas2 e Csn1, bem como dois elementos de ARN de não-

codificação, ARNtracr e uma disposição característica de sequências repetitivas (repetições diretas) espaçadas por alongamentos pequenos de sequências não repetitivas (separadores, cerca de 30 pb cada). Neste sistema, a quebra de filamento duplo (DSB) de ADN escolhido como alvo é gerada em quatro etapas subsequentes (Figura 2A). Em primeiro lugar, dois ARNs de não-codificação, a disposição de pré-ARNcr e ARNtracr são transcritos a partir do *locus* de CRISPR. Em segundo lugar, o ARNtracr é hibridizado para as repetições diretas de pré-ARNcr, que é depois processado em ARNscr maduros que contêm sequências individuais de separador. Em terceiro lugar, o complexo ARNcr:ARNtracr maduro direciona Cas9 para o alvo de ADN que consiste no *protospacer* e PAM correspondente através de formação heterodupla entre a região de separador do ARNcr e o ADN de *protospacer*. Finalmente, Cas9 medeia a clivagem de ADN-alvo a montante de PAM para criar uma DSB no *protospacer* (Figura 2A). Este exemplo descreve um processo de exemplo para adaptar este sistema de nuclease de ARN programável para direcionar a atividade do complexo CRISPR nos núcleos de células eucarióticas.

Para melhorar a expressão de componentes de CRISPR nas células de mamíferos, foram otimizados os codões de dois genes do *locus* 1 SF370 de *Streptococcus pyogenes* (*S. pyogenes*), *Cas9* (*SpCas9*) e *RNase III* (*SpRNase III*). Para facilitar a localização nuclear, foi incluído um sinal de localização nuclear (NLS) nos terminais amino (N) ou carboxilo (C) de ambas as *SpCas9* e *SpRNase III* (Figura 2B). Para facilitar a visualização da expressão de proteínas, um marcador de proteína fluorescente também foi incluído nos terminais N ou C de ambas as proteínas (Figura 2B). Igualmente, foi gerada uma versão de *SpCas9* com um NLS ligado a ambos os terminais N e C (2xNLS-*SpCas9*). As construções que contêm *SpRNase III* e *SpCas9* fundidas em NLS foram transfetadas em células de rim embrionário humano

(HEK) 293FT, e verificou-se que o posicionamento relativo da NLS para SpCas9 e SpRNase III afeta a respectiva eficiência de localização nuclear. Ao passo que a NLS de C-terminal foi suficiente para escolher a SpRNase III como alvo para o núcleo, a ligação de uma única cópia desta NLS específica ao N-terminal ou C-terminal de SpCas9 não conseguiu alcançar a localização nuclear adequada neste sistema. Neste exemplo, a NLS de C-terminal correspondia à da nucleoplasmina (KRPAATKKAGQAKKKK), e a NLS de C-terminal correspondia à NLS do T-antigénio grande SV40 (PKKKRKV). Das versões de SpCas9 testadas, apenas 2xNLS-SpCas9 apresentou localização nuclear (Figura 2B).

O ARNtracr do locus de CRISPR de *S. pyogenes* SF370 tinha dois locais de início de transcrição, originando duas transcrições de 89 nucleótidos (nt) e 171 nt que são subsequentemente processadas nos ARNstracr maduros de 75 nt idênticos. O ARNtracr de 89 nt mais pequeno foi selecionado para a expressão nas células de mamífero (construções de expressão ilustradas na Figura 6, com funcionalidade conforme determinado pelos resultados do ensaio Surveyor ilustrados na Figura 6B). Os locais de início de transcrição estão marcados como +1, e o terminador de transcrição e a sequência examinada por *Northern Blot* são igualmente indicados. A expressão de ARNtracr processado também foi confirmada por *Northern Blot*. A Figura 7C ilustra resultados de uma análise por *Northern Blot* de ARN total extraído de células 293FT transfectadas com construções de expressão U6 que transportam ARNtracr longo ou pequeno, bem como SpCas9 e DR-EMX1(1)-DR. Os painéis esquerdo e direito correspondem às células 293FT transfectadas sem ou com SpRNase III, respetivamente. U6 indica um controlo de carregamento transferido eletroforeticamente (*blotted*) com uma sonda que escolhe como alvo ARNpn U6 humano. A transfeção da construção de expressão de ARNtracr pequeno originou níveis abundantes da

forma processada de ARNtracr (~75 pb). Quantidades muito pequenas de ARNtracr longo são detetadas na análise por *Northern blot*.

Para provocar uma iniciação de transcrição precisa, foi selecionado o promotor U6 baseado em ARN polimerase III para ativar a expressão de ARNtracr (Figura 2C). De forma semelhante, uma construção baseada no promotor U6 foi desenvolvida para expressar uma disposição de pré-ARNcr que consiste num único separador flanqueado por duas repetições diretas (DRs, igualmente abrangidas pelo termo "sequências tracr-mate"; Figura 2C). O separador inicial foi concebido para escolher como alvo um local-alvo de 33 pares de bases (pb) (uma sequência de *protospacer* de 30 pb mais um motivo CRISPR de 3 pb (PAM) que cumpre o motivo de reconhecimento NGG de Cas9) no *locus* de *EMX1* humano (Figura 2C), um gene essencial no desenvolvimento no córtex cerebral.

Para testar se a expressão heteróloga do sistema CRISPR (SpCas9, SpRNase III, ARNtracr e pré-ARNcr) nas células de mamífero pode alcançar a clivagem escolhida como alvo de cromossomas de mamífero, as células HEK 293FT foram transfetadas com combinações de componentes de CRISPR. Uma vez que as DSBs nos núcleos de mamíferos são parcialmente reparadas pela via de junção terminal não homóloga (NHEJ), que origina a formação de indels (inserções/deleções), o ensaio Surveyor foi utilizado para detetar a potencial atividade de clivagem no *locus* de *EMX1*-alvo (consulte, p. ex. Guschin et al., 2010, *Methods Mol Biol* 649: 247). A cotransfecção de todos os quatro componentes de CRISPR foi capaz de induzir até 5,0% de clivagem no *protospacer* (consulte a Figura 2D). A cotransfecção de todos os componentes de CRISPR menos SpRNase III induziu igualmente até 4,7% de indel no *protospacer*, sugerindo que podem existir RNases endógenas de mamífero que são capazes de ajudar na maturação de ARNcr, tal como por exemplo as enzimas Dicer e Drosha relacionadas. A remoção de qualquer

um dos restantes três componentes anulou a atividade de clivagem do genoma do sistema CRISPR (Figura 2D). A sequenciação de Sanger de *amplicons* com o *locus*-alvo observou a atividade de clivagem: em 43 clones sequenciados, e foram detetados 5 alelos mutantes (11,6%). Experiências semelhantes utilizando uma variedade de sequências-guia produziram percentagens de indel tão elevadas como 29% (consulte as Figuras 4 a 8, 10 e 11). Estes resultados definem um sistema de três componentes para uma modificação eficiente de genoma mediado por CRISPR nas células de mamífero.

Para otimizar a eficiência de clivagem, os Requerentes também testaram se diferentes isoformas de ARNtracr afetaram a eficiência de clivagem e verificaram que, neste sistema de exemplo, apenas a forma de transcrição curta (89 pb) foi capaz de mediar a clivagem do *locus* genómico de *EMX1* humano. A Figura 9 fornece uma análise por *Northern blot* adicional de processamento de ARNcr nas células de mamífero. A Figura 9A ilustra um esquema que ilustra o vetor de expressão para um único separador flanqueado por duas repetições diretas (DR-*EMX1*(1)-DR). O separador de 30 pb que escolhe como alvo o *protospacer* 1 do *locus* de *EMX1* humano e as sequências de repetições diretas são ilustrados na sequência por baixo da Figura 9A. A linha indica a região cuja sequência de complemento reverso foi utilizada para gerar sondas de *Northern blot* para a deteção de ARNcr *EMX1*(1). A Figura 9B ilustra uma análise por *Northern Blot* de ARN total extraído de células 293FT transfectadas com construções de expressão U6 que transportam DR-*EMX1*(1)-DR. Os painéis esquerdo e direito correspondem às células 293FT transfectadas sem ou com SpRNase III, respetivamente. DR-*EMX1*(1)-DR foi processado em ARNscr maduros apenas na presença de SpCas9 e ARNtracr pequeno, e não se encontrava dependente da presença de SpRNase III. O ARNcr maduro detetado a partir do ARN total 293FT transfectado tem ~33 pb

e é mais pequeno do que o ARNcr maduro de 39 a 42 pb de *S. pyogenes*. Estes resultados demonstram que um sistema CRISPR pode ser transplantado nas células eucarióticas e reprogramado 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 um esquema que ilustra o locus 1 de CRISPR de *Streptococcus pyogenes* SF370 e um mecanismo de clivagem de ADN mediada por CRISPR proposto por este sistema. O ARNcr maduro processado a partir da disposição de repetições diretas-separadores direciona Cas9 para alvos genómicos que consistem em *protospacers* complementares e num *protospacer*-motivo adjacente (PAM). Nos pares de bases de separador-alvo, a Cas9 medeia uma quebra de filamento duplo no ADN-alvo. A Figura 2B ilustra a engenharia de Cas9 de *S. pyogenes* (SpCas9) e RNase III (SpRNase III) com sinais de localização nuclear (NLSs) para permitir a importação para o núcleo mamífero. A Figura 2C ilustra a expressão de mamífero de SpCas9 e SpRNase III ativada pelo promotor EF1a constitutivo e a disposição de ARNtracr e pré-ARNcr (DR-Separador-DR) ativada pelo promotor U6 Pol3 de ARN para provocar a iniciação e terminação precisas da transcrição. Um *protospacer* do locus de *EMX1* humano com uma sequência PAM satisfatória é utilizado como o separador na disposição de pré-ARNcr. A Figura 2D ilustra o ensaio de nuclease Surveyor para inserções e deleções menores mediadas por SpCas9. A SpCas9 foi expressada com e sem SpRNase III, ARNtracr e uma disposição de pré-ARNcr que transporta o separador de *EMX1*-alvo. A Figura 2E ilustra uma representação esquemática de pares de bases entre o locus-alvo e ARNcr de *EMX1*-alvo, bem como um cromatograma de exemplo que mostra uma microdeleção adjacente ao local de clivagem de SpCas9. A Figura 2F ilustra alelos mutantes identificados a partir da análise de sequenciação de 43

amplicons clonais que mostram uma variedade de microinserções e microdeleções. Os traços indicam bases eliminadas, e as bases desalinhadas ou dissonantes indicam inserções ou mutações. Barra de escala = 10 μ m.

Para simplificar mais o sistema de três componentes, foi adaptada uma conceção de híbrido ARNcr quimérico-ARNtracr, em que um ARNcr maduro (compreendendo uma sequência-guia) é fundido num ARNtracr parcial através de uma estrutura em pedúnculo para simular a dupla ARNcr:ARNtracr natural (Figura 3A).

As sequências-guia podem ser inseridas entre locais BbsI utilizando oligonucleótidos temperados. Os *protospacers* nos filamentos *sense* e *anti-sense* são indicados acima e abaixo das sequências de ADN, respetivamente. Uma taxa de modificação de 6,3% e 0,75% foi alcançada para os *loci* de *PVALB* humano e *Th* de ratinho respetivamente, demonstrando a vasta aplicabilidade do sistema CRISPR na modificação de diferentes *loci* em múltiplos organismos. Embora a clivagem só tenha sido detetada com um de três separadores para cada *locus* utilizando as construções quiméricas, todas as sequências-alvo foram clivadas com a eficiência de produção de indels a atingir 27% ao utilizar a disposição de pré-ARNcr coexpressado (Figuras 4 e 5).

A Figura 5 fornece uma outra ilustração em que é possível reprogramar SpCas9 para escolher como alvo múltiplos *loci* genómicos nas células de mamífero. A Figura 5A fornece um esquema dos *locis* de *EMX1* humano que mostra a localização de cinco *protospacers*, indicados pelas sequências sublinhadas. A Figura 5B fornece um esquema do complexo pré-ARNcr/ARNtracr que mostra a hibridização entre a região de repetições diretas do pré-ARNcr e ARNtracr (parte superior), e um esquema de uma conceção de ARN quimérico compreendendo uma sequência-guia de 20 pb, e sequências tracr-mate e tracr que consistem numa repetição

direta parcial e sequências de ARNtracr hibridizadas numa estrutura em gancho (parte inferior). Os resultados de um ensaio Surveyor que comparam a eficácia da clivagem mediada por Cas9 em cinco *protospacers* no *locus* de *EMX1* humano são ilustrados na Figura 5C. Cada *protospacer* é escolhido como alvo utilizando o complexo pré-ARNcr/ARNtracr (ARNcr) processado ou o ARN quimérico (ARNqui).

Uma vez que a estrutura secundária de ARN pode ser crucial para interações intermoleculares, foi utilizado um algoritmo de previsão de estrutura baseado no conjunto de estrutura com peso de Boltzmann e energia livre mínima para comparar a estrutura secundária putativa de todas as sequências-guia utilizadas na experiência de escolha de genomas-alvo (Figura 3B) (consulte p. ex. Gruber et al., 2008, *Nucleic Acids Research*, 36: W70). A análise revelou que, na maioria dos casos, as sequências-guia eficazes no contexto de ARNcr quimérico eram substancialmente livres de motivos de estrutura secundária, ao passo que as sequências-guia ineficazes tinham mais probabilidade de formar estruturas secundárias internas que podem prevenir pares de bases com o ADN de *protospacer*-alvo. Deste modo, é possível que a variabilidade na estrutura secundária de separador afete a eficiência da interferência mediada por CRISPR ao utilizar um ARNcr quimérico.

A Figura 3 ilustra vetores de expressão de exemplo. A Figura 3A fornece um esquema de um vetor bicistrônico para ativar a expressão de uma quimera de ARNcr sintético-ARNtracr (ARN quimérico), bem como SpCas9. O ARN-guia quimérico contém uma sequência-guia de 20 pb correspondente ao *protospacer* existente no local genómico-alvo. A Figura 3B fornece um esquema que mostra sequências-guia que escolhem como alvo os *loci* de *EMX1* humano, *PVALB* e *Th* de ratinho, bem as respectivas estruturas secundárias previstas. A eficiência de modificação em cada local-alvo é indicada por baixo da figura da estrutura secundária de ARN

(*EMX1*, $n = 216$ leituras de sequenciação de *amplicons*; *PYALB*, $n = 224$ leituras; *Th*, $n = 265$ leituras). O algoritmo de dobramento produziu um resultado com cada base colorida de acordo com a respetiva probabilidade de assumir a estrutura secundária prevista, conforme indicado por uma escala de arco-íris que é reproduzida na Figura 3B na escala de cinzentos. Outras conceções de vetor para SpCas9 são ilustradas na Figura 3A, incluindo vetores de expressão única que incorporam um promotor U6 ligado a um local de inserção para um oligo-guia, e um promotor Cbh ligado à sequência de codificação de SpCas9.

Para testar se os separadores que contêm estruturas secundárias são capazes de funcionar nas células procarióticas onde as CRISPRs atuam naturalmente, a interferência de transformação de plasmídeos que suportam *protospacers* foi testada numa estirpe *E. coli* que expressa de forma heteróloga o *locus* 1 de CRISPR SF370 de *S. pyogenes* (Figura 3C). O *locus* de CRISPR foi clonado para um vetor de expressão *E. coli* de poucas cópias e a disposição de ARNcr foi substituída por um único separador flanqueado por um par de DRs (pCRISPR). As estirpes de *E. coli* que albergam diferentes plasmídeos de pCRISPR foram transformadas com plasmídeos de estímulo que contêm o *protospacer* e as sequências PAM correspondentes (Figura 3C). No ensaio bacteriano, todos os separadores facilitaram uma interferência de CRISPR eficaz (Figura 3C). Estes resultados sugerem que podem existir fatores adicionais que afetam a eficiência da atividade de CRISPR nas células de mamífero.

Para investigar a especificidade da clivagem mediada por CRISPR, o efeito de mutações de nucleótido único na sequência-guia na clivagem de *protospacer* no genoma de mamífero foi analisado utilizando uma série de ARNscr quiméricos de *EMX1*-alvo com mutações pontuais únicas (Figura 4A). A Figura 4B ilustra os resultados de um ensaio

de nuclease Surveyor que comparam a eficiência de clivagem de Cas9 quando emparelhada com diferentes ARNs quiméricos mutantes. A dissonância de base única até 5' de 12 pb do PAM revogou substancialmente a clivagem genômica por SpCas9, ao passo que os separadores com mutações em posições a montante mais distantes mantiveram a atividade contra o alvo de *protospacer* original (Figura 4B). Além do PAM, a SpCas9 tem especificidade de base única nos últimos 12 pb do separador. Além disso, CRISPR é capaz de mediar a clivagem genômica tão eficientemente como um par de nucleases TALE (TALEN) que escolhem como alvo o mesmo *protospacer* de *EMX1*. A Figura 4C fornece um esquema que mostra a concepção de TALENs que escolhem *EMX1* como alvo, e a Figura 4D ilustra um gel Surveyor que compara a eficiência de TALEN e Cas9 (n=3).

Após o estabelecimento de um conjunto de componentes para alcançar a edição de genes mediada por CRISPR nas células de mamífero através do mecanismo NHEJ sujeito a erros, foi testada a capacidade de CRISPR para estimular a recombinação homóloga (HR), uma via de reparação de genes de alta-fidelidade para efetuar edições precisas no genoma. A SpCas9 de tipo selvagem pode mediar DSBs específicas do local que podem ser reparadas através de NHEJ e HR. Além disso, uma substituição de aspartato por alanina (D10A) no domínio catalítico RuvC I de SpCas9 foi produzida para converter a nuclease numa *nickase* (SpCas9n; ilustrada na Figura 5A) (consulte p. ex. Sapranasaks et al., 2011, *Cucleic Acis Research*, 39: 9275; Gasiunas et al., 2012, *Proc. Natl. Acad. Sci. E.U.A.*, 109:E2579), de modo a que o ADN genômico entalhado seja submetido à reparação direcionada para homologia (HDR) de alta-fidelidade. O ensaio Surveyor confirmou que SpCas9n não gera indels no alvo de *protospacer* de *EMX1*. Conforme ilustrado na Figura 5B, a coexpressão de ARNcr quimérico de *EMX1*-alvo com SpCas9 produziu indels no local-alvo, ao passo que a

coexpressão com SpCas9n não produziu (n=3). Além do mais, a sequenciação de 327 *amplicons* não detetou nenhuma indel induzida por SpCas9n. O mesmo *locus* foi selecionado para testar HR mediada por CRISPR através da cotransfecção de células HEK 293FT com o ARN quimérico que escolhe como alvo *EMX1*, hSpCas9 ou hSpCas9n, bem como um modelo HR para introduzir um par de locais de restrição (*HindIII* e *NheI*) junto ao *protospacer*. A Figura 5C fornece uma ilustração esquemática da estratégia HR, com localizações relativas de pontos de recombinação e sequências de *têmpera* de *primers* (setas). Na realidade, SpCas9 e SpCas9n catalisaram a integração do modelo HR para o *locus* de *EMX1*. A amplificação PCR da região-alvo seguida pelo resumo de restrição com *HindIII* revelou produtos de clivagem correspondentes a tamanhos de fragmento esperados (setas na análise do gel de polimorfismo de comprimento de fragmento de restrição ilustradas na Figura 5D), com SpCas9 e SpCas9n a mediar níveis semelhantes de eficiências HR. Os Requerentes verificaram ainda HR utilizando a sequenciação de Sanger de *amplicons* genômicos (Figura 5E). Estes resultados demonstram a utilidade de CRISPR para facilitar a inserção de genes-alvo no genoma de mamífero. Dada a especificidade do alvo de 14 pb (12 pb do separador e 2 pb do PAM) da SpCas9 de tipo selvagem, a disponibilidade de uma *nickase* pode reduzir significativamente a probabilidade de modificações fora do alvo, uma vez que as quebras de filamento único não são substratos para a via NHEJ sujeita a erros.

As construções de expressão que simulam a arquitetura natural de *loci* de CRISPER com separadores dispostos por ordem (Figura 2A) foram construídas para testar a possibilidade de escolha de sequências-alvo multiplexada. Mediante a utilização de uma única disposição de CRISPR que codifica um par de separadores de escolha de *EMX1* e *PVALB* como alvo, foi detetada uma clivagem eficiente em ambos os

loci (Figura 4F, que ilustra uma concepção esquemática da disposição de ARNcr e um *blot* Surveyor que ilustra uma mediação eficiente de clivagem). A deleção-alvo de regiões genômicas maiores através de DSBs simultâneas utilizando separadores contra dois alvos no *EMX1* espaçado por 119 pb foi igualmente testada, e foi detetada uma eficácia de deleção de 1,6% (3 de 182 *amplicons*; Figura 5G). Isto demonstra que o sistema CRISPR pode mediar a edição multiplexada num único genoma.

Exemplo 2: Alternativas e modificações do sistema CRISPR

A capacidade para utilizar o ARN para programar a clivagem de ADN específica da sequência define uma nova classe de ferramentas de engenharia de genomas para uma variedade de aplicações industriais e de investigação. Diversos aspetos do sistema CRISPR podem ser ainda melhorados para aumentar a eficiência e a versatilidade da escolha de CRISPR como alvo. Uma atividade de Cas9 ideal pode depender da disponibilidade de Mg^{2+} livre em níveis superiores aos presentes no núcleo mamífero (consulte p. ex. Jinek et al., 2012, *Science*, 337:816), e a preferência por um motivo NGG imediatamente a jusante do *protospacer* restringe a capacidade de escolher como alvo em média todos os 12 pb no genoma humano. Alguns destes constrangimentos podem ser superados explorando a diversidade de *loci* de CRISPR no metagenoma microbiano (consulte p. ex. Makarova et al., 2011, *Nat Rev Microbiol*, 9:467). Outros *loci* de CRISPR podem ser transplantados no meio celular de mamífero através de um processo semelhante ao descrito no Exemplo 1. A eficiência de modificação em cada local-alvo é indicada por baixo das estruturas secundárias de ARN. O algoritmo que gera as estruturas pinta cada base de acordo com a respetiva probabilidade de assumir a estrutura secundária prevista. Os separadores de guia ARN 1 e 2 induziram 14% e 6,4%, respetivamente. A análise estatística da atividade de clivagem na réplica biológica nestes dois locais de

protospacer é igualmente fornecida na Figura 7.

Exemplo 3: Algoritmo de seleção de sequência-alvo de amostra

Um programa de *software* foi concebido para identificar sequências-alvo candidatas de CRISPR em ambos os filamentos de uma sequência de ADN de entrada com base no comprimento desejado da sequência-guia e numa sequência de motivo CRISPER (PAM) para uma enzima de CRISPR especificada. Por exemplo, os locais-alvo para Cas9 de *S. pyogenes*, com sequências PAM NGG, podem ser identificados procurando por 5'-N_x-NGG-3' tanto na sequência de entrada como no complemento reverso da entrada. Igualmente, os locais-alvo para Cas9 de *S. thermophilus* CRISPR1, com sequência PAM NNAGAAW, podem ser identificados procurando por 5'-N_x-NNAGAAW-3' tanto na sequência de entrada como no complemento reverso da entrada. Igualmente, os locais-alvo para Cas9 de *S. thermophilus* CRISPR3, com sequência PAM NGGNG, podem ser identificados procurando por 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 utilizador, tal como 20.

Uma vez que múltiplas ocorrências no genoma do local-alvo de ADN podem originar uma edição de genomas não específica, após a identificação de todos os locais potenciais, o programa filtra as sequências com base no número de vezes que as mesmas aparecem no genoma de referência relevante. Para as enzimas de CRISPR para as quais a especificidade da sequência é determinada por uma sequência "semente" (*seed*), tal como 5' de 11 a 12 pb a partir da sequência PAM, incluindo a própria sequência PAM, a etapa de filtração pode basear-se na sequência semente. Deste modo, para evitar a edição nos *loci* genómicos adicionais, os resultados são filtrados com base no número de ocorrências da sequência semente:PAM no genoma relevante. O utilizador pode ter permissão para escolher o

comprimento da sequência semente. O utilizador também pode ter permissão para especificar o número de ocorrências da sequência semente:PAM num genoma para fins de passagem do filtro. A predefinição é procurar sequências exclusivas. O nível de filtração é mudado alterando tanto o comprimento da sequência semente como o número de ocorrências da sequência no genoma. O programa pode, além disso ou em alternativa, fornecer a sequência de uma sequência-guia complementar às sequências-alvo apresentadas fornecendo o complemento reverso das sequências-alvo identificadas.

Exemplo 4: Avaliação de múltiplos híbridos de ARNcr quimérico-ARNtracr

Este exemplo descreve os resultados obtidos para ARNs quiméricos (ARNsqui; compreendendo uma sequência-guia, uma sequência tracr-mate, e uma sequência tracr numa única transcrição) com sequências tracr que incorporam diferentes comprimentos de sequência de ARNtracr de tipo selvagem. A Figura 18a ilustra um esquema de um vetor de expressão bicistrónico para ARN quimérico e Cas9. A Cas9 é ativada pelo promotor CBh e o ARN quimérico é ativado por um promotor U6. O ARN-guia quimérico consiste numa sequência-guia de 20 pb (Ns) unida à sequência tracr (desde o primeiro "U" do filamento inferior até ao fim da transcrição), que se encontra truncada em várias posições conforme indicado. As sequências-guia e tracr são separadas pela sequência tracr-mate GUUUUAGAGCUA seguida pela sequência de estrutura GAAA. Os resultados de ensaios SURVEYOR para indels mediadas por Cas9 nos *loci* de *EMX1* humano e *PVALB* são ilustrados nas Figuras 18b e 18c, respetivamente. As setas indicam os fragmentos SURVEYOR esperados. Os ARNsqui são indicados pela respetiva designação "+n", e ARNcr refere-se a um ARN híbrido onde as sequências-guia e tracr são expressadas como transcrições separadas. A quantificação destes resultados, efetuada em triplicado, é ilustrada através do histograma nas Figuras

11a e 11b, correspondendo às Figuras 10b e 10c, respectivamente ("N.D." indica "nenhum indel detectado"). Os IDs de *protospacer* e respectivo alvo genômico correspondente, a sequência *protospacer*, a sequência PAM e a localização de filamento são fornecidos na Tabela D. As sequências-guia foram concebidas para serem complementares para toda a sequência *protospacer* em caso de transcrições separadas no sistema de híbridos, ou apenas para a parte sublinhada em caso de ARNs quiméricos.

Tabela D

ID de <i>proto-spacer</i>	alvo-genômico	sequência <i>protospacer</i> (5' a 3')	PAM	Filamento
1	<i>EMX1</i>	<u>GGACATCGATGTCACCTCCAATGACTAG</u> GG	TGG	+
2	<i>EMX1</i>	CATTGGAGGT <u>GACATCGATGTCCTCCCC</u> AT	TGG	-
3	<i>EMX1</i>	<u>GGAAGGGCCTGAGTCCGAGCAGAAGAA</u> GAA	GGG	+
4	<i>PVALB</i>	<u>GGTGGCGAGAGGGGCCGAGATTGGGTGT</u> TC	AGG	+
5	<i>PYALB</i>	ATGCAGGAGGGT <u>GGCGAGAGGGGCCGA</u> GAT	TGG	+

Transfecção e cultura de células

A linha celular de rim embrionário humano (HEK) 293FT (Life Technologies) foi mantida no Meio de Eagle Modificado por Dulbecco (DMEM) complementado com 10% de soro bovino fetal (HyClone), 2 mM de GlutaMAX (Life Technologies), 100U/mL de penicilina e 100 µg/mL de estreptomicina a 37 °C com 5% de incubação em CO₂. As células 293ET foram semeadas em placas de 24 poços (Corning) 24 horas antes da transfecção numa 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

utilizado um total de 500 ng de plasmídeo.

Ensaio SURVEYOR para modificação de genomas

As células 293FT foram transfetadas com ADN de plasmídeo conforme descrito acima. As células foram incubadas a 37 °C durante 72 horas pós-transfecção antes da extração de ADN genómico. O ADN genómico foi extraído utilizando a Solução de Extração de ADN QuickExtract (Epicentre) seguindo o protocolo do fabricante. Em resumo, as células peletizadas foram novamente suspensas na solução QuickExtract e incubadas a 65 °C durante 15 minutos e a 98 °C durante 10 minutos. A região genómica que flanqueia o local-alvo de CRISPR para cada gene foi amplificada por PCR (*primers* indicados na Tabela E), e os produtos foram purificados utilizando a Coluna Giratória QiaQuick (Qiagen) seguindo o protocolo do fabricante. Um total de 400 ng dos produtos PCR purificados foi misturado com 2 µl de tampão PCR de Taq ADN Polimerase 10X (Enzymatics) e água ultrapura até um volume final de 20 µl, e submetido a um processo de nova *têmpera* para permitir a formação heterodupla: 95 °C durante 10 min, 95 °C a 85 °C subindo em - 2 °C/s, 85 °C a 25 °C em - 0,25 °C/s, e 25 °C em pausa durante 1 minuto. Após a nova *têmpera*, os produtos foram tratados com nuclease SURVEYOR e intensificador S SURVEYOR (Transgenomics) seguindo o protocolo recomendado pelo fabricante, e analisados em 4 a 20% de geles de poliacrilamida TBE Novex (Life Technologies). Os geles foram coloridos com coloração de ADN SYBR Gold (Life Technologies) durante 30 minutos e representados por meio de imagens com um sistema de imagens de geles Gel Doc (Bio-rad). A quantificação baseou-se em intensidades de banda relativas.

Tabela E

nome do primer	alvo-genômico	sequência primer (5' a 3')
Sp-EMX1-F	<i>EMX1</i>	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	<i>EMX1</i>	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	<i>PVALB</i>	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	<i>PVALB</i>	GGCAGCAAACCTCCTTGTCT

Identificação Computacional de locais-alvo exclusivos de CRISPR

Para identificar locais-alvo exclusivos para a enzima Cas9 de *S. pyogenes* SF370 (SpCas9) no genoma humano, de ratinho, de rato, de peixe-zebra, de mosca da fruta e de *C. elegans*, foi desenvolvido um pacote de software para examinar ambos os filamentos de uma sequência de ADN e identificar todos os locais-alvo de SpCas9 possíveis. Para este exemplo, cada local-alvo de SpCas9 foi definido de modo operacional como uma sequência de 20 pb seguida por uma sequência *protospacer*-motivo adjacente (PAM) NGG, e foram identificadas todas as sequências que cumprem esta definição 5'-N₂₀-NGG-3' em todos os cromossomos. Para prevenir a edição de genomas não específica, após a identificação de todos os locais potenciais, todos os locais-alvo foram filtrados com base no número de vezes que aparecem no genoma de referência relevante. Para tirar partido da especificidade da sequência da atividade de Cas9 conferida por uma sequência "semente", que pode ser, por exemplo, uma sequência 5' de aproximadamente 11 a 12 pb a partir da sequência PAM, as sequências 5'-NNNNNNNNNN-NGG-3' foram selecionadas para serem exclusivas no genoma relevante. Todas as sequências genômicas foram transferidas a partir do UCSC Genome Browser (hg19 do genoma humano, mm9 do genoma de ratinho, rn5 do genoma de rato, danRer7 do genoma de peixe-zebra, dm4 do genoma *D. melanogaster* e cel0 do genoma *C. elegans*). Os resultados de toda a pesquisa estão disponíveis para pesquisa utilizando as informações

do UCSC Genome Browser. A Figura 22 fornece uma visualização de exemplo de alguns locais-alvo no genoma humano.

Inicialmente, foram escolhidos como alvo três locais no *locus* de *EMX1* nas células HEK 293FT humanas. A eficiência da modificação de genomas de cada ARNqui foi avaliada utilizando o ensaio de nuclease SURVEYOR, que detecta mutações resultantes das quebras de filamento duplo (DSBs) de ADN e respetiva reparação subsequente através da via de reparação de danos de ADN de junção terminal não homóloga (NHEJ). As construções designadas por ARNqui(+n) indicam que o nucleótido até ao +n de ARNtracr de tipo selvagem está incluído na construção de ARN quimérico, com valores de 48, 54, 67 e 85 utilizados para n. Os ARNs quiméricos que contêm fragmentos mais longos de ARNtracr de tipo selvagem (ARNqui(+67) e ARNqui(+85)) mediaram a clivagem de ADN em todos os três locais-alvo de *EMX1*, em que ARNqui(+85), em particular, demonstra níveis significativamente mais elevados de clivagem de ADN do que os híbridos de ARNcr/ARNtracr correspondentes que expressaram sequências-guia e tracr em transcrições separadas (Figuras 10b e 10a). Dois locais no *locus* de *PVALB* que não produziram nenhuma clivagem detetável utilizando o sistema de híbridos (sequência-guia e sequência tracr expressadas como transcrições separadas) foram igualmente escolhidos como alvo utilizando ARNsqui. ARNqui(+67) e ARNqui(+85) foram capazes de mediar uma clivagem significativa nos dois *protospacers* de *PVALB* (Figuras 10c e 10b).

Para todos os cinco alvos nos *loci* de *EMX1* e *PVALB*, foi observado um aumento consistente na eficiência da modificação de genomas com um comprimento de sequência tracr crescente. Sem ser pretendida qualquer limitação a uma teoria, a estrutura secundária formada pela extremidade 3' do ARNtracr pode desempenhar um papel no melhoramento da

taxa de formação do complexo CRISPR. A Figura 21 fornece uma ilustração de estruturas secundárias previstas para cada um dos ARNs quiméricos utilizados neste exemplo. A estrutura secundária foi prevista utilizando dobraARN (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) utilizando a energia livre mínima e o algoritmo de função de partição. Uma pseudocor para cada base (reproduzida em escala de cinzentos) indica a probabilidade de formação de pares. Uma vez que os ARNsqui com sequências tracr mais longas foram capazes de clivar alvos que não foram clivados por híbridos de ARNcr/ARNtracr CRISPR nativos, é possível que o ARN quimérico seja carregado para Cas9 de forma mais eficiente do que o respetivo equivalente híbrido nativo. Para facilitar a aplicação de Cas9 para a edição de genomas específica do local nas células eucarióticas e nos organismos, todos os locais-alvo exclusivos previstos para a Cas9 de *S. pyogenes* foram identificados de forma computacional nos genomas humanos, de ratinho, de rato, de peixe-zebra, de *C. elegans* e de *D. melanogaster*. Os ARNs quiméricos podem ser concebidos para enzimas Cas9 a partir de outros micróbios para expandir o espaço-alvo de nucleases de ARN programável CRISPR.

As Figuras 11 e 21 ilustram vetores de expressão bicistrônicos exemplares para a expressão de ARN quimérico incluindo até ao nucleótido +85 da sequência ARN tracr de tipo selvagem, e SpCas9 com sequências de localização nuclear. A SpCas9 é expressada a partir de um promotor CBh e terminada com o sinal poliA bGH (pA bGH). A sequência expandida ilustrada imediatamente abaixo do esquema corresponde à região que circunda o local de inserção de sequência-guia e inclui, de 5' a 3', a parte 3' do promotor U6 (primeira região sombreada), locais de clivagem BbsI (setas), repetição direta parcial (sequência tracr-mate GTTTAGAGCTA, sublinhada), sequência de estrutura GAAA, e sequência tracr +85 (sequência sublinhada a seguir à

sequência de estrutura). Uma sequência-guia exemplar é ilustrada por baixo do local de inserção de sequência-guia, com nucleótidos da sequência-guia para um alvo selecionado representado por um "N".

As sequências descritas nos exemplos acima são conforme apresentado em seguida (as sequências de polinucleótidos são de 5' a 3'):

ARNtracr de U6 pequeno (*Streptococcus pyogenes* SF370):

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAAT
AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

(negrito = sequência de ARNtracr; sublinhado = sequência de terminador)

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

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTT
 CTTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTTGGAACC
 ATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG
 TGGCACCGAGTCGGTGCTTTTTTTT

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

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACCGGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAA
ACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

Estrutura principal de U6-ARN quimérico-BbsI (*Streptococcus pyogenes* SF370)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAT
GGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
TTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAAT
AGCAAGTTAAAATAAGGCTAGTCCG

NLS-SpCas9-EGFP :

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGIHGVPAADKKYSIGLDIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKADLRLLYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNEFL
AEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGAQSFIERMTN
FDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN

RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKHKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
 KTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
 ILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
 LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID
 NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
 LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVRINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
 IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERSSFENPIDFLEAKGYKEVKKDLIILPKYSLFELE
 NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFYK
 FDTTIDRKRYTSTKEVL DATLIHQSTGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILV
 ELDGDVNGHKFSVS GEGEDATY GKLTLKFICTTGKLPVPWPTLVTTLT YGVQCFSRYP
 DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
 GNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
 VLLPDNHYLSTQSALS KDPNEKRDMVLLFVTAAGITLGMDELYK

SpGas9-EGFP-NLS :

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHIPF
 GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFG
 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY
 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
 HAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASASQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPA
 FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL
 KHKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQG

DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
 SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
 YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLIT
 QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYG
 DYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
 VWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPK
 YGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKE
 VKKDLIIKLPKYSIFELENGRRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGS
 PEDNEQKQLFVEQHKHYLDEHIEQISEFSKRVLADANLDKVLSAYNKHRRDKPIREQAENI
 IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGDAA
 AVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWP
 TLVTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG
 DTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGVS
 QLADHYQQNTPIGDGPVLLPDNHYLSTQSALSCKPNEKRDHMLLEFVTAAGITLGMD
 ELYKKRPAATKAGQAKKKK

NLS-SpCas9-EGFP-NLS :

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIIHGVPAAADKKYSIGLDIGTNSVG
 WAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
 NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
 LRKKLVDS'TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
 ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
 AEDAKLQLSKD'TYD'DDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
 ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFEDQSKNGYAGYIDGGASQEEFYKFIK
 PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNR
 EKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
 FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRISRKLINGIRDKQSG
 KTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPAIKKG
 ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI

LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDNRLSDYDVDHIVPQSFLKDDSID
 NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
 LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
 IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKKGSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIKLPKYSLEFELE
 NGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGSPEDNEQKQLFVEQHKH
 YLDEHIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIHLFTLTNLGAPAAFKY
 FDTITDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDAAAVSKGEELFTGVVPILV
 EL DGDVNGHKFSVS GEGEGDATY GKLTLKFICTTGKLPVPWPTLVTTLT YGVQCFSRYP
 DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRA EVKFEGDTLVNRIELKGIDFKED
 GNILGHKLEYNYN SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
 VLLPDNHVYSTQSALSKDPNEKRDMVLLFEVTAAGITLGMDELYKKRPAATKKAGQA
 KKKK

NLS-SpCas9-NLS:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVG IHHGVPAADKKYSIGLDIGTNSVG
 WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
 NRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
 LRKKLV DSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFE
 ENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNEKSNFDL
 AEDAKLQLSKD TYDDDLNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLS
 ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK
 PILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNR
 EKIEKILTFRIPYYVGPLARONS RFAWMTRKSEETTPWNFE EVVDKGASAQSFIERMTN
 FDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
 KTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKG
 ILQTVKVVD ELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
 LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDNRLSDYDVDHIVPQSFLKDDSID

NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVVYDVRKMIKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKH
YLDEHIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK

NLS-mCherry-SpRNase3:

MFLFLSLTSLSSRTLVSKEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYE
GTQTAKLKVTKGGLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVM
NFEDGGVVTVTQDSSLQDGEFYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPED
GALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVE
QYERAEGRHSTGGMDELYKGSQLEELLSTSFDIQFNDLTLLTETAFTHTSYANEHRLN
VSHNERLEFLGDAVLQLIISEYLFAYPKKTEGDMSKLRSMIVREESLAGFSRFCSDAYI
KLKGEEKSGGRRRDTILGDLFEAFLGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVK
DYKTCLQEFLQTKGDVAIDYQVISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQD
AAKNALAQLSEV

SpRNase3-mCherry-NLS:

MKQLEELLSTSFDIQFNDLTLLTETAFTHTSYANEHRLNVSHNERLEFLGDAVLQLIISEY
LFAKYPKKTEGDMSKLRSMIVREESLAGFSRFCSDAYIKLKGEEKSGGRRRDTILGDL
FEAFLGALLLDKGIDAVRRFLKQVMIPQVEKGNFERVKDYKTCLQEFLQTKGDVAIDYQ
VISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQDAAKNALAQLSEVGSVSKGEE
DNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGLPFAWDIL
SPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFI
YKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAE
VKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKKRP
AATKKAGQAKKKK

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

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLaIGTNSVGW
 AVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
 RICYLQEIPSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
 RKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
 NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
 EDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSA
 SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI
 LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFPFLKDNREK
 IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFD
 KNLPNKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
 KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILE
 DIVLTILTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
 KTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
 ILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIIEGIKELGSQI
 LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID
 NKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
 LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
 IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIJKLPKYSLEFE
 NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKY
 FDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGDKRPAATKKAGQAKKKK

hEMX1-HR HindII-NheI Modelo:

GAATGCTGCCCTCAGACCCGCTTCCTCCCTGTCTGTCTGTCCAAGGAGAATGAGG
 TCTCACTGGTGGATTTCCGACTACCTGAGGAGCTGGCACCTGAGGGACAAGGCCC
 CCCACCTGCCAGCTCCAGCCTCTGATGAGGGGTGGGAGAGAGCTACATGAGGTTG
 CTAAGAAAGCCTCCCTGAAGGAGACCACACAGTGTGTGAGGTTGGAGTCTCTAGC
 AGCGGGTCTGTGCCCCAGGGATAGTCTGGCTGTCCAGGCACCTGCTCTTGATATAA
 ACACCACCTCCTAGTTATGAAACCATGCCCATCTGCCTCTCTGTATGGAAAAGAGC

ATGGGGCTGGCCCGTGGGGTGGTGTCCACTTTAGGCCCTGTGGGAGATCATGGGAA
 CCCACGCAGTGGGTCATAGGCTCTCTCATTTACTACTCACATCCACTCTGTGAAGAA
 GCGATTATGATCTCTCTCTAGAAACTCGTAGAGTCCCATGTCTGCCGGCTTCCAGA
 GCCTGCACTCCTCCACCTTGGCTTGGCTTTGCTGGGGCTAGAGGAGCTAGGATGCAC
 AGCAGCTCTGTGACCCCTTTGTTTGAGAGGAACAGGAAAACCACCCTTCTCTCTGGCC
 CACTGTGTCTCTTCTGCCCTGCCATCCCCCTTCTGTGAATGTTAGACCCATGGGAGC
 AGCTGGTCAGAGGGGACCCCGGCCTGGGGCCCCCTAACCTATGTAGCCTCAGTCTTC
 CCATCAGGCTCTCAGCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTGCTCTGGGGGCC
 TCCTGAGTTTCTCATCTGTGCCCCCTCCCTCCCTGGCCCCAGGTGAAGGTGTGGTTCCAG
 AACCGGAGGACAAAGTACAAACGGCAGAAAGCTGGAGGAGGAAGGGCCTGAGTCCG
 AGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAG
 GCCAATGGGGAGGACATCGATGTCACTCCAATGACaagcttgctageGGTGGGCAACCAC
 AAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCTGCGTGGGCCCAA
 GCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGGCCCCA
 CAGGGCTTGAAGCCCGGGGCCGCAATTGACAGAGGGACAAGCAATGGGCTGGCTGA
 GGCTGGGACCACTTGGCCTTCTCCTCGGAGAGCCTGCCCTGCCTGGGCGGGCCCGCC
 CGCCACCGCAGCCTCCCAGCTGCTCTCCGTGTCTCCAATCTCCCTTTTGTTTTGATGC
 ATTTCTGTTTTAATTTATTTTCCAGGCACCACTGTAGTTTAGTGATCCCCAGTGTCCC
 CCTTCCCTATGGGAATAATAAAAGTCTCTCTCTTAATGACACGGGCATCCAGCTCCA
 GCCCCAGAGCCTGGGGTGGTAGATTCCGGCTCTGAGGGCCAGTGGGGGCTGGTAGA
 GCAAACGCGTTCAAGGCCTGGGAGCCTGGGGTGGGGTACTGGTGGAGGGGGTCAAG
 GGTAATTCATTAACTCCTCTCTTTTGTGGGGGACCCTGGTCTCTACCTCCAGCTCCA
 CAGCAGGAGAAACAGGCTAGACATAGGGAAGGGCCATCCTGTATCTTGAGGGAGGA
 CAGGCCCAGGTCTTTCTTAACGTATTGAGAGGTGGGAATCAGGCCCAGGTAGTTCAA
 TGGGAGAGGGAGAGTGCTTCCCTCTGCCTAGAGACTCTGGTGGCTTCTCCAGTTGAG
 GAGAAACCAGAGGAAAGGGGAGGATTGGGGTCTGGGGGAGGGAACACCATTACAA
 AAGGCTGACGCTTCCAGTCCGAAGTCGTGGGCCCCACCAGGATGCTCACCTGTCTTG
 GAGAACCGCTGGGCAGGTTGAGACTGCAGAGACAGGGCTTAAGGCTGAGCCTGCAA
 CCAGTCCCCAGTGACTCAGGGCCTCCTCAGCCCAAGAAAGAGCAACGTGCCAGGGC
 CCGCTGAGCTCTTGTGTTACCTG

NLS-StCsn1-NLS:

MKRPAATKKAGQAKKKKSDLVLGLDIGIGSVGVGHILNKVTGEIHKNSRIFPAAQAENN
 LVRRTNRQGRRLARRKKHRRVRLNRLFEEESGLITDFTKISINLNPYQLRVKGLTDELSNE
 ELFIALKNMVXHRGISYLLDDASDDGNSSVGDYAQIVKENSQKLETKTPGGQIQLERYQTY
 GQLRGDFTVEKDGGKHLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKR
 KYHGPNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLND
 LNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKS
 GKAEHITFEAYRKMKTLETLDIEQMDRETLDKLAYVLTNTEREGIQEALEHEFADGSFS
 QKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSS
 SNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNVIEMARETNEDDEKK
 AIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLY
 TGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKQRTPYQALDSMD
 DAWSFRELKAFVRESKTLNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNA
 LQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIAASSQLNLWKK
 QKNTLVSYSEDQLLDIETGELISDDDEYKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSK
 FNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMY
 RHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKNGPEIKSLK
 YYDSKLGHNHIDITPKDSNNKVVLQSVSPWRADVFNKTTGKYEILGLKYADLQFEKGT
 GTYKISQEKYNDIKKKEGVDSSEFKFTLYKNDLLLVDKTETKEQQLRFLSRTMPKQK
 HYVELKPYDKQKFEGGEALIKVLGNVANSQGCKKGLGKSNISIYKVRTDVLGNQHIKN
 EGDKPKLDFKRPAATKKAGQAKKKK

U6-St_ARNtracr (7-97) :

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGA
 GAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
 ACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAATATGTTTTAAAT
 GGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATC
 TTGTGGAAAGGACGAAACACCGTTACTTAAATCTTGCAGAAGCTACAAAGATAAGG
 CTTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTTCGTTATTTAA

U6-DR-separador-DR (*S. pyogenes* SF370)

gagggcctatttcccatgattccttcataatgtgcatatacgatacaaggctgtt
 agagagataattggaattaatgtgactgtaaacacaaagatat
 tagtacaaaatacgtgacgtagaaagtaataatgttcttgggtagtttgcagttttaaaa
 ttatgttttaaaatggactatcatatgcttaccgtaactt

gaaagtatatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccgggttt
tagagctatgctgtttgaatgggtcccaaacNNN

NNgtttttagagctatgcttgttttgaatgggtccc
aaaac**TTTTTTT** (sublinhado em minúsculas = repetição direta;
N = sequência-guia; negrito = terminador)

ARN quimérico contendo +48 ARN tracr (*S. pyogenes* SF370)

gagggcctatttcccatgattccttcataatttgcatatacgatacaaggctgtt
agagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaa
ttatgtttttaaaatggactatcatatgcttaccgtaactt
gaaagtatatttcgatttcttggctttatatatcttgtggaaaggacgaaacaceNNNNNN
NNgtttttaga gctagaaatagcaagt-
taaaataaggctagtagccg**TTTTTTT** (N = sequência-guia; primeiro
sublinhado = sequência tracr-mate; segundo sublinhado =
sequência tracr; negrito = terminador)

ARN quimérico contendo +54 ARN tracr (*S. pyogenes* SF370)

gagggcctatttcccatgattccttcataatttgcatatacgatacaaggctgtt
agagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaa
ttatgtttttaaaatggactatcatatgcttaccgtaactt
gaaagtatatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNN
NNgtttttaga
gctagaaatagcaaggttaaaataaggctagtagccgttatca**TTTTTTTTT** (N =
sequência-guia; primeiro sublinhado = sequência tracr-mate;
segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico contendo +67 ARN tracr (*S. pyogenes* SF370)

gagggcctatttcccatgattccttcataatttgcatatacgatacaaggctgtt
agagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaa
ttatgtttttaaaatggactatcatatgcttaccgtaactt
gaaagtatatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNN
NNgtttttaga
gctagaaatagcaaggttaaaataaggctagtagccgttatcaacttgaaaaagt**TTTTTTT**
T (N = sequência-guia; primeiro sublinhado = sequência
tracr-mate; segundo sublinhado = sequência tracr; negrito =

terminador)

ARN quimérico contendo +85 ARN tracr (*S. pyogenes* SF370)

gagggcctatttcccatgattccttcataatttgcatatacgatacaaggctggt
agagagataattggaattaatttgactgtaaacacaaagatat
tagtacaaaatacgtgacgtagaaagtaataatttcttggtagtttgagttttaaaa
ttatgttttaaaaatggactatcatatgcttaccgtaactt
gaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccNNNNNN
NNNNNNNNNNNNNNgttttaga
gctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccg
agtcggtgc**TTTTTTT** (N = sequência-guia; primeiro sublinhado =
sequência tracr-mate; segundo sublinhado = sequência tracr;
negrito = terminador)

CBh-NLS-SpCas9-NLS

CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCC
ATIGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG
ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA

TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCA
 TTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTA
 GTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCT
 CCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCG
 ATGGGGGGCGGGGGGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGGCGGGGGCGGGGGCGAG
 GGGCGGGGGCGGGGGCGAGGCGGAGAGGTGCGGGCGGCAGCCAATCAGAGCGGCGCGC
 TCCGAAAAGTTTCCTTTTATGGCGAGGCGGGCGGGCGGGCGGGCCCTATAAAAAGCGA
 AGCGCGCGGGCGGGGGAGTCGCTGCGACGCTGCCTTCGCCCCGTGCCCCGCTCCG
 CCGCGCCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCACAGGTGA
 GCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAGGTAAGGG
 TTTAAGGGATGGTTGGTTGGTTGGGGTATTAATGTTTAATTACCTGGAGCACCTGCCT
 GAAATCACTTTTTTTTCAGGTTGGacgggtgccaccATGGACTATAAGGACCACGACGGAGA
CTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATC
GGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCGCGTGATCACCGACGAGTACAA
GGTGGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAACATCAAGA
AGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGGCGAAACAGCCGAGGCCACCCGG
CTGAAGAGAACCGCCAGAAAGATACACCAGACCGGAAGAACCGGATCTGCTATCT
GCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGAC
TGGAAGAGTCCTTCCTGGTGGAAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTC
GGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCT
GAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGGCTGATCTATCTGG
CCCTGGCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACC
CCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAG
CTGTTTCGAGGAAAACCCCATCAACGCCAGCGGGCGTGGACGCCAAGGCCATCCTGTC
TGCCAGACTGAGCAAGAGCAGACGGCTGGAAAAATCTGATCGCCCAGCTGCCCGGCG
AGAAGAAGAATGGCCTGTTCCGGCAACCTGATTGCCCTGAGCCTGGGGCTGACCCCC
AACCTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGA
CACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGGCGACCAGTACGCCG
ACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATAC

GACGAGCACCAACCAGGACCTGACCCTGCTGAAAAGCTCTCGTGCGGCAGCAGCTGCC
TGAGAAGTACAAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACA
TTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAA
AAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCG
GAAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGAGAGC
TGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCTGAAGGACAACCGGG
AAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCCTCTGGCCA
GGGGAACAGCAGATTTCGCTGGATGACCAGAAAAGAGCGAGGAAACCATCACCCC
CTGGAACTTCGAGGAAGTGTTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAUC
GGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGC
CTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACC
GAGGGAATGAGAAAAGCCCCGCTTCCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGG
ACCTGCTGTTCAAGACCAACCGGAAAAGTGACCGTGAAGCAGCTGAAAGAGGACTAC
TTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTT
AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTT
CCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACAC
TGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTC
GACGACAAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGC
TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
GATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGAC
GACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGA
TAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCA
TCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAAGTGATGGGCCGGCACAAG
CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGAC
AGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
CAGCCAGATCCTGAAAAGAACCCCCGTGGAAAAACCCAGCTGCAGAACGAGAAG
CTGTACCTGTACTACCTGCAGAAATGGCCGGGATATGTACGTGGACCAGGAACTGGA
CATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAA
GGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAG
AGCGACAACGTGCCCTCCGAAGAGGTGCTGAAGAAATGAAGAATACTGGCGGCA
GCTGCTGAACGCCAAGCTGATTACCCAGAGAAAAGTTCGACAATCTGACCAAGGCCG

AGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACAC
TAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGT
CCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCA
ACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTA
CGACGTGCGGAAGATGATCGCCAAGAGCGGAGCAGGAAATCGGCAAGGCTACCGCC
AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGA
TCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCC
CAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC
CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAG
CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCC
CTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACCTTCCTGTACCTGGCCA
GCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAACAGCTGTTT
GTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTC
CAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCCTGCCGCCITCAAGTACTTTGACACCACCATCGACCGG
AAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCAT
CACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACTTCTTTT
TCTTAGCTTGACCAGCTTCTTAGTAGCAGCAGGACGCTTTAA

(sublinhado = NLS-hSpCas9-NLS)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatctt
 scagaagctacaaagataaggctt
 catgccgaaatcaacaccctgtcatttttatggcagggtgttttcggtatttta**TTTTTT**

(N = sequência-guia; primeiro sublinhado = sequência tracr-mate; segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S.*

thermophilus (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaa
gataaggcttcatgccgaaatca
acaccctgtcattttatggcaggggtgttttcggtatttaa**TTTTTT** (N =
 sequência-guia; primeiro sublinhado = sequência de tracr-
 mate; segundo sublinhado = sequência tracr; negrito =
 terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaa
gataaggcttcatgccgaaatca acaccctgt-cattttatggcaggggtgt**TTTTTT**
 (N = sequência-guia; primeiro sublinhado = sequência tracr-
 mate; segundo sublinhado = sequência tracr; negrito =
 terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatctt
gcagaagctacaaagataaggctt
catgccgaaatcaacaccctgtcattttatggcaggggtgttttcggtatttaa**TTTTTT**
 (N = sequência-guia; primeiro sublinhado = sequência tracr-
 mate; segundo sublinhado = sequência tracr; negrito =
 terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaa
gataaggcttcatgccgaaatc
aacaccctgtcattttatggcaggggtgttttcggtatttaa**TTTTTT** (N =
 sequência-guia; primeiro sublinhado = sequência tracr-mate;
 segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaaa
gataaggcttcatgccgaaatc aacaccctgt-cattttatggcaggggtgt**TTTTTT**
 (N = sequência-guia; primeiro sublinhado = sequência tracr-
 mate; segundo sublinhado = sequência tracr; negrito =

terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaagatttaGAAAtaaatcttgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcaggggtgttttcggtattttaaTTTTT
(N = sequência-guia; primeiro sublinhado = sequência tracr-mate; segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcaggggtgttttcggtattttaaTTTTT (N = sequência-guia; primeiro sublinhado = sequência tracr-mate; segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico de exemplo para Cas9 CRISPR1 LMD-9 de *S. thermophilus* (com PAM de NNAGAAW)

NNNNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagctacaatgataaggcttcatgccgaaatcaacaccctgtcattttatggcaggggtgt**TTTTTT**
(N = sequência-guia; primeiro sublinhado = sequência tracr-mate; segundo sublinhado = sequência tracr; negrito = terminador)

ARN quimérico de exemplo para Cas9 CRISPR3 LMD-9 de *S. thermophilus* (com PAM de NGGNG)

NNNNNNNNNNNNNNNNNNNNNNgtttttagagctgtgGAAAcacagcgagttaaaataaggcttagtccgtactcaactt gaaaaggt-ggcaccgattcggtgt**TTTTTT** (N = sequência-guia; primeiro sublinhado = sequência tracr-mate; segundo sublinhado = sequência tracr; negrito = terminador)

Versão com codões otimizados de Cas9 de locus de CRISPR3 LMD-9 de *S. thermophilus* (com uma NLS nas extremidades 5' e 3')

ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGACCA
AGCCCTACAGCATCGGCCTGGACATCGGCACCAATAGCGTGGGCTGGGCCGTGACC
ACCGACAACCTACAAGGTGCCAGCAAGAAAATGAAGGTGCTGGGCAACACCTCCAA
GAAGTACATCAAGAAAAACCTGCTGGGCGTGCTGCTGTTCGACAGCGGCATTACAG
CCGAGGGCAGACGGCTGAAGAGAACCGCCAGACGGCGGTACACCCGGCGGAGAAA
CAGAATCCTGTATCTGCAAGAGATCTTCAGCACCGAGATGGCTACCCTGGACGACG
CCTTCTTCAGCGGCTGGACGACAGCTTCCTGGTGCCCGACGACAAGCGGGACAGC
AAGTACCCCATCTTCGGCAACCTGGTGGAAGAGAAGGCCTACCACGACGAGTTCCC
CACCATCTACCACCTGAGAAAGTACCTGGCCGACAGCACCAAGAAGGCCGACCTGA
GACTGGTGTATCTGGCCCTGGCCACATGATCAAGTACCGGGGCCACTTCCTGATCG
AGGGCGAGTTCAACAGCAAGAACAACGACATCCAGAAGAACTTCAGGACTTCCTG

GACACCTACAACGCCATCTTCGAGAGCGACCTGTCCCTGGAAAACAGCAAGCAGCT
GGAAGAGATCGTGAAGGACAAGATCAGCAAGCTGGAAAAGAAGGACCGCATCCTG
AAGCTGTTCCCCGGCGAGAAGAACAGCGGAATCTTCAGCGAGTTTCTGAAGCTGAT
CGTGGGCAACCAGGCCGACTTCAGAAAAGTGCTTCAACCTGGACGAGAAAAGCCAGCC
TGCACCTCAGCAAAGAGAGCTACGACGAGGACCTGGAAACCCTGCTGGGATATATC
GGCGACGACTACAGCGACGTGTTCTGAAGGCCAAGAAGCTGTACGACGCTATCCT
GCTGAGCGGCTTCCTGACCGTGACCGACAACGAGACAGAGGCCCCACTGAGCAGCG
CCATGATTAAGCGGTACAACGAGCACAAAGAGGATCTGGCTCTGCTGAAAGAGTAC
ATCCGGAACATCAGCCTGAAAACCTACAATGAGGTGTTCAAGGACGACACCAAGAA
CGGCTACGCCGGCTACATCGACGGCAAGACCAACCAGGAAGATTTCTATGTGTACC
TGAAGAAGCTGCTGGCCGAGTTCGAGGGGGCCGACTACTTTCTGGAAAAAATCGAC
CGCGAGGATTTCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCTACCA
GATCCATCTGCAGGAAATGCGGGCCATCCTGGACAAGCAGGCCAAGTTCTACCCAT
TCCTGGCCAAGAACAAGAGCGGATCGAGAAGATCCTGACCTTCCGCATCCCTTACT
ACGTGGGCCCCCTGGCCAGAGGCCAACAGCGATTTTGCCTGOTCCATCCGGAAGCGC
AATGAGAAGATCACCCCTGGAACCTTCGAGGACGTGATCGACAAAGAGTCCAGCGC
CGAGGCCCTTCATCAACCGGATGACCAGCTTCGACCTGTACCTGCCCGAGGAAAAGG
TGCTGCCCAAGCACAGCCTGCTGTACGAGACATTCAATGTGTATAACGAGCTGACCA
AAGTGCGGTTTATCGCCGAGTCTATGCGGGACTACCAGTTCTCTGGACTCCAAGCAGA
AAAAGGACATCGTGCGGCTGTACTTCAAGGACAAGCGGAAAGTGACCGATAAGGAC
ATCATCGAGTACCTGCACGCCATCTACGGCTACGATGGCATCGAGCTGAAGGGCAT
CGAGAAGCAGTTCAACTCCAGCCTGAGCACATAACCACGACCTGCTGAACATTATCA
ACGACAAAGAATTTCTGGACGACTCCAGCAACGAGGCCATCATCGAAGAGATCATC
CACACCCTGACCATCTTTGAGGACCGCGAGATGATCAAGCAGCGGCTGAGCAAGTT
CGAGAACATCTTCGACAAGAGCGTGCTGAAAAAGCTGAGCAGACGGCACTACACCG
GCTGGGGCAAGCTGAGCGCCAAGCTGATCAACGGCATCCGGGACGAGAAGTCCGGC
AACACAATCCTGGACTACCTGATCGACGACGGCATCAGCAACCGGAACCTCATGCA
GCTGATCCACGACGACGCCCTGAGCTTCAAGAAGAAGATCCAGAAGGCCCAGATCA
TCGGGGACGAGGACAAGGGCAACATCAAAGAAGTCGTGAAGTCCCTGCCCGGCAGC
CCCCCATCAAGAAGGGAATCCTGCAGAGCATCAAGATCGTGGACGAGCTCGTGAA
AGTGATGGGCGGCAGAAAGCCCGAGAGCATCGTGGTGGAATGGCTAGAGAGAAC

CAGTACACCAATCAGGGCAAGAGCAACAGCCAGCAGAGACTGAAGAGACTGGAAA
AGTCCCTGAAAGAGCTGGGCAGCAAGATTCTGAAAGAGAATATCCCTGCCAAGCTG
TCCAAGATCGACAACAACGCCCTGCAGAACGACCGGCTGTACCTGTACTACCTGCA
GAATGGCAAGGACATGTATACAGGCGACGAOCTGGATATCGACCGCCTGAGCAACT
ACGACATCGACCATATTATCCCCCAGGCCTTCCTGAAAGACAACAGCATTGACAAC
AAAGTGCTGGTGTCTCCGCCAGCAACCGCGGCAAGTCCGATGATGTGCCCAGCCT
GGAAGTCGTGAAAAAGAGAAAGACCTTCTGGTATCAGCTGCTGAAAAGCAAGCTGA
TTAGCCAGAGGAAGTTCGACAACCTGACCAAGGCCGAGAGAGGCGGCCTGAGCCCT
GAAGATAAGGCCGGCTTCATCCAGAGACAGCTGGTGGAAACCCGGCAGATCACCAA
GCACGTGGCCAGACTGCTGGATGAGAAGTTTAACAACAAGAAGGACGAGAACAACC
GGGCCGTGCGGACCGTGAAAGATCATCACCCCTGAAGTCCACCCTGGTGTCCCAGTTCC
GGAAGGACTTCGAGCTGTATAAAGTGCGCGAGATCAATGACTTTCACCACGCCAC
GACGCCTACCTGAATGCCGTGGTGGCTTCCGCCCTGCTGAAGAAGTACCCTAAGCTG
GAACCCGAGTTTCGTGTACGGCGACTACCCCAAGTACAACCTCCTTCAGAGAGCGGAA
GTCCGCCACCGAGAAGGTGTACTTCTACTCCAACATCATGAATATCTTTAAGAAGTC
CATCTCCCTGGCCGATGGCAGAGTGATCGAGCGGCCCTGATCGAAGTGAACGAAG
AGACAGGCGAGAGCGTGTGGAACAAAGAAAGCGACCTGGCCACCGTGCGGCGGGT
GCTGAGTTATCCTCAAGTGAATGTCTGGAAGAAGGTGGAAGAACAGAACCACGGCC
TGGATCGGGGCAAGCCCAAGGGCCTGTTCAACGCCAACCTGTCCAGCAAGCCTAAG
CCCAACTCCAACGAGAATCTCGTGGGGGCCAAAGAGTACCTGGACCCTAAGAAGTA
CGGCGGATACGCCGGCATCTCCAATAGCTTCACCGTGCTCGTGAAGGGCACAATCG
AGAAGGGCGCTAAGAAAAAGATCACAAACGTGCTGGAATTTACAGGGGATCTCTATC
CTGGACCGGATCAACTACCGGAAGGATAAGCTGAACTTTCTGCTGGAAAAAGGCTA
CAAGGACATTGAGCTGATTATCGAGCTGCCTAAGTACTCCCTGTTGAACTGAGCGA
CGGCTCCAGACGGATGCTGGCCTCCATCCTGTCCACCAACAACAAGCGGGGCGAGA
TCCACAAGGGAAACCAGATCTTCCTGAGCCAGAAATTTGTGAAACTGCTGTACCACG
CCAAGCGGATCTCCAACACCATCAATGAGAACCACCGGAAATACGTGGAAAACAC
AAGAAAGAGTTTGAGGAACTGTTCTACTACATCCTGGAGTTCAACGAGAACTATGTG
GGAGCCAAGAAGAACGGCAAACTGCTGAACTCCGCCTTCCAGAGCTGGCAGAACCA
CAGCATCGACGAGCTGTGCAGCTCCTTCATCGGCCCTACCGGCAGCGAGCGGAAGG
GACTGTTTGAGCTGACCTCCAGAGGCTCTGCCGCCGACTTTGAGTTCCTGGGAGTGA

AGATCCCCCGGTACAGAGACTACACCCCTCTAGTCTGCTGAAGGACGCCACCCTGA
TCCACCAGAGCGTGACCGGCCTGTACGAAACCCGGATCGACCTGGCTAAGCTGGGC
GAGGGAAAGCGTCCTGCTGCTACTAAGAAAGCTGGTCAAGCTAAGAAAAAGAAATA

A

Exemplo 5: Otimização do ARN-guia para Cas9 de Streptococcus pyogenes (referida como SpCas9).

Os Requerentes mudaram o ARNtracr e as sequências de repetição direta, ou mudaram o ARN-guia quimérico para melhorar os ARNs nas células.

A otimização baseia-se na observação de que existem alongamentos de timinas (Ts) no ARNtracr e ARN-guia, o que pode originar a terminação de transcrição precoce através do promotor pol 3. Por conseguinte, os Requerentes geraram as seguintes sequências otimizadas. O ARN otimizado e a repetição direta otimizada correspondente são apresentados aos pares.

ARNtracr 1 otimizado (mutação sublinhada):

**GGAACCATTCaAACAGCATAGCAAGTTAtAATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTT**

Repetição direta otimizada 1 (mutação sublinhada):

GTTaTAGAGCTATGCTGTTaTGAATGGTCCCAAAAC

ARNtracr 2 otimizado (mutação sublinhada):

**GGAACCATTCAAiACAGCATAGCAAGTTAAiATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTT**

Repetição direta otimizada 2 (mutação sublinhada):

GTaTTAGAGCTATGCTGTaTTGAATGGTCCCAAAAC

Os Requerentes otimizaram igualmente o ARN-guia quimérico para uma atividade ideal nas células eucarióticas.

ARN-guia original:

NNNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

Sequência 1 de ARN-guia quimérico otimizado:

NNNNNNNNNNNNNNNNNNNNNNNGTATTAGAGCTAGAAATAGCAAGTTAAATAAGGC
TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

Sequência 2 de ARN-guia quimérico otimizado:

NNNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTGGAAACAAAACAGCA
TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG
GTGCTTTTTTT

Sequência 3 de ARN-guia quimérico otimizado:

NNNNNNNNNNNNNNNNNNNNNNNGTATTAGAGCTATGCTGTATTGGAAACAATACAGC
ATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
GGTGCTTTTTTT

Os Requerentes mostraram que o ARN-guia quimérico otimizado funciona melhor conforme indicado na Figura 9. A experiência foi realizada através da cotransfecção de células 293FT com Cas9 e uma cassete de ADN de ARN-guia U6 para expressar uma das quatro formas de ARN ilustradas acima. O alvo do ARN-guia é o mesmo local-alvo no *locus* de Emx1 humano: "GTCACCTC-CAATGACTAGGG"

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

Os Requerentes conceberam ARNs-guia quiméricos conforme ilustrado na Figura 12.

Os ARNs-guia de St1Cas9 podem ser submetidos ao mesmo tipo de otimização dos ARNs-guia de Spas9, quebrando os alongamentos das politiminas (Ts).

Exemplo 7: Melhoramento do sistema Cas9 para aplicação in

vivo

Os Requerentes realizaram uma pesquisa metagenômica para uma Cas9 com baixo peso molecular. A maioria dos homólogos de Cas9 é razoavelmente grande. Por exemplo, a SpCas9 tem cerca de 1368aa de comprimento, o que é demasiado grande para ser facilmente empacotada nos vetores virais para distribuição. Algumas das sequências podem ter sido mal anotadas e, por conseguinte, a frequência exata para cada comprimento pode não ser necessariamente precisa. Contudo, é fornecida uma breve visão da distribuição de proteínas Cas9 e sugerida a existência de homólogos de Cas9 mais pequenos.

Através de análise computacional, os Requerentes verificaram que na estirpe bacteriana *Campylobacter* existem duas proteínas Cas9 com menos de 1000 aminoácidos. A sequência para uma Cas9 de *Campylobacter jejuni* é apresentada abaixo. Com este comprimento, a CjCas9 pode ser facilmente empacotada em AAV, lentivírus, Adenovírus e outros vetores virais para a distribuição vigorosa nas células primárias e *in vivo* nos modelos de animal.

>Cas9 de *Campylobacter jejuni* (CjCas9)

```
MARILAFDIGISSIGWAFSENDELKDCGVRIPTKVENPKTGESLALPRRLARSARKRLARR
KARLNHILKHLIANEFKLNIEDYQSFDSESLAKAYKGSLSIPYELRFRALNELLSKQDFAR
VILHIARRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSK
EFTNVRNKKESYERCIASFLKDELKLIFKKQREFGFSFSKKFEEVLSVAFYKRALKDFS
HLVGNCSEFTDEKRAPKNSPLAFMFVALTRIINLNLKNTGILYTKDDLNLALLNEVLK
NGTLTYKQTKLLGLSDDYEFKGEKGTIFYEFKKYKEFIKALGEHNSQDDLNEIAKDIT
LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPMLLEGKKYDEACNE
LNLKVAINEDKKDFLPAFNETYKDEVTPVVLRAIKEYRKVLNALLKKYGVKHINIE
LAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGKINSKNILKLRLFKEQKEFCA
YSGEKIKISDLQDEKMLEIDHIYPYSRSDSYMNVLVFTKQNOEKLNTPFEEAFGND
AKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDRYIARLVNLNTKDY
DFLPLSDDENTKLNLTQKGSKVHVEAKSGMLTSALRHTWGFSKDRNNHLLHDAVI
IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLKIDEIF
VSKPERKKPSGALHEETFRKEEEFYQSYGKGEVLKALELGKIRKVNKIVKNGDMFR
VDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYK
DSLILIQTKDMQEPEFVYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKS
IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK.
```

O elemento ARNtracr putativo para esta CjCas9 é:

**TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTG
CGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT**

A sequência de Repetição Direta é:

ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

A estrutura de codobrimento do ARNtracr e repetição direta é fornecida na Figura 6.

Um exemplo de um ARN-guia quimérico para CjCas9 é:

**NNNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUU
UGCGGGACUCUGCGGGGUACAAUCCCUAAAACCGCUUUU**

Os Requerentes otimizaram igualmente o ARN-guia de Cas9 utilizando métodos *in vitro*. A Figura 18 ilustra dados da otimização de ARN-guia quimérico de StlCas9 *in vitro*.

Embora as formas de realização preferidas da presente invenção tenham sido aqui ilustradas e descritas, será óbvio para os peritos na técnica que essas formas de realização são fornecidas apenas como exemplo. Diversas variações, alterações e substituições irão ocorrer aos peritos na técnica sem sair da invenção. Convém compreender que é possível utilizar na aplicação da invenção várias alternativas às formas de realização da invenção aqui descritas. É pretendido que as seguintes reivindicações definam o âmbito da invenção e que os métodos e as estruturas no âmbito destas reivindicações e respectivos equivalentes sejam assim abrangidos.

Exemplo 8: Otimização de ARN_{gu} Sa

Os Requerentes conceberam cinco variantes de ARN_{gu} para SaCas9 para uma arquitetura truncada ideal com uma maior eficiência de clivagem. Além disso, o sistema duplo de repetição direta nativa:tracr foi testado juntamente com

ARNsgu. As guias com comprimentos indicados foram cotransfetadas com SaCas9 e testadas em células HEK 293FT relativamente à atividade. Um total de 100 ng de *amplicons* U6-PCR de ARNsgu (ou 50 ng de repetições diretas e 50 ng de ARNtracr) e 400 ng de plasmídeo de SaCas9 foi cotransfetado em 200.000 hepatócitos de ratinho Hepal-6, e o ADN foi colhido 72 horas após a transfeção para análise SURVEYOR. Os resultados são ilustrados na Fig. 23.

Referências:

1. Umov, 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 a 646 (2010).
2. Bogdanove, A.J. & Voytas, D.F. "TAL effectors: customizable proteins for DNA targeting." *Science* 333, 1843 a 1846 (2011).
3. Stoddard, B.L. "Homing endonuclease structure and function." *Q. Rev. Biophys.* 38, 49 a 95 (2005).
4. Bae, T. & Schneewind, O. "Allelic replacement in *Staphylococcus aureus* with inducible counter-selection." *Plasmid* 55, 58 a 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 a 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 a 223 (2009).
7. Jinek, M. et al. "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." *Science* 337, 816 a 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 a 493 (2010).
9. Horvath, P. & Banangou, R. "CRISPR-Cas, the immune

system of bacteria and archaea." *Science* 327, 167 a 170 (2010).

10. Terns, M.P. & Terns, R.M. "CRISPR-based adaptive immune systems." *Curr. Opin. Microbiol.* 14, 321 a 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 a 407 (2009).

12. Brouns, S.J. et al. "Small CRISPR RNAs guide antiviral defense in prokaryotes." *Science* 321, 60 a 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 a 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. E.U.A.* 108, 21.218 a 21.222 (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 a 1358 (2010).

17. Deveau, H. et al. "Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*." *J. Bacteriol.* 190, 1390 a 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. E.U.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 a 838 (2012).

21. Brouns, S.J. "Molecular biology. A Swiss army knife of immunity." *Science* 337, 808 a 809 (2012).

22. Carroll, D. A CRISPR "Approach to Gene Targeting." *Mol. Ther.* 20, 1658 a 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 a 186 (2012).

24. Saprunauskas, 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. E.U.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. E.U.A.* (2011).

27. Zahner, D. & Hakenbeck, R. "The *Streptococcus pneumoniae* beta-galactosidase is a surface protein." *J. Bacteriol.* 182, 5919 a 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 a 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 S12 enhances protein synthesis activity during the late growth phase in *Escherichia coli*." *Mol. Genet. Genomics* 271, 317 a 324 (2004).
31. Costantino, N. & Court, D.L. "Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants." *Proc. Natl. Acad. Sci. E.U.A.* 100, 15.748 a 15.753 (2003).
32. Edgar, R. & Qimron, U. "The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction." *J. Bacteriol.* 192, 6291 a 6294 (2010).
33. Marraffini, L.A. & Sontheimer, E.J. "Self versus non-self discrimination during CRISPR RNA-directed immunity." *Nature* 463, 568 a 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, 33.351 a 33.363 (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 a 49 (2011).
36. Wang, H.H. et al. "Genome-scale promoter engineering by coselection MAGE." *Nat Methods* 9, 591 a 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 a 5717 (2001).

40. Havarstein, L.S., Coomaraswamy, G. & Morrison, D.A. "An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*." *Proc. Natl. Acad. Sci. E.U.A.* 92, 11.140 a 11.144 (1995).
41. Horinouchi, S. & Weisblum, B. "Nucleotide sequence and functional map of pCl94, a plasmid that specifies inducible chloramphenicol resistance." *J. Bacteriol.* 150, 815 a 825 (1982).
42. Horton, R.M. "In Vitro Recombination and Mutagenesis of DNA : SOEing Together Tailor-Made Genes." *Methods Mol. Biol.* 15, 251 a 261 (1993).
43. Podbielski, A., Spellerberg, B., Woischnik, M., Pohl, B. & Lutticken, R. "Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS)." *Gene* 177, 137 a 147 (1996).
44. Husmann, L.K., Scott, J.R., Lindahl, G. & Stenberg, L. "Expression of the Alp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*." *Infection und immunity* 63, 345 a 348 (1995).
45. Gibson, D.G. et al. "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nat Methods* 6, 343 a 345 (2009).
46. Tangri S, et al. "Rationally engineered therapeutic proteins with reduced immunogenicity" *J. Immunol.* 15 de março de 2005;174(6):3187 a 96.

Embora as formas de realização preferidas da presente invenção tenham sido aqui ilustradas e descritas, será óbvio para os peritos na técnica que essas formas de realização são fornecidas apenas como exemplo. Diversas variações, alterações e substituições irão ocorrer aos peritos na técnica sem sair da invenção. Convém compreender que é possível utilizar na aplicação da invenção várias alternativas às formas de realização da invenção aqui

EP2771468B1

descritas.

DOCUMENTOS REFERIDOS NA DESCRIÇÃO

Esta lista de documentos referidos pelo autor do presente pedido de patente foi elaborada apenas para informação do leitor. Não é parte integrante do documento de patente europeia. Não obstante o cuidado na sua elaboração, o IEP não assume qualquer responsabilidade por eventuais erros ou omissões.

Documentos de patente referidos na descrição

- US 4873316 A [0057]
- EP 264166 A [0057]
- US 20110059502 A [0072]
- US 5449386 A [0074]
- US 4946787 A [0074] [0075]
- US 4897355 A [0074]
- WO 9117424 A [0074]
- WO 9116024 A [0074]
- US 4186183 A [0075]
- US 4217344 A [0075]
- US 4235871 A [0075]
- US 4261975 A [0075]
- US 4485054 A [0075]
- US 4501728 A [0075]
- US 4774085 A [0075]
- US 4837028 A [0075]
- US 9405700 W [0077]
- US 4797368 A [0078]
- WO 9324641 A [0078]
- US 5173414 A [0078]
- US 20030087817 A [0079]
- US 6603061 B [0085]
- US 7868149 B [0085]
- US 20090100536 A [0085]

Literatura não relacionada com patentes, citada na descrição

- **GOEDDEL.** GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY. Academic Press, 1990, vol. 185 [0007] [0051]
- **BOSHART et al.** *Cell*, 1985, vol. 41, 521-530 [0007]
- *Mol. Cell. Biol.*, 1988, vol. 8 (1), 466-472 [0007]
- *Proc. Natl. Acad. Sci. USA.*, 1981, vol. 78 (3), 1527-31 [0007]
- **SAMBROOK ; FRITSCH ; MANIATIS.** MOLECULAR CLONING: A LABORATORY MANUAL. 1989 [0050]
- CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. 1987 [0050]
- METHODS IN ENZYMOLOGY. Academic Press, Inc, [0050]
- PCR 2: A PRACTICAL APPROACH. 1995 [0050]
- ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE. 1987 [0050]
- **SMITH; JOHNSON.** *Gene*, 1988, vol. 67, 31-40 [0052]
- **AMRANN et al.** *Gene*, 1988, vol. 69, 301-315 [0053]
- **STUDIER et al.** GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY. Academic Press, 1990, vol. 185, 60-89 [0053]
- **BALDARI et al.** *EMBO J.*, 1987, vol. 6, 229-234 [0054]
- **KUIJAN ; HERSKOWITZ.** *Cell*, 1982, vol. 30, 933-943 [0054]
- **SCHULTZ et al.** *Gene*, 1987, vol. 54, 113-123 [0054]
- **SMITH et al.** *Mol. Cell. Biol.*, 1983, vol. 3, 2156-2165 [0055]
- **LUCKLOW ; SUMMERS.** *Virology*, 1989, vol. 170, 31-39 [0055]
- **SEED.** *Nature*, 1987, vol. 329, 840 [0056]
- **KAUFMAN et al.** *EMBO J.*, 1987, vol. 6, 187-195 [0056]
- **SAMBROOK et al.** MOLECULAR CLONING: A LABORATORY MANUAL. Cold Spring Harbor Laboratory Press, 1989 [0056]
- **PINKERT et al.** *Genes Dev.*, 1987, vol. 1, 268-277 [0057]

- **CALAME ; EATON.** *Adv. Immunol.*, 1988, vol. 43, 235-275 [0057]
- **WINOTO; BALTIMORE.** *EMBD J.*, 1988, vol. 8, 729-733 [0057]
- **BANEIJI et al.** *Cell*, 1983, vol. 33, 729-740 [0057]
- **QUEEN ; BALTIMORE.** *Cell*, vol. 33, 741-748 [0057]
- **BYRNE ; RUDDLE.** *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, 5473-5477 [0057]
- **EDLUND et al.** *Science*, 1985, vol. 230, 912-916 [0057]
- **KESSEL ; GRUSS.** *Science*, 1990, vol. 249, 374-379 [0057]
- **CAMPES ; TILGHMAN.** *Genes Dev.*, 1989, vol. 3, 537-546 [0057]
- **ISHINO et al.** *J. Bacteriol.*, 1987, vol. 169, 5429-5433 [0058]
- **NAKATA et al.** *J. Bacteriol.*, 1989, vol. 171, 3553-3556 [0058]
- **GROENEN et al.** *Mol. Microbiol.*, 1993, vol. 10, 1057-1065 [0058]
- **HOE et al.** *Emerg. Infect. Dis.*, 1999, vol. 5, 254-263 [0058]
- **MASEPOHL et al.** *Biochim. Biophys. Acta*, 1996, vol. 1307, 26-30 [0058]
- **MOJICA et al.** *Mol. Microbiol.*, 1995, vol. 17, 85-93 [0058]
- **JANSSEN et al.** *OMICS J. Integ. Biol.*, 2002, vol. 6, 23-33 [0058]
- **MOJICA et al.** *Mol. Microbiol.*, 2000, vol. 36, 244-246 [0058]
- **VAN EMBDEN et al.** *J. Bacteriol.*, 2000, vol. 182, 2393-2401 [0058]
- **JANSEN et al.** *Mol. Microbiol.*, 2002, vol. 43, 1565-1575 [0058]
- **NAKAMURA, Y. et al.** Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucl. Acids Res.*, 2000, vol. 28, 292 [0064]

- **ZUKER ; STIEGLER.** *Nucleic Acids Res.*, 1981, vol. 9, 133-148 [0069]
- **A.R. GRUBER et al.** *Cell*, 2008, vol. 106 (1), 23-24 [0069]
- **PA CARR ; GM CHURCH.** *Nature Biotechnology*, 2009, vol. 27 (12), 1151-62 [0069]
- **ANDERSON.** *Science*, 1992, vol. 256, 808-813 [0073]
- **NABEL ; FELGNER.** *TIBTECH*, 1993, vol. 11, 211-217 [0073]
- **MITANI ; CASKEY.** *TIBTECH*, 1993, vol. 11, 162-166 [0073]
- **DILLON.** *TIBTECH*, 1993, vol. 11, 167-175 [0073]
- **MILLER.** *Nature*, 1992, vol. 357, 455-460 [0073]
- **VAN BRUNT.** *Biotechnology*, 1988, vol. 6 (10), 1149-1154 [0073]
- **VIGNE.** *Restorative Neurology and Neuroscience*, 1995, vol. 8, 35-36 [0073]
- **KREMER ; PERRICAUDET.** *British Medical Bulletin*, 1995, vol. 51 (1), 31-44 [0073]
- **HADDADA et al.** *Current Topics in Microbiology and Immunology* [0073]
- **YU et al.** *Gene Therapy*, 1994, vol. 1, 13-26 [0073]
- **CRYSTAL.** *Science*, 1995, vol. 270, 404-410 [0075]
- **BLAESE et al.** *Cancer Gene Ther.*, 1995, vol. 2, 291-297 [0075]
- **BEHR et al.** *Bioconjugate Chem.*, 1994, vol. 5, 382-389 [0075]
- **REMY et al.** *Bioconjugate Chem.*, 1994, vol. 5, 647-654 [0075]
- **GAO et al.** *Gene Therapy*, 1995, vol. 2, 710-722 [0075]
- **AHMAD et al.** *Cancer Res.*, 1992, vol. 52, 4817-4820 [0075]
- **BUCHSCHER et al.** *J. Virol.*, 1992, vol. 66, 2731-2739 [0077]
- **JOHANN et al.** *J. Virol.*, 1992, vol. 66, 1635-1640 [0077]
- **SOMMNERFELT et al.** *Virol.*, 1990, vol. 176, 58-59 [0077]

- **WILSON et al.** *J. Virol.*, 1989, vol. 63, 2374-2378 [0077]
- **MILLER et al.** *J. Virol.*, 1991, vol. 65, 2220-2224 [0077]
- **WEST et al.** *Virology*, 1987, vol. 160, 38-47 [0078]
- **KOTIN.** *Human Gene Therapy*, 1994, vol. 5, 793-801 [0078]
- **MUZYCZKA.** *J. Clin. Invest.*, 1994, vol. 94, 1351 [0078]
- **TRATSCHIN et al.** *Mol. Cell. Biol.*, 1985, vol. 5, 3251-3260 [0078]
- **TRATSCHIN et al.** *Mol. Cell. Biol.*, 1984, vol. 4, 2072-2081 [0078]
- **HERMONAT ; MUZYCZKA.** *PNAS*, 1984, vol. 81, 6466-6470 [0078]
- **SAMULSKI et al.** *J. Virol.*, 1989, vol. 63, 03822-3828 [0078]
- **MORRELL et al.** Crop genomics: advances and applications. *Nat Rev Genet.*, 29 December 2011, vol. 13 (2), 85-96 [0085]
- **ROBERT D. WELLS ; TETSUO ASHIZAWA.** Genetic Instabilities and Neurological Diseases. Academic Press, 13 October 2011 [0096]
- **MCIVOR EI ; POLAK U ; NAPIERALA M.** New insights into repeat instability: role of RNA•DNA hybrids. *RNA Biol.*, September 2010, vol. 7 (5), 551-8 [0096]
- Genetics of Epilepsy and Genetic Epilepsies. Mariani Foundation Paediatric Neurology, 2009, 20 [0097]
- Genetic Diseases of the Eye. Oxford University Press, 2012 [0098]
- **TANGRI S et al.** Rationally engineered therapeutic proteins with reduced immunogenicity. *J Immunol.*, 15 March 2005, vol. 174 (6), 3187-96 [0119] [0205]
- **SAPRANAUSAKS et al.** *Cucleic Acis Research*, 2011, vol. 39, 9275 [0137]
- **GASIUNAS et al.** *Proc. Natl. Acad. Sci. USA*, 2012, vol. 109, 2579 [0137]

- **JINEK et al.** *Science*, 2012, vol. 337, 816 [0139]
- **MAKAROVA et al.** *Nat Rev Microbiol*, 2011, vol. 9, 467 [0139]
- **UMOV, F.D.; REBAR, E.J. ; HOLMES, M.C. ; ZHANG, H.S. ; GREGORY, P.D.** Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, 2010, vol. 11, 636-646 [0205]
- **BOGDANOVE, A.J. ; VOYTAS, D.F.** TAL effectors: customizable proteins for DNA targeting. *Science*, 2011, vol. 333, 1843-1846 [0205]
- **STODDARD, B.L.** Homing endonuclease structure and function. *Q. Rev. Biophys*, 2005, vol. 38, 49-95 [0205]
- **BAE, T. ; SCHNEEWIND, O.** Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid*, 2006, vol. 55, 58-63 [0205]
- **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.*, 2001, vol. 67, 5190-5196 [0205]
- **SHARAN, S.K. ; THOMASON, L.C. ; KUZNETSOV, S.G. ; COURT, D.L.** Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.*, 2009, vol. 4, 206-223 [0205]
- **JINEK, M. et al.** A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, vol. 337, 816-821 [0205]
- **DEVEAU, H. ; GARNEAU, J.E. ; MOINEAU, S.** CRISPR-Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.*, 2010, vol. 64, 475-493 [0205]
- **HORVATH, P. ; BANANGO, R.** CRISPR-Cas, the immune system of bacteria and archaea. *Science*, 2010, vol. 327, 167-170 [0205]
- **TERNS, M.P. ; TERNS, R.M.** CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.*, 2011, vol. 14, 321-327 [0205]

- **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.*, 2009, vol. 34, 401-407 [0205]
- **BROUNS, S.J. et al.** Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008, vol. 321, 60-964 [0205]
- **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.*, 2008, vol. 22, 3489-3496 [0205]
- **DELTCHEVA, E. et al.** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011, vol. 471, 602-607 [0205]
- **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.*, 2011, vol. 108, 21218-21222 [0205]
- **HAURWITZ, R.E. ; JINEK, M. ; WIEDENHEFT, B. ; ZHOU, K. ; DOUDNA, J.A.** Sequence and structure-specific RNA processing by a CRISPR endonuclease. *Science*, 2010, vol. 329, 1355-1358 [0205]
- **DEVEAU, H. et al.** Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, 2008, vol. 190, 1390-1400 [0205]
- **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 [0205]
- **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.*, 2011, vol. 6, 38 [0205]

- **BARRANGOU, R.** RNA-mediated programable DNA cleavage. *Nat. Biotechnol.*, 2012, vol. 30, 836-838 [0205]
- **BROUNS, S.J.** Molecular biology. A Swiss army knife of immunity. *Science*, 2012, vol. 337, 808-809 [0205]
- **CARROLL, D.** A CRISPR Approach to Gene Targeting. *Mol. Ther.*, 2012, vol. 20, 1658-1660 [0205]
- **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*, 2012, vol. 12, 177-186 [0205]
- **SAPRANAUSKAS, R. et al.** The *Streptococcus thermophilus* CRISPR-Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.*, 2011 [0205]
- **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 [0205]
- **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 [0205]
- **ZAHNER, D. ; HAKENBECK, R.** The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J. Bacteriol.*, 2000, vol. 182, 5919-5921 [0205]
- **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.*, 2006, vol. 70, 192-221 [0205]
- **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.*, 1999, vol. 13, 2889-2903 [0205]
- **HOSAKA, T. et al.** The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during

the late growth phase in *Escherichia coli*. *Mol. Genet. Genomics*, 2004, vol. 271, 317-324 [0205]

- **COSTANTINO, N. ; COURT, D.L.** Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.*, 2003, vol. 100, 15748-15753 [0205]
- **EDGAR, R. ; QIMRON, U.** The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J. Bacteriol.*, 2010, vol. 192, 6291-6294 [0205]
- **MARRAFFINI, L.A. ; SONTHEIMER, E.J.** Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 2010, vol. 463, 568-571 [0205]
- **FISCHER, S. et al.** An archaeal immune system can detect multiple Protospacer Adjacent Motifs (PAMs) to target invader DNA. *J. Biol. Chem.*, 2012, vol. 287, 33351-33363 [0205]
- **GUDBERGSDOTTIR, 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.*, 2011, vol. 79, 35-49 [0205]
- **WANG, H.H. et al.** Genome-scale promoter engineering by coselection MAGE. *Nat Methods*, 2012, vol. 9, 591-593 [0205]
- **CONG, L. et al.** Multiplex Genome Engineering Using CRISPR-Cas Systems. *Science*, 2013 [0205]
- **MALI, P. et al.** RNA-Guided Human Genome Engineering via Cas9. *Science*, 2013 [0205]
- **HOSKINS, J. et al.** Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.*, 2001, vol. 183, 5709-5717 [0205]
- **HAVARSTEIN, L.S. ; COOMARASWAMY, G. ; MORRISON, D.A.** An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus*

pneumonia. *Proc. Natl. Acad. Sci. U.S.A.*, 1995, vol. 92, 11140-11144 [0205]

- **HORINOUCI, S. ; WEISBLUM, B.** Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.*, 1982, vol. 150, 815-825 [0205]
- **HORTON, R.M.** In Vitro Recombination and Mutagenesis of DNA : SOEing Together Tailor-Made Genes. *Methods Mol. Biol.*, 1993, vol. 15, 251-261 [0205]
- **PODBIELSKI, A. ; SPELLERBERG, B. ; WOISCHNIK, M. ; POHL, B. ; LUTTICKEN, R.** Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS). *Gene*, 1996, vol. 177, 137-147 [0205]
- **HUSMANN, L.K. ; SCOTT, J.R. ; LINDAHL, G. ; STENBERG, L.** Expression of the Alp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*. *Infection und immunity*, 1995, vol. 63, 345-348 [0205]
- **GIBSON, D.G. et al.** Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, vol. 6, 343-345 [0205]

Lisboa, 6 de Maio de 2015

REIVINDICAÇÕES

1. Uma composição produzida ou que ocorre de forma não natural compreendendo:

uma sequência de polinucleótidos de ARN quimérico (ARNqui) do sistema associado a CRISPR (Cas) (CRISPR - Repetições Palindrômicas Curtas Agrupadas e Regularmente Interpoladas) (CRISPR-Cas), em que a sequência de polinucleótidos compreende:

- (a) uma sequência-guia com um comprimento de 10 a 30 nucleótidos, capaz de hibridizar para uma sequência-alvo numa célula eucariótica,
- (b) uma sequência tracr-mate, uma
- (c) uma sequência de ARNtracr,

em que (a), (b) e (c) são dispostos numa orientação de 5' para 3',

em que, quando transcrita, a sequência tracr-mate é hibridizada para a sequência ARNtracr e a sequência-guia direciona a ligação específica da sequência de um complexo CRISPR para a sequência-alvo,

em que o complexo CRISPR compreende uma proteína Cas9 de Tipo II complexada com (1) a sequência-guia que é hibridizada para a sequência-alvo, e (2) a sequência tracr-mate que é hibridizada para a sequência ARNtracr, em que a sequência ARNtracr tem um comprimento de 50 ou mais nucleótidos.

2. Um sistema de vetores associado a CRISPR (Cas) (CRISPR - Repetições Palindrômicas Curtas Agrupadas e Regularmente Interpoladas) (CRISPR-Cas) compreendendo um ou mais vetores compreendendo:

I. um primeiro elemento regulador ligado de modo funcional a uma sequência de nucleótidos que codifica uma sequência de polinucleótidos de ARN quimérico (ARNqui) do sistema CRISPR-Cas conforme definido na reivindicação 1,

e

II. um segundo elemento regulador ligado de modo funcional a uma sequência de nucleótidos que codifica uma proteína Cas9 de Tipo II compreendendo uma ou mais sequências de localização nuclear, de força suficiente para ativar a acumulação da referida proteína Cas9 numa quantidade detetável no núcleo de uma célula eucariótica; em que os componentes I e II se encontram situados nos mesmos ou em diferentes vetores do sistema.

3. A composição de acordo com a reivindicação 1, em que a enzima CRISPR compreende uma ou mais sequências de localização nuclear de força suficiente para ativar a acumulação da referida enzima Cas9 numa quantidade detetável no núcleo de uma célula eucariótica.

4. A composição de acordo com a reivindicação 1, ou o sistema de vetores de acordo com a reivindicação 2, em que a proteína Cas9 é uma nuclease que direciona a clivagem de ambos os filamentos do *locus* do polinucleótido.

5. A composição de acordo com a reivindicação 1, ou o sistema de vetores de acordo com a reivindicação 2, em que a proteína Cas9 compreende uma ou mais mutações num domínio catalítico, e corresponde a uma *nickase* que cliva apenas um filamento do *locus* do polinucleótido.

6. O sistema de vetores de acordo com a reivindicação 2, ou qualquer outra reivindicação dependente da mesma, em que a sequência de nucleótidos que codifica a proteína Cas9 tem codões otimizados para a expressão numa célula eucariótica.

7. O sistema de vetores de acordo com a reivindicação 2, ou qualquer reivindicação dependente da mesma, em que os vetores correspondem a vetores virais.

8. O sistema de vetores de acordo com a reivindicação 7, em que os vetores virais correspondem a vetores retrovirais, lentivirais, adenovirais, adeno-associados ou virais de herpes simplex.

9. O sistema de vetores de acordo com a reivindicação 2, ou qualquer reivindicação dependente da mesma, em que o sistema de vetores compreende codificação de sequências de nucleótidos para dois ou mais sinais de localização nuclear (NLSs) expressados com a sequência de nucleótidos que codifica a proteína Cas9.

10. O sistema de vetores de acordo com a reivindicação 9, em que, quando expressado, pelo menos um NLS se encontra no, ou próximo do, amino-terminal da proteína Cas9 e/ou pelo menos um NLS se encontra no, ou próximo do, carboxi-terminal da proteína Cas9.

11. O sistema de vetores de acordo com a reivindicação 9, em que pelo menos um NLS se encontra no, ou próximo do, amino-terminal da proteína Cas9 e pelo menos um NLS se encontra no, ou próximo do, carboxi-terminal da proteína Cas9.

12. Utilização da composição de acordo com a reivindicação 1, ou do sistema de vetores de acordo com a reivindicação 2 ou qualquer reivindicação dependente das mesmas para a engenharia de genomas, desde que a referida utilização não seja um método para o tratamento do corpo humano ou animal mediante cirurgia ou terapia, e desde que a referida utilização não seja um processo para a modificação da identidade genética da linha germinal de seres humanos.

13. A utilização de acordo com a reivindicação 12, em que a engenharia de genomas compreende a modificação de um

polinucleótido-alvo numa célula eucariótica, a modificação da expressão de um polinucleótido numa célula eucariótica, a geração de uma célula eucariótica modelo que compreende um gene de doença mutante, ou a inativação (*knocking out*) de um gene.

14. A utilização de acordo com a reivindicação 12, em que a utilização compreende ainda a reparação do referido polinucleótido-alvo clivado inserindo um polinucleótido de modelo exógeno, em que a referida reparação resulta numa mutação compreendendo uma inserção, deleção ou substituição de um ou mais nucleótidos do referido polinucleótido-alvo.

15. A utilização de acordo com a reivindicação 12, em que a utilização compreende ainda a edição do referido polinucleótido-alvo clivado inserindo um polinucleótido de modelo exógeno, em que a referida edição resulta numa mutação compreendendo uma inserção, deleção ou substituição de um ou mais nucleótidos do referido polinucleótido-alvo.

16. A utilização de acordo com a reivindicação 13 ou 14, em que a inserção é mediante recombinação homóloga.

17. Utilização da composição de acordo com a reivindicação 1, ou do sistema de vetores de acordo com a reivindicação 2 ou qualquer reivindicação dependente das mesmas, na produção de um animal transgénico não humano ou uma planta transgénica.

Lisboa, 6 de Maio de 2015

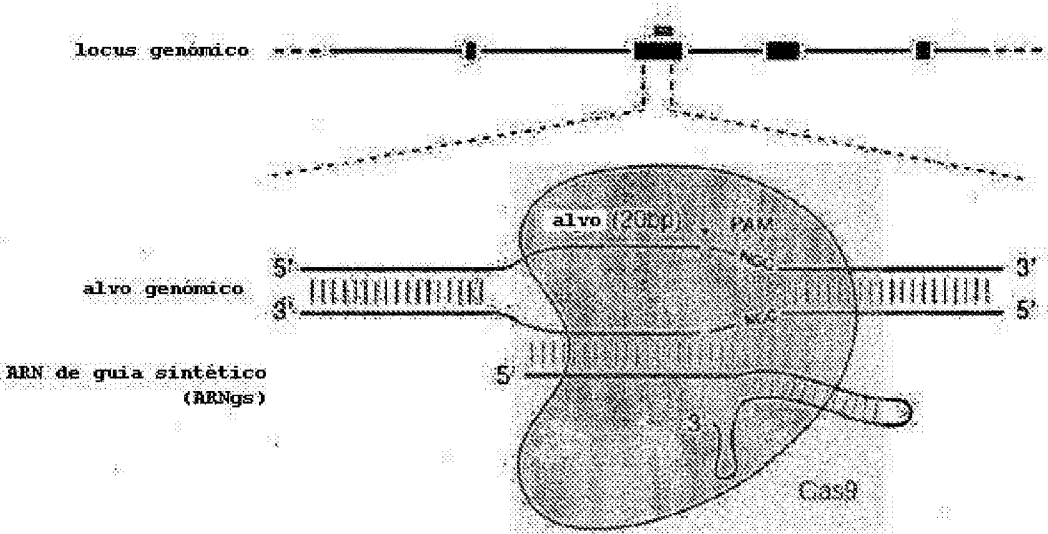


FIG. 1

A

locus 1 de crisper *Streptococcus pyogenes* SF370

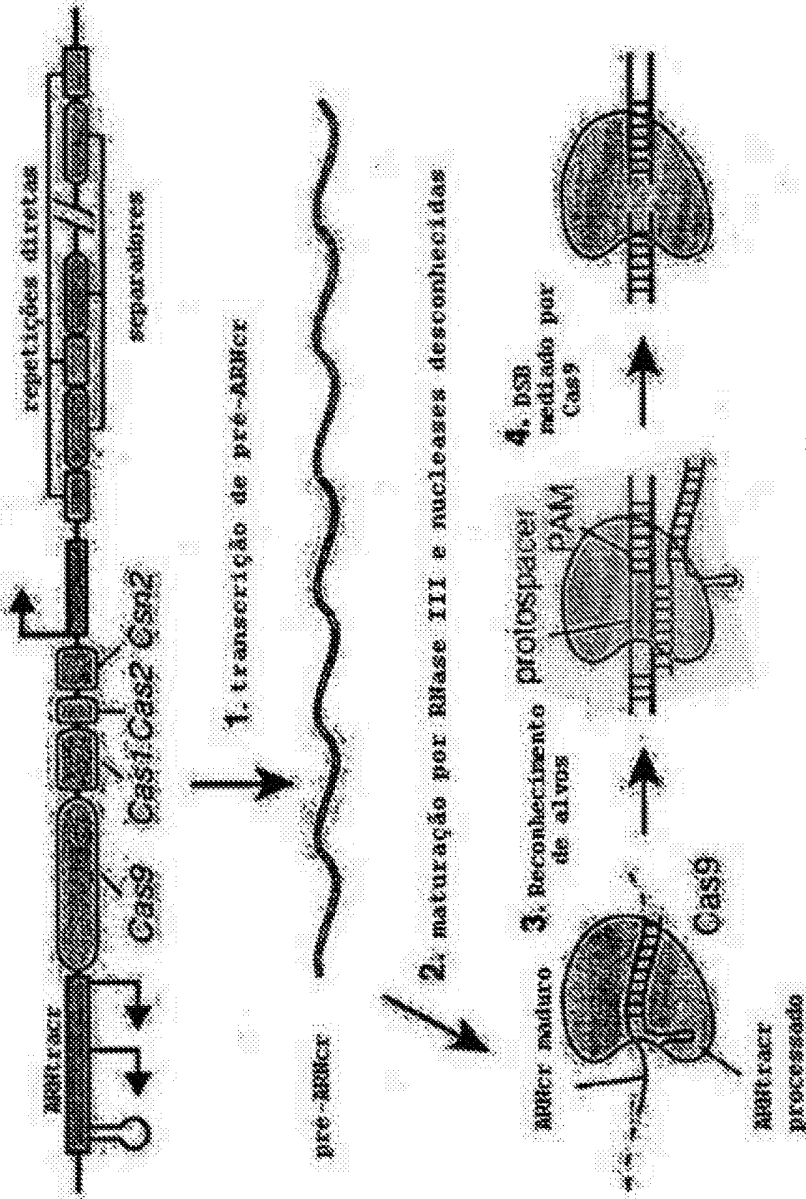


FIG. 2A

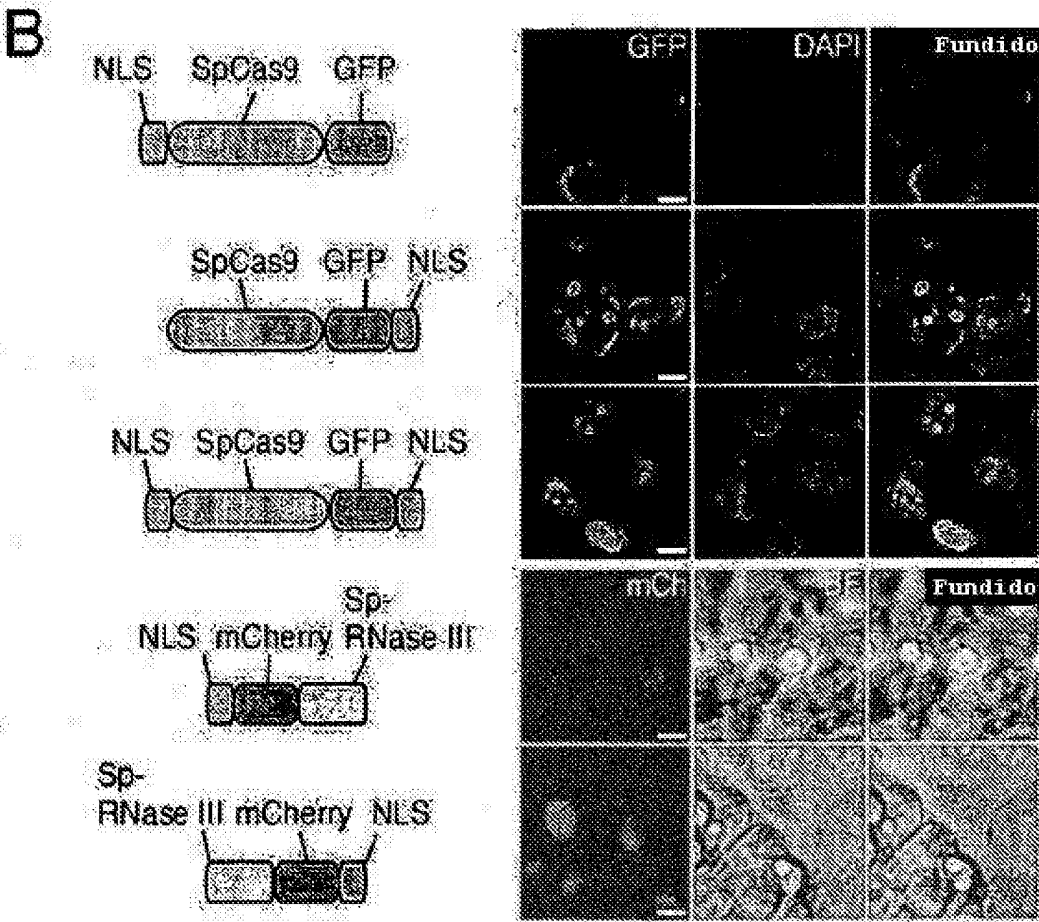


FIG. 2B

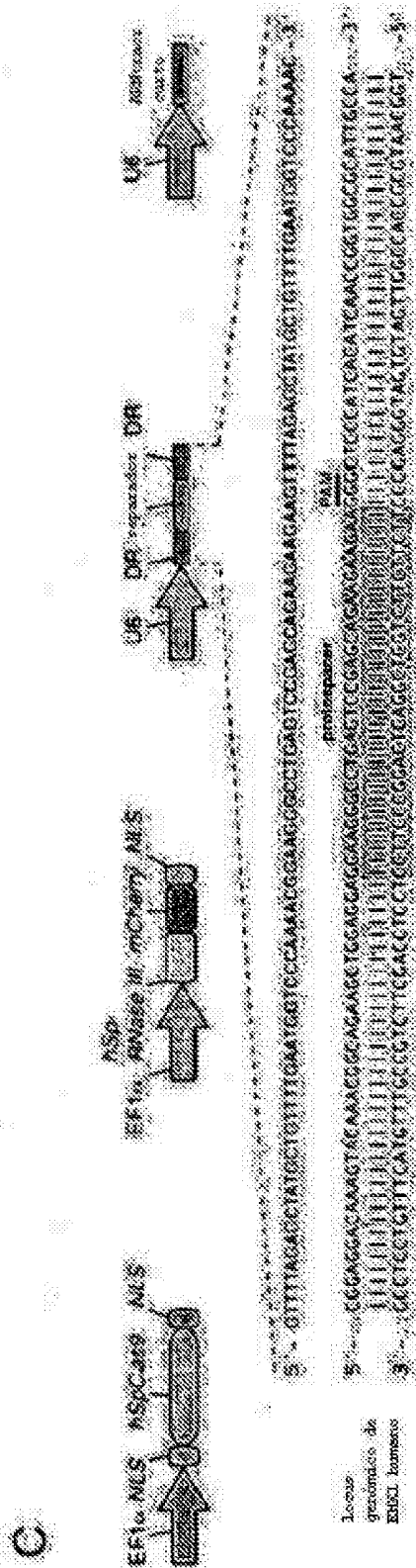
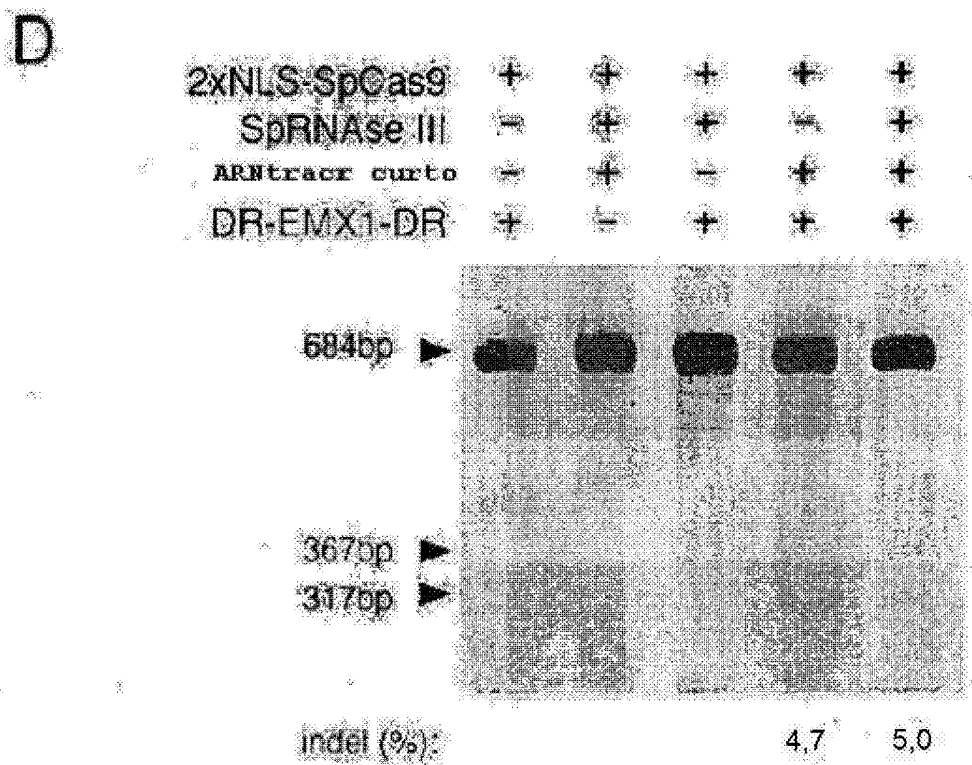
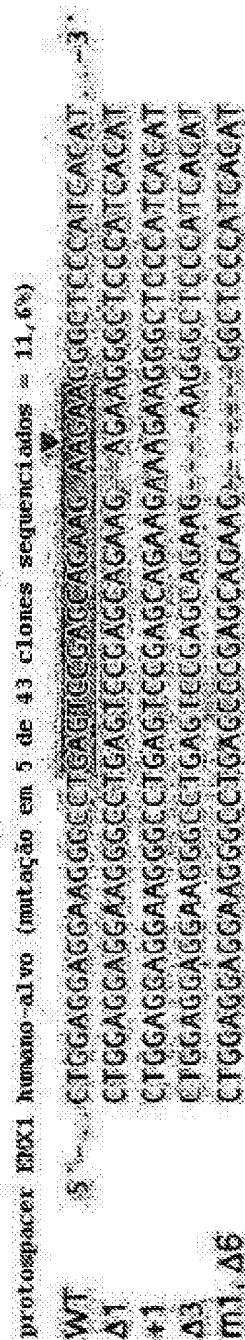
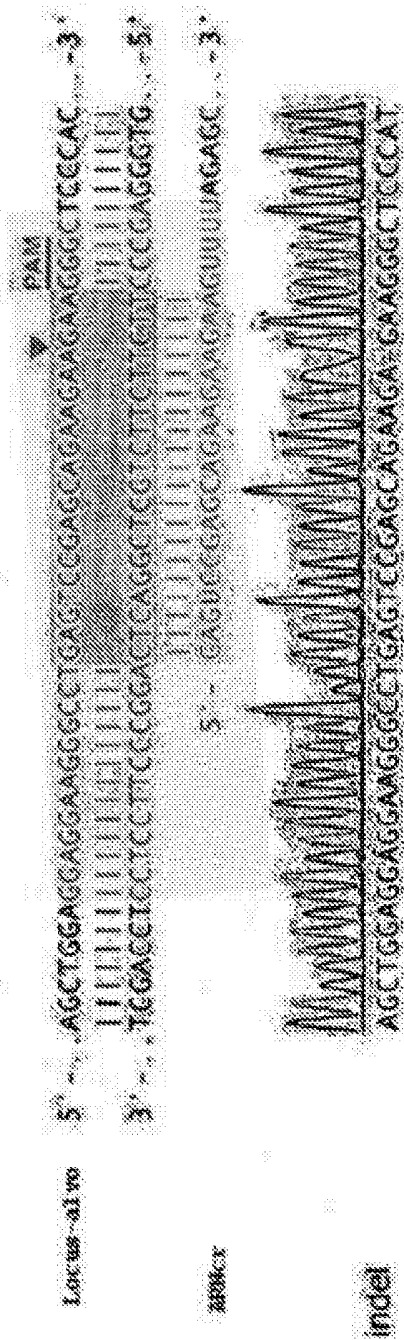


FIG 2C





HC-21-F



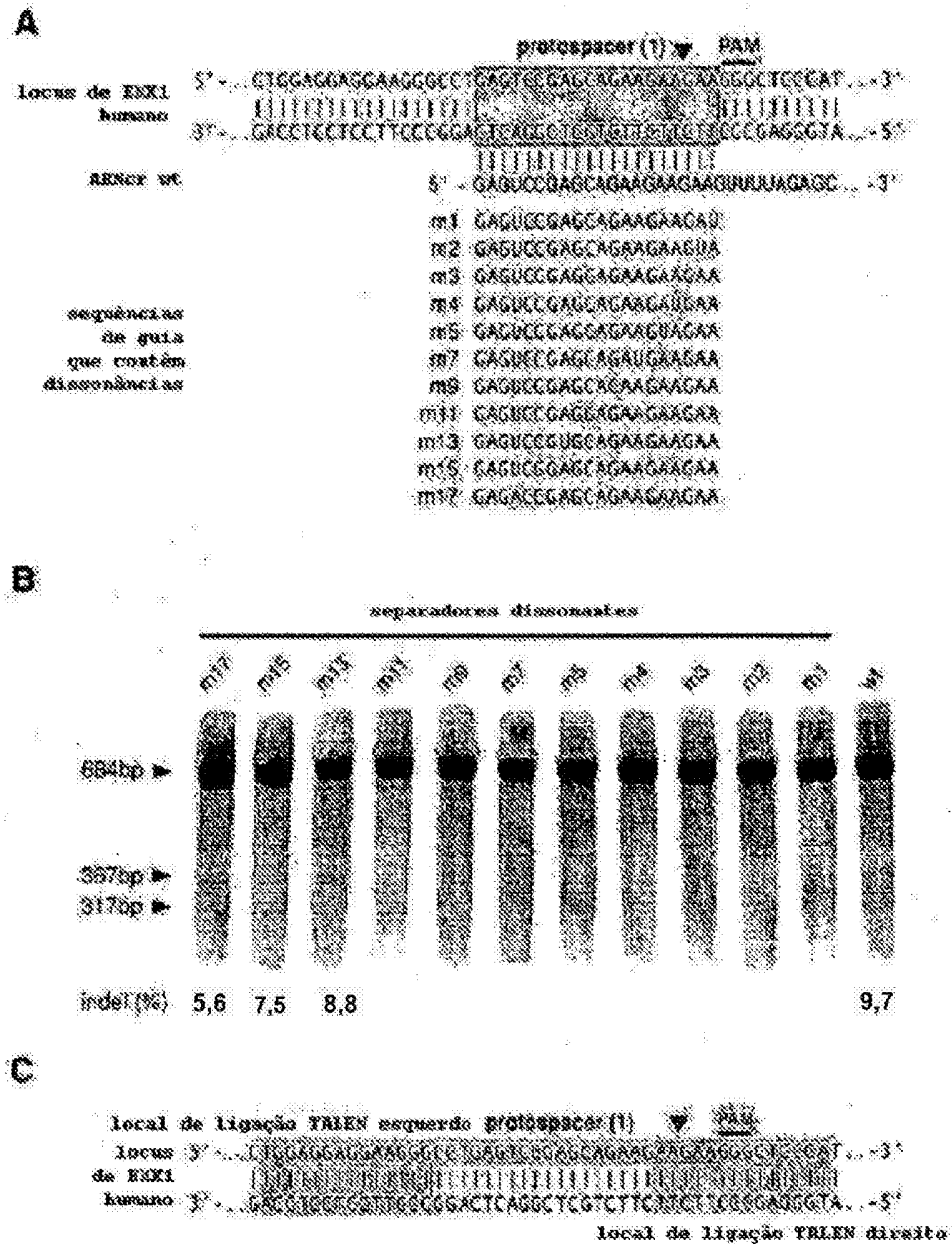


FIG. 4A-C

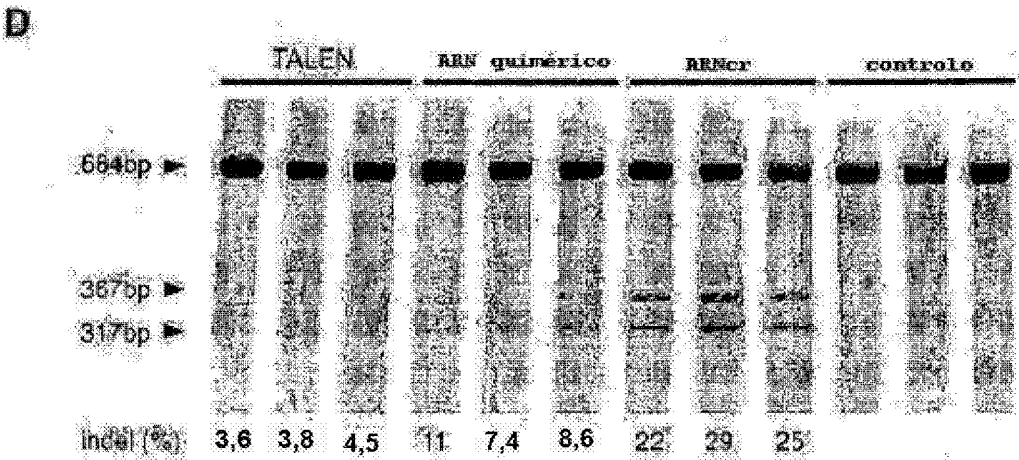
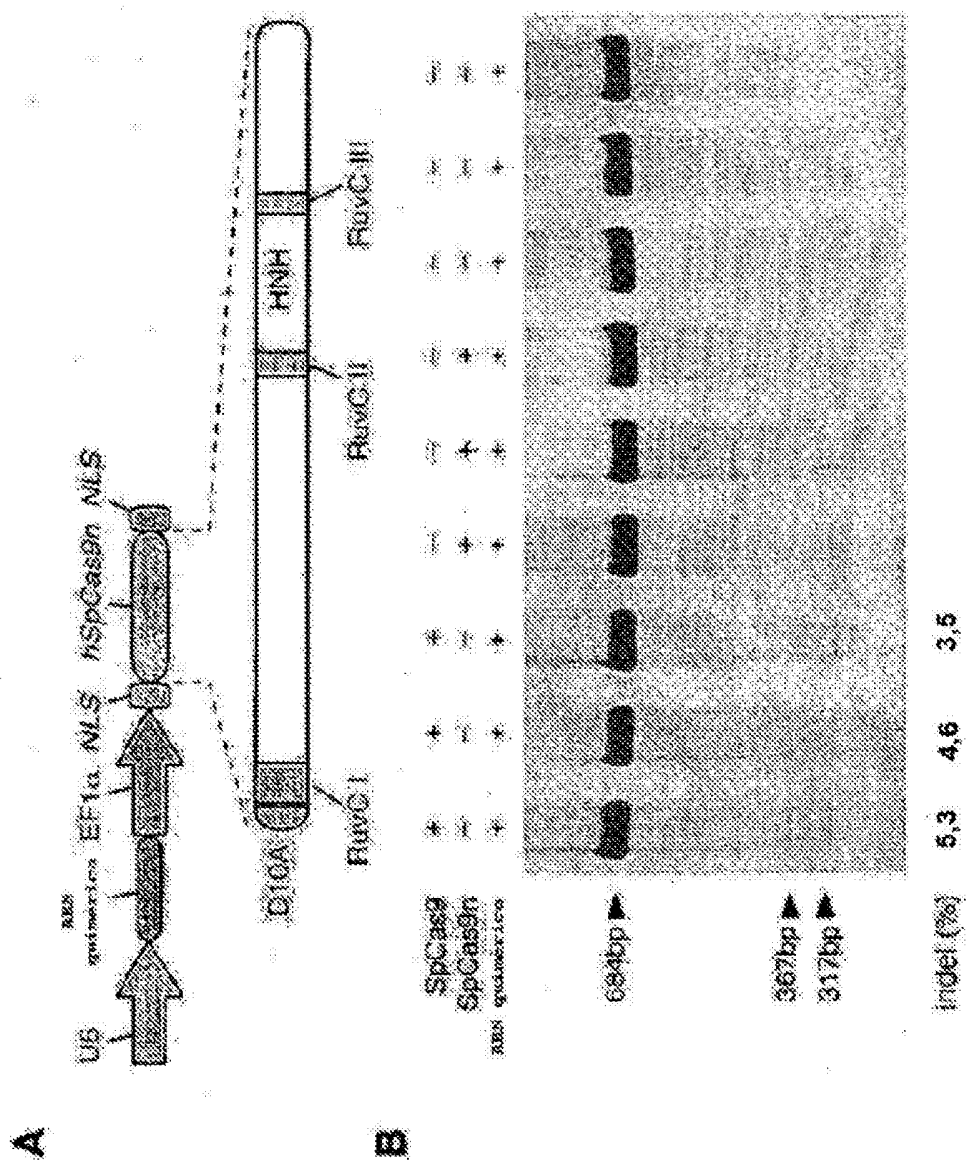


FIG. 4D



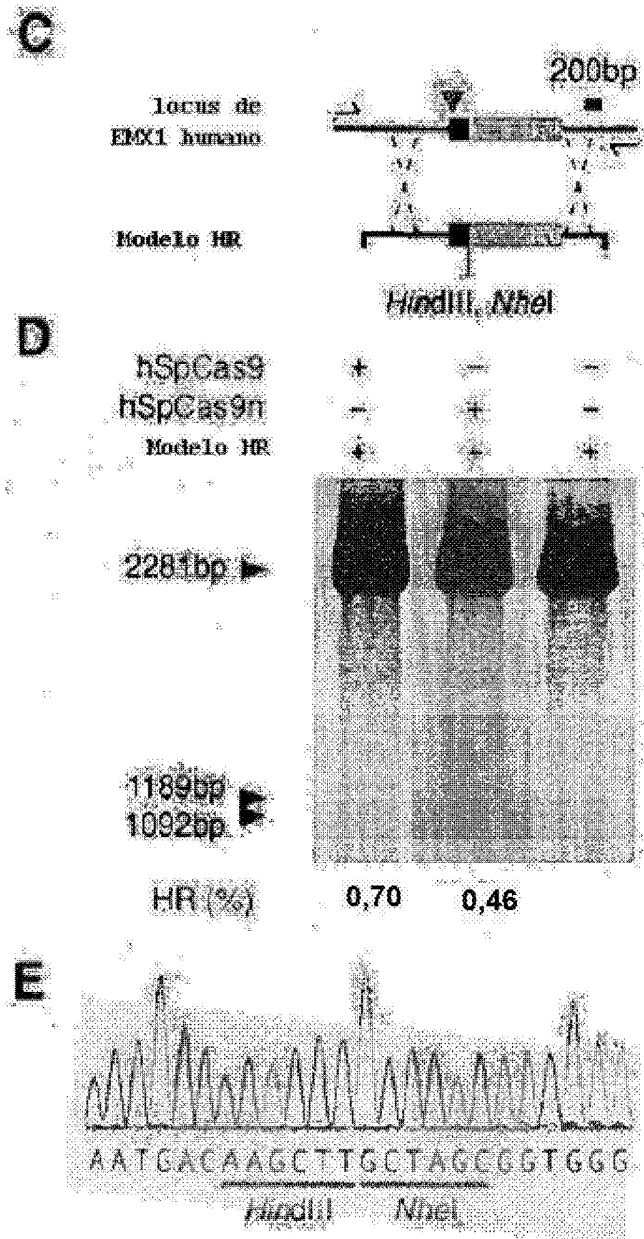


FIG. 5C-E

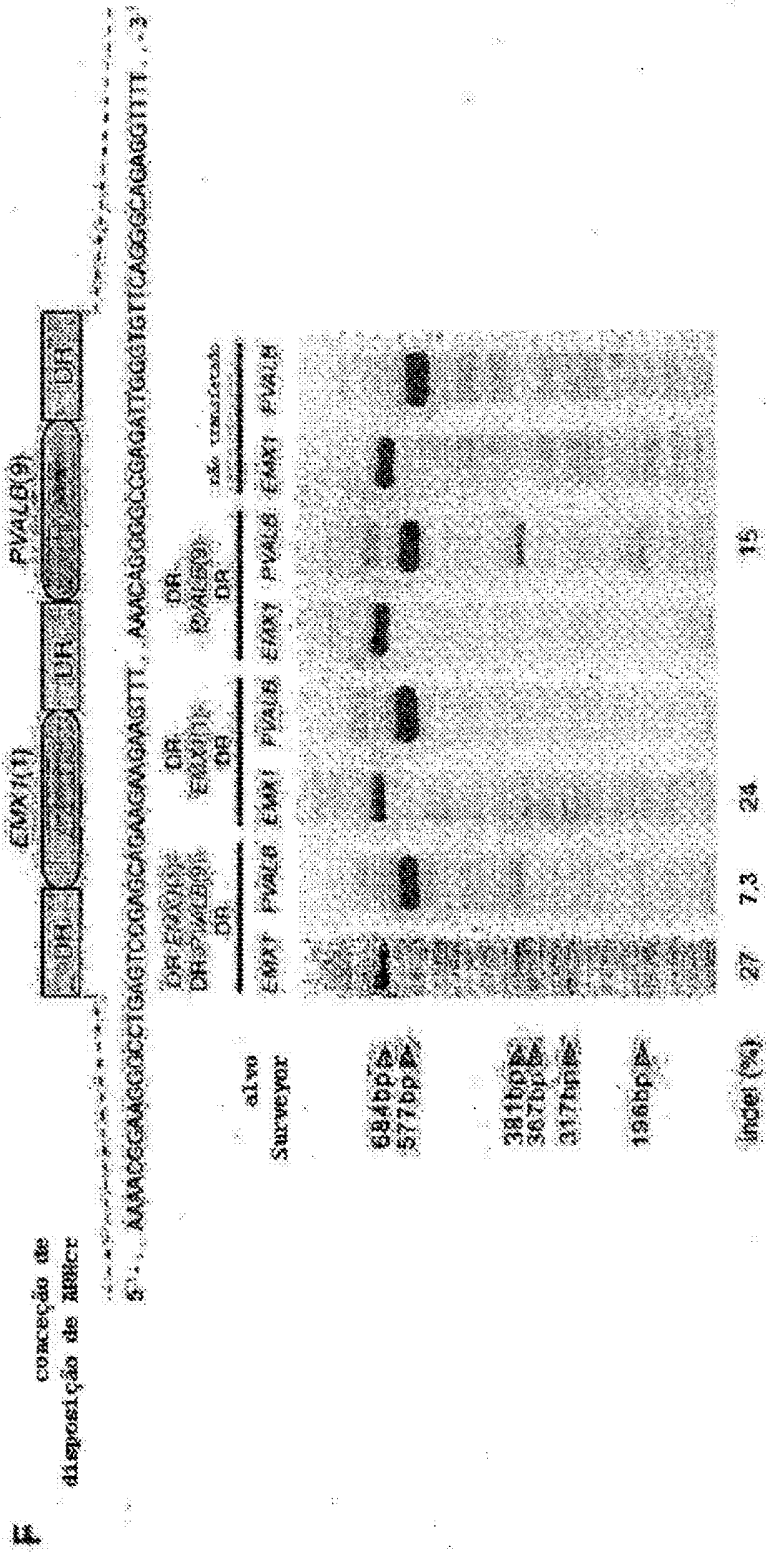
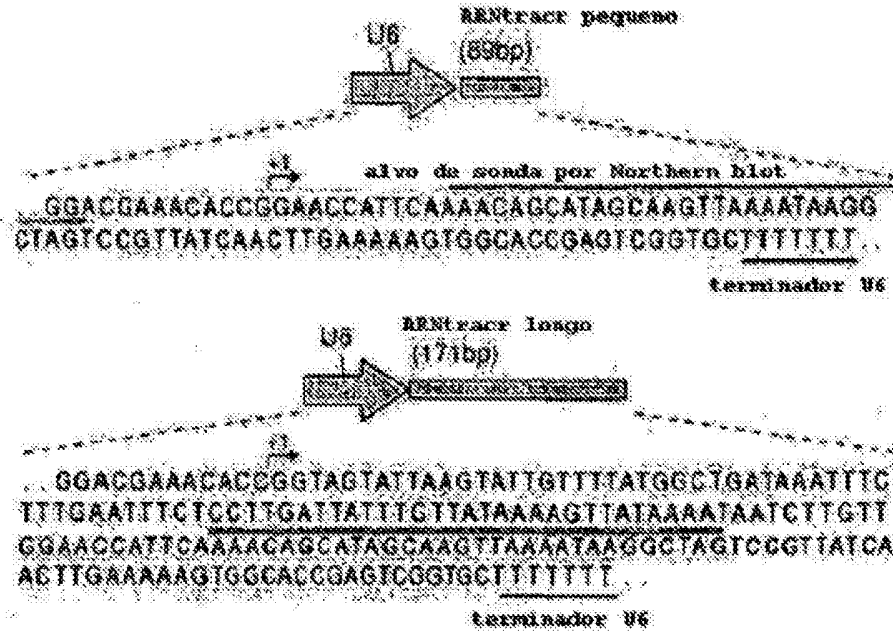


FIG. 5F

A



B

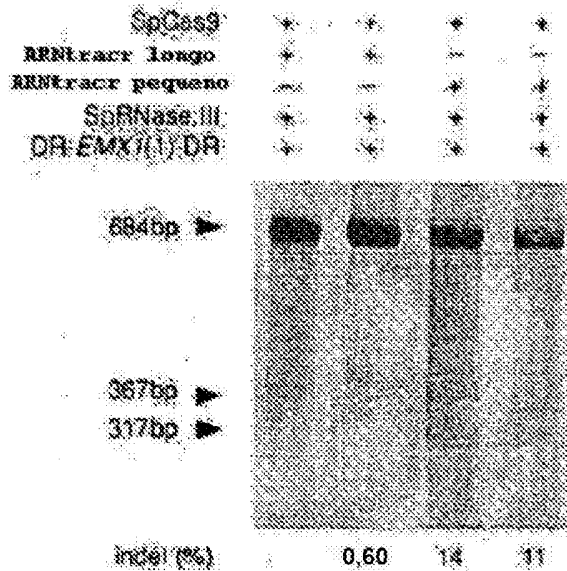


FIG. 6A-B

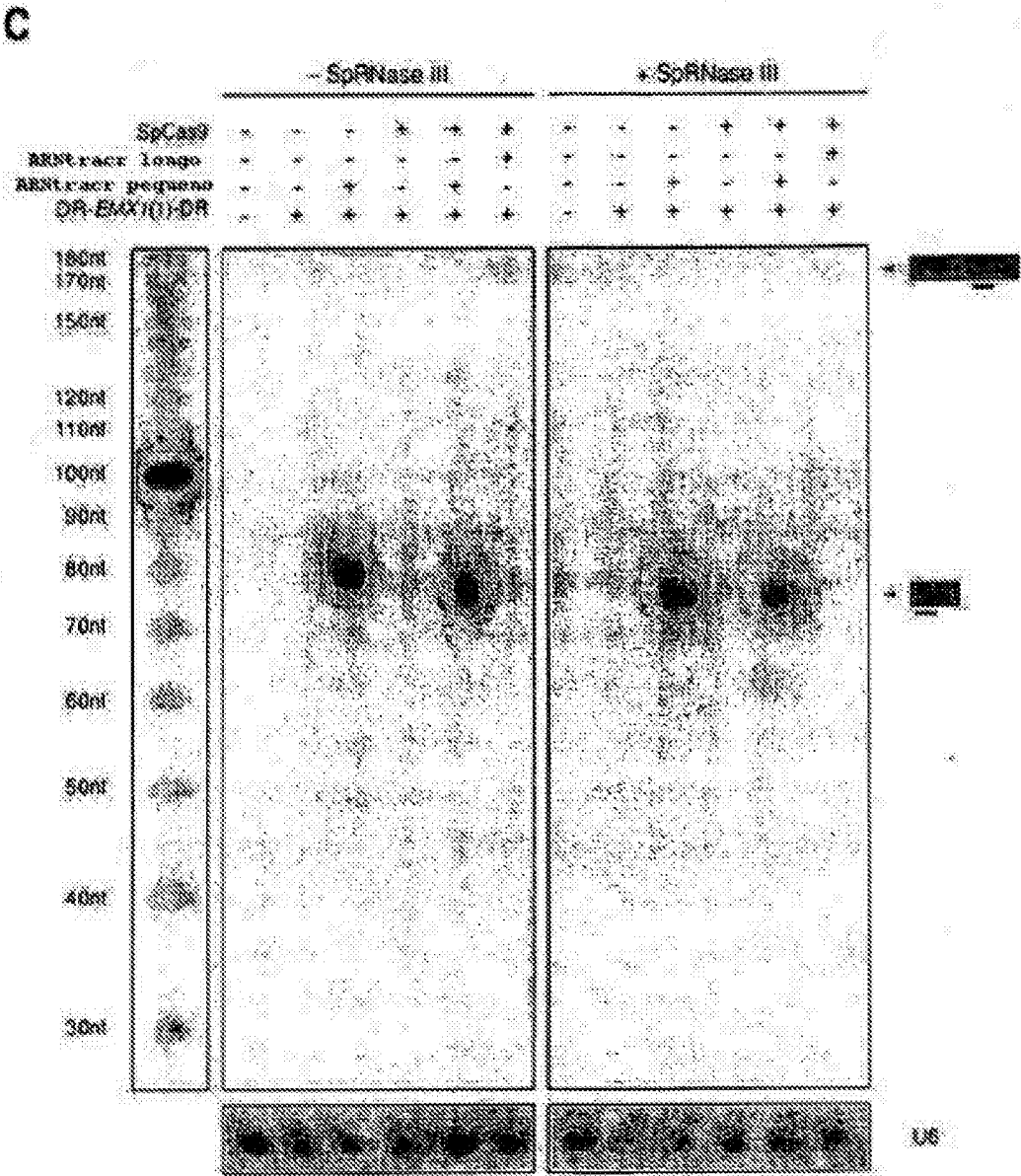


FIG. 6C

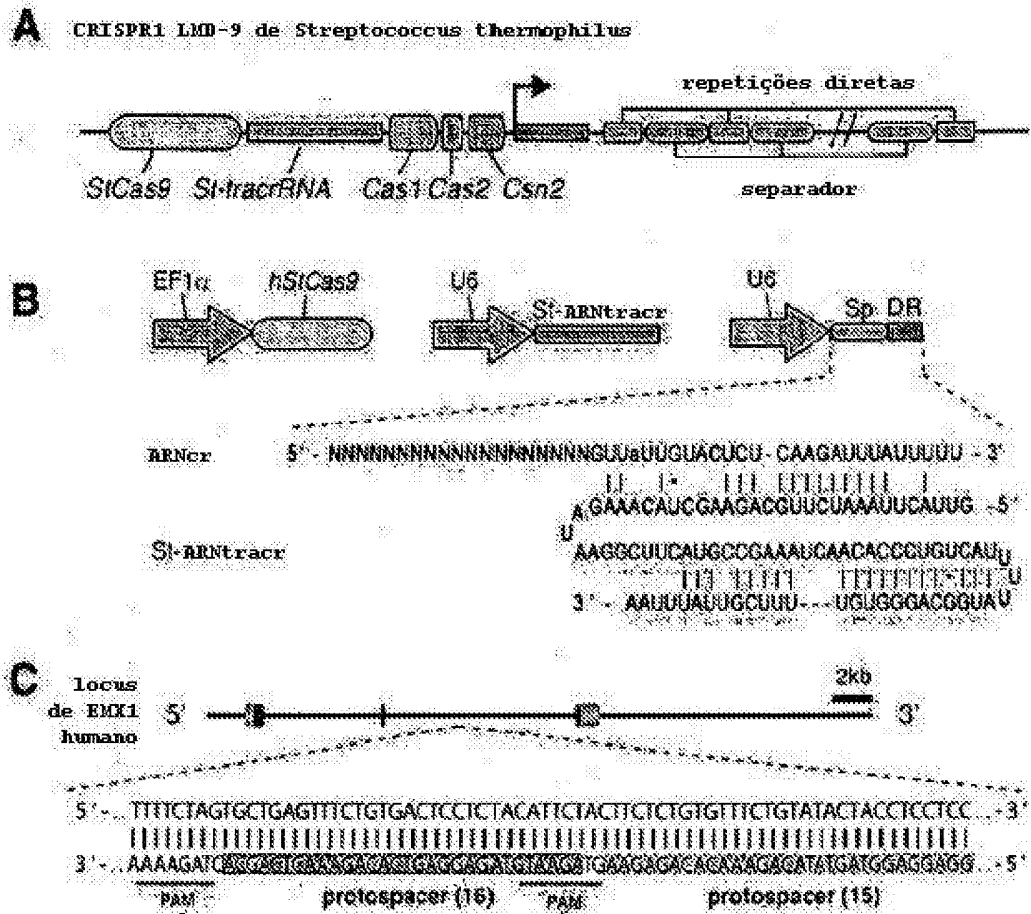


FIG. 7A-C

D

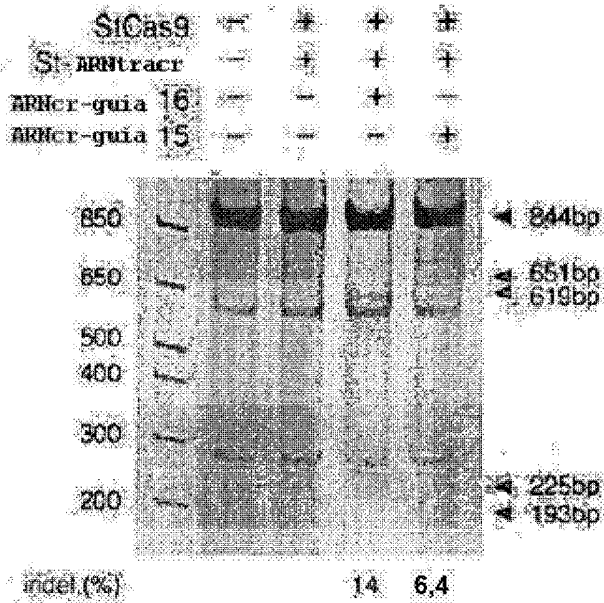


FIG. 7D

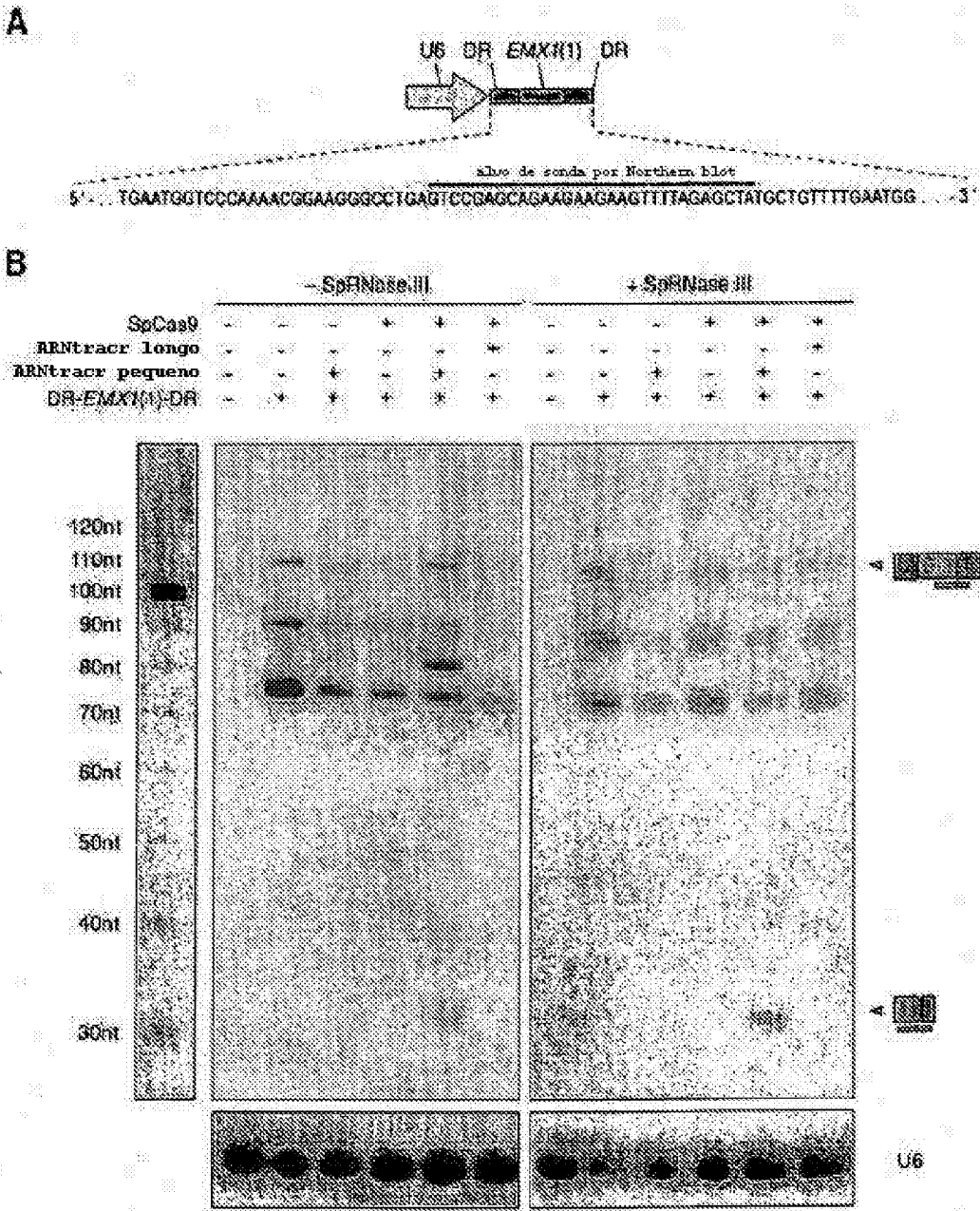


FIG. 9A-B

b

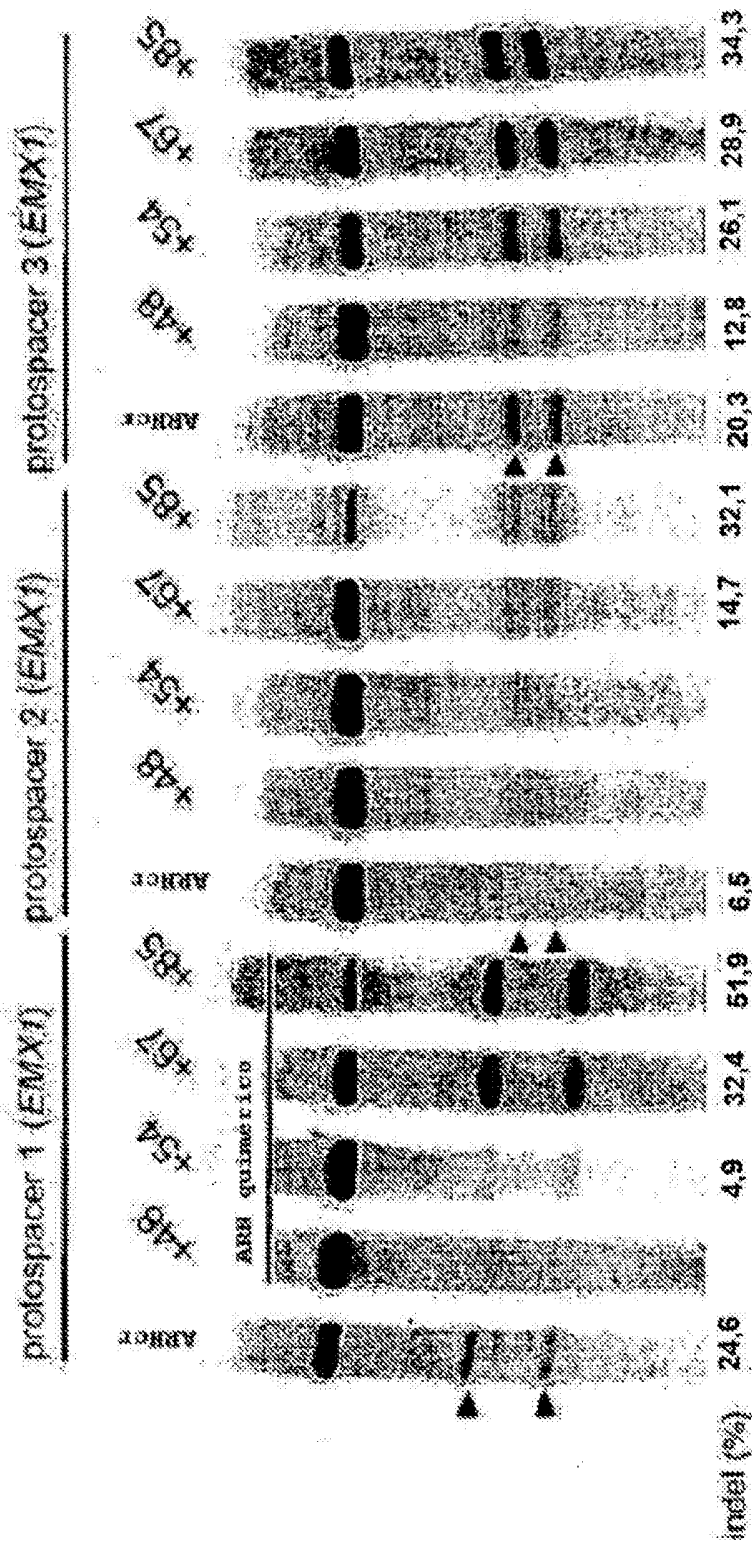


FIG. 10B

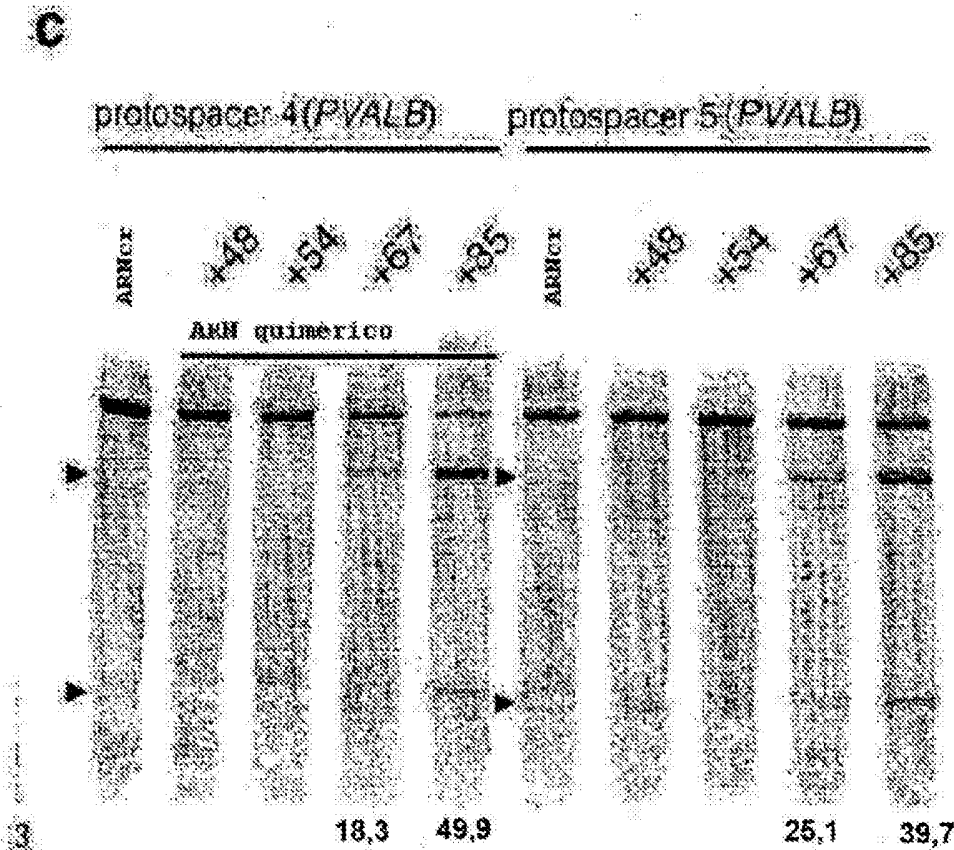


FIG. 10C

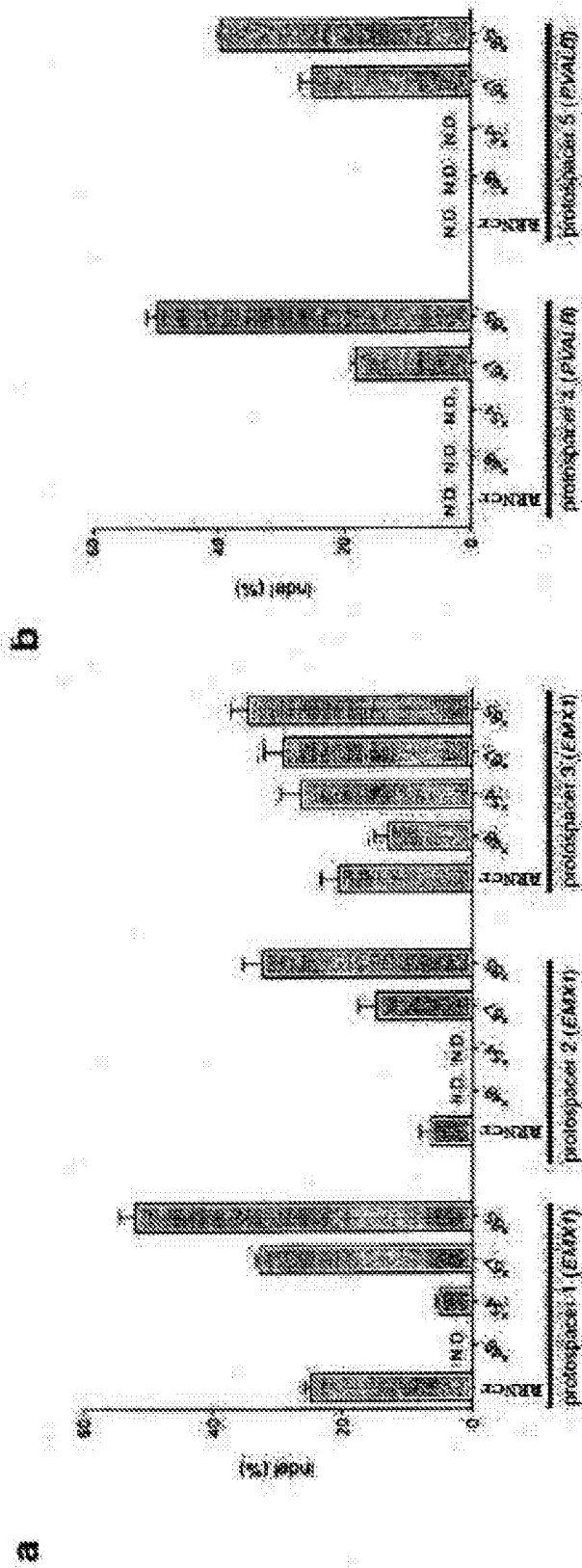


FIG. 11A-B

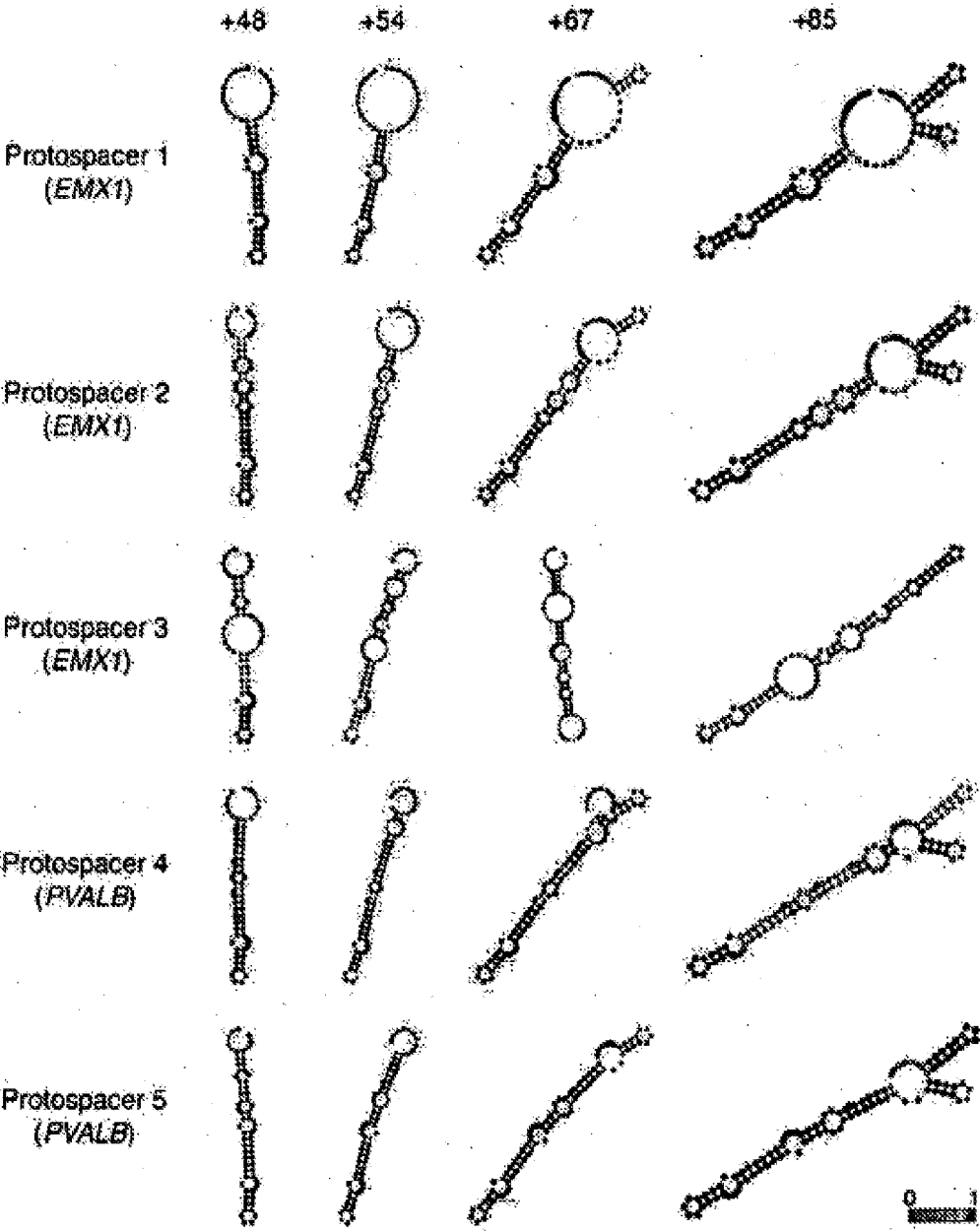
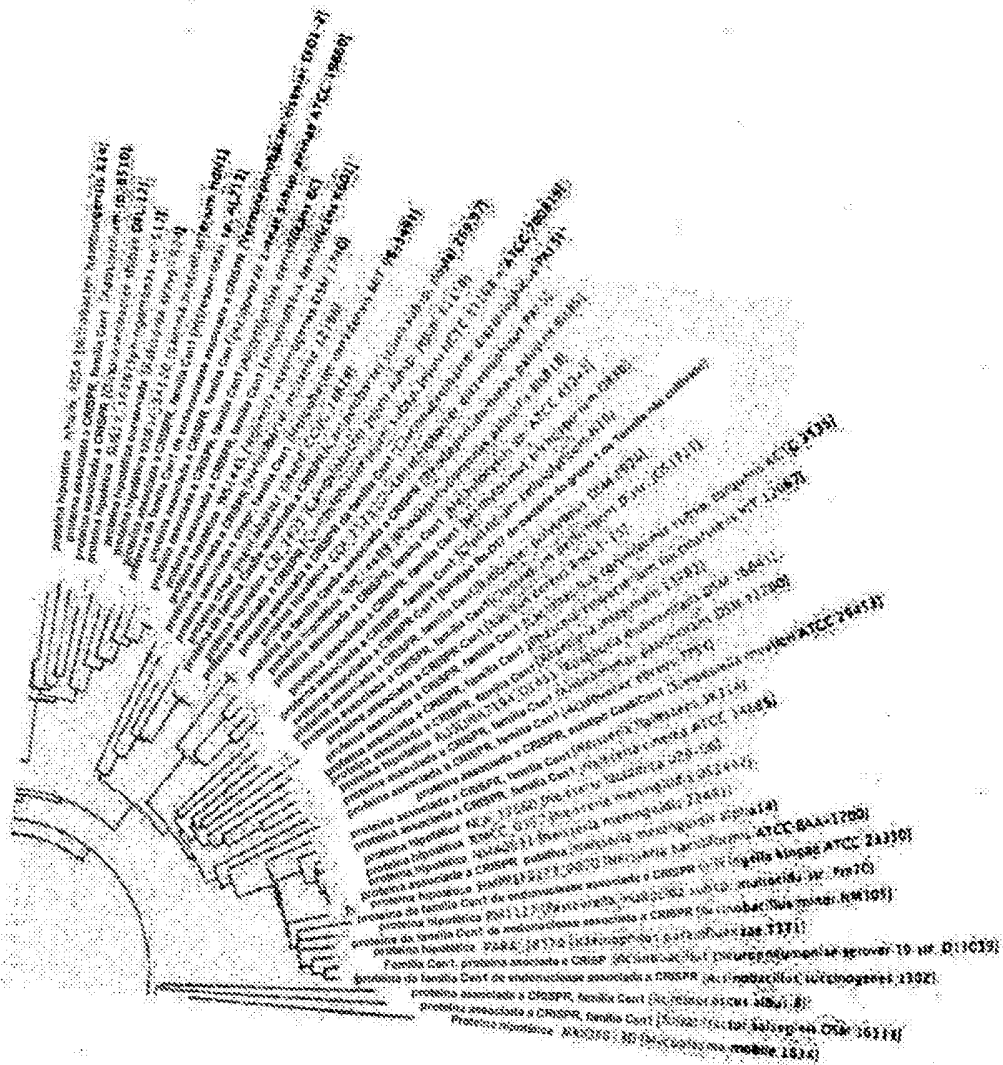


FIG. 12



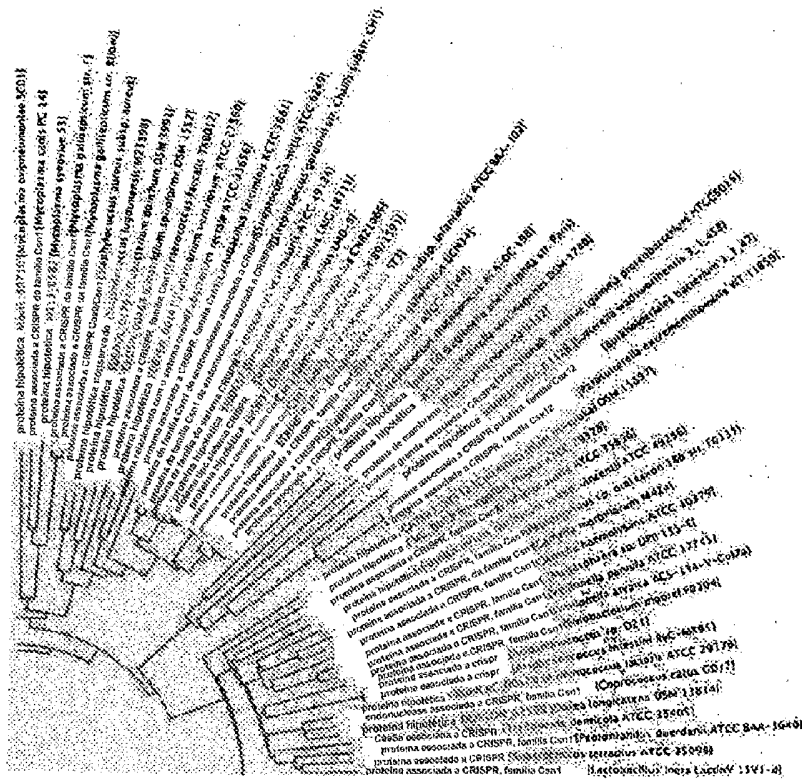


FIG. 13D

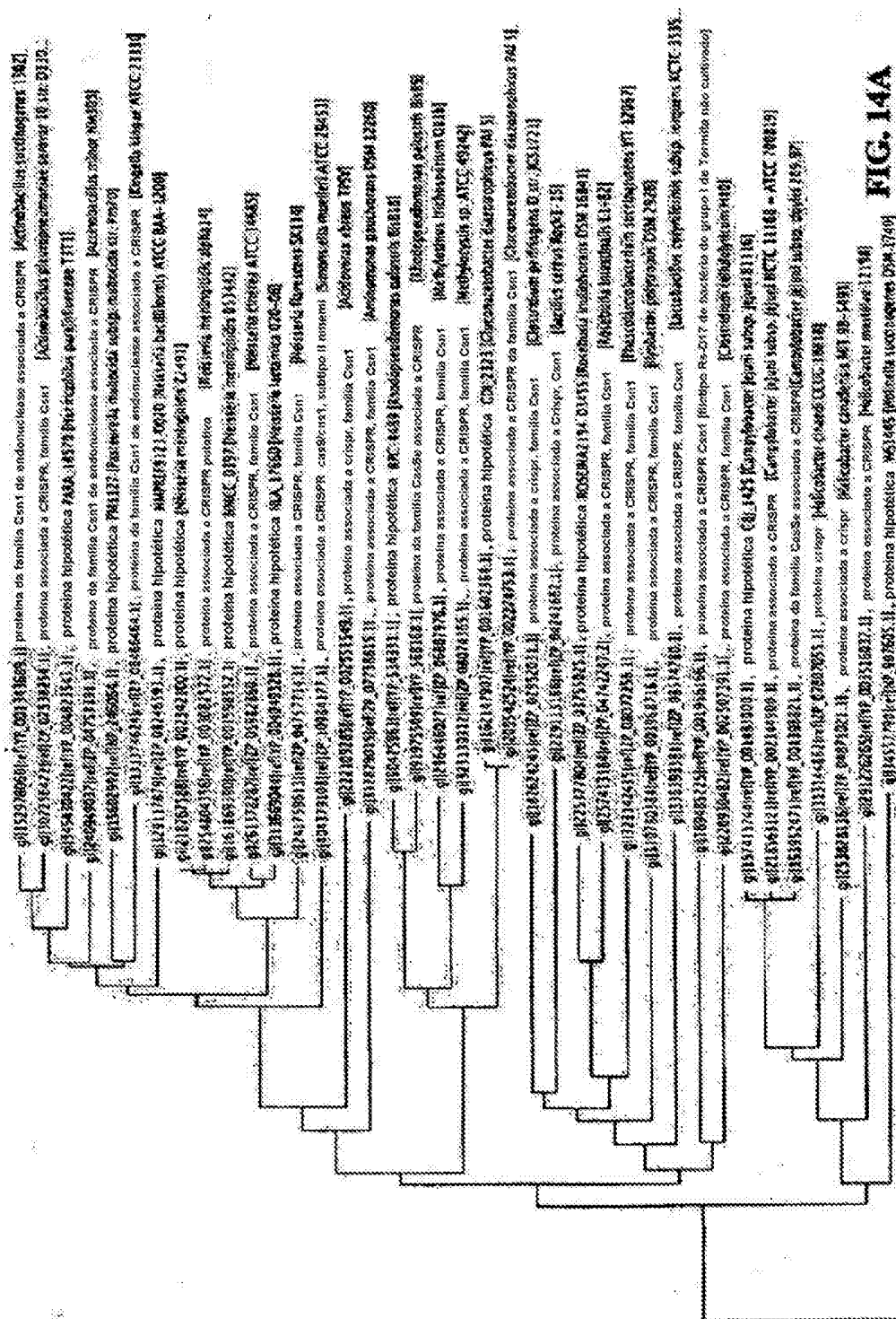






FIG 14C

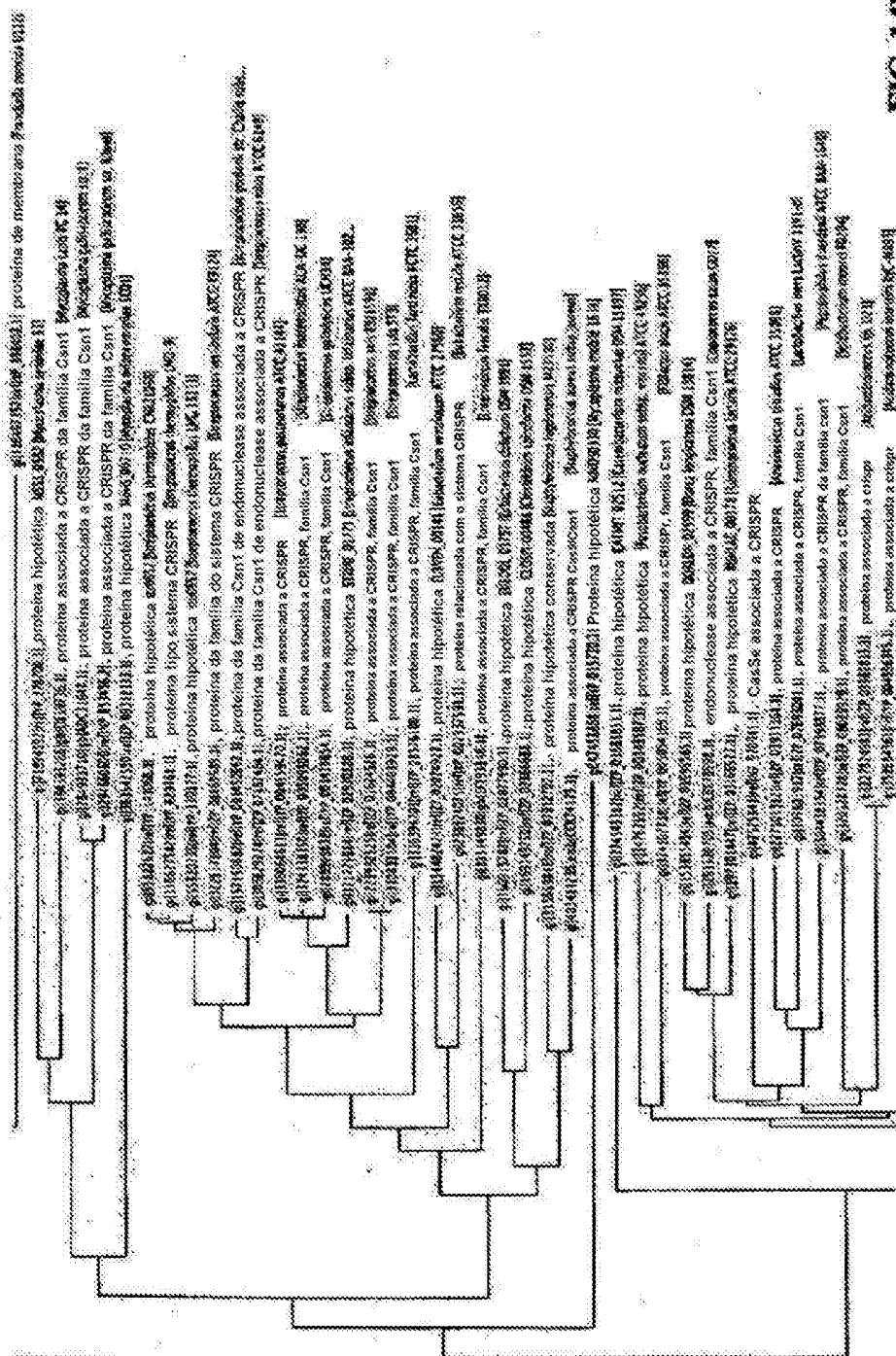
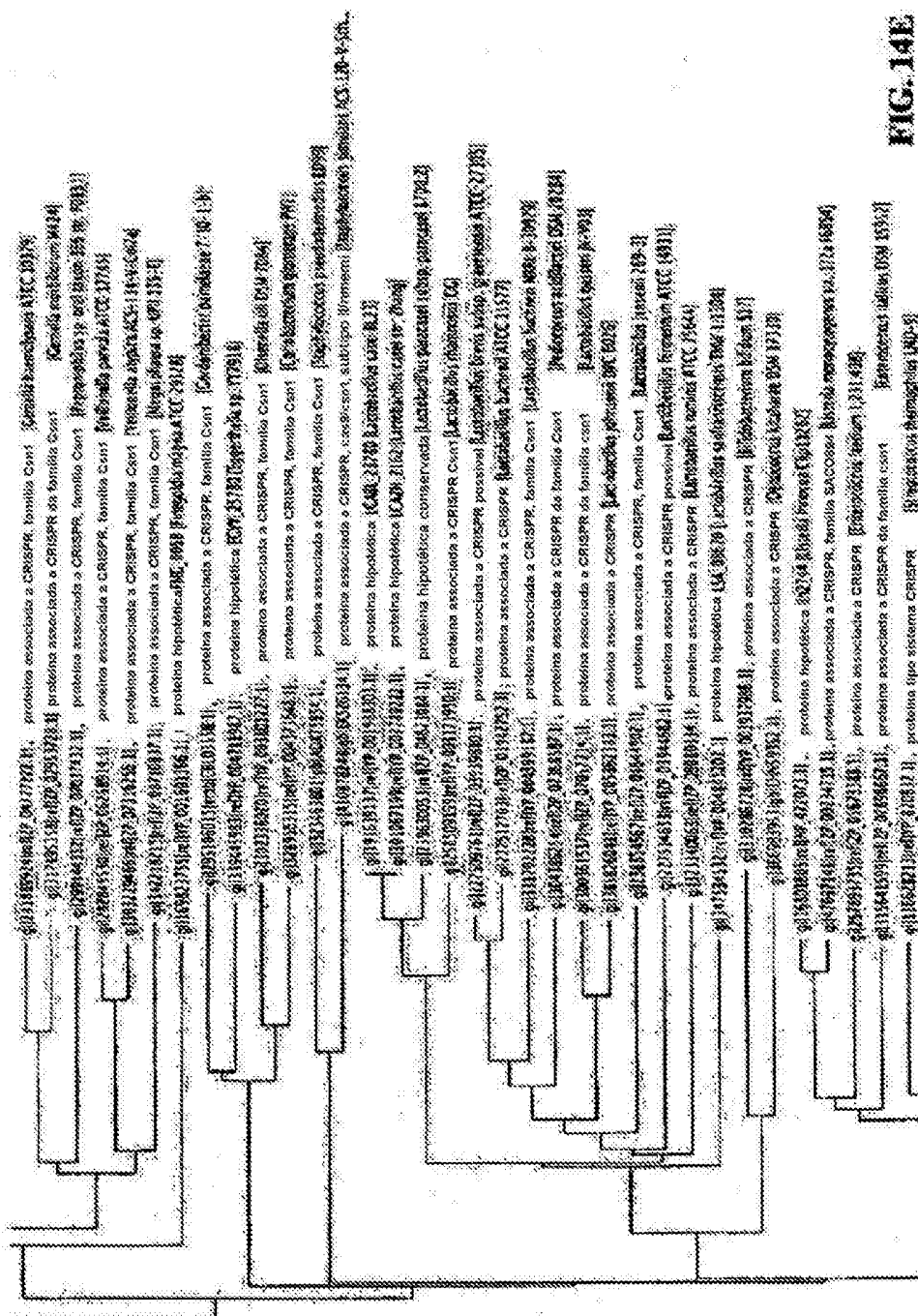


FIG. 14D



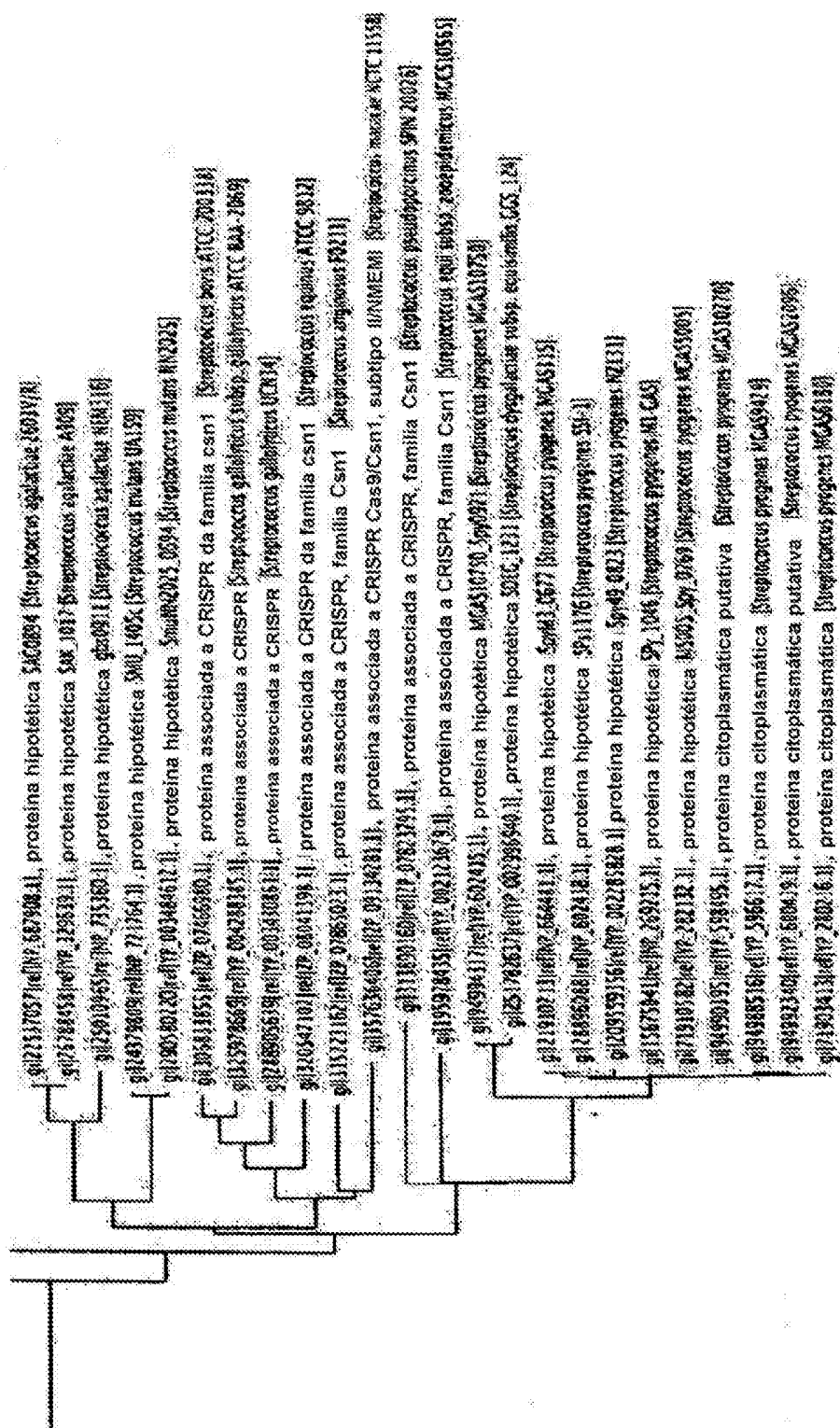
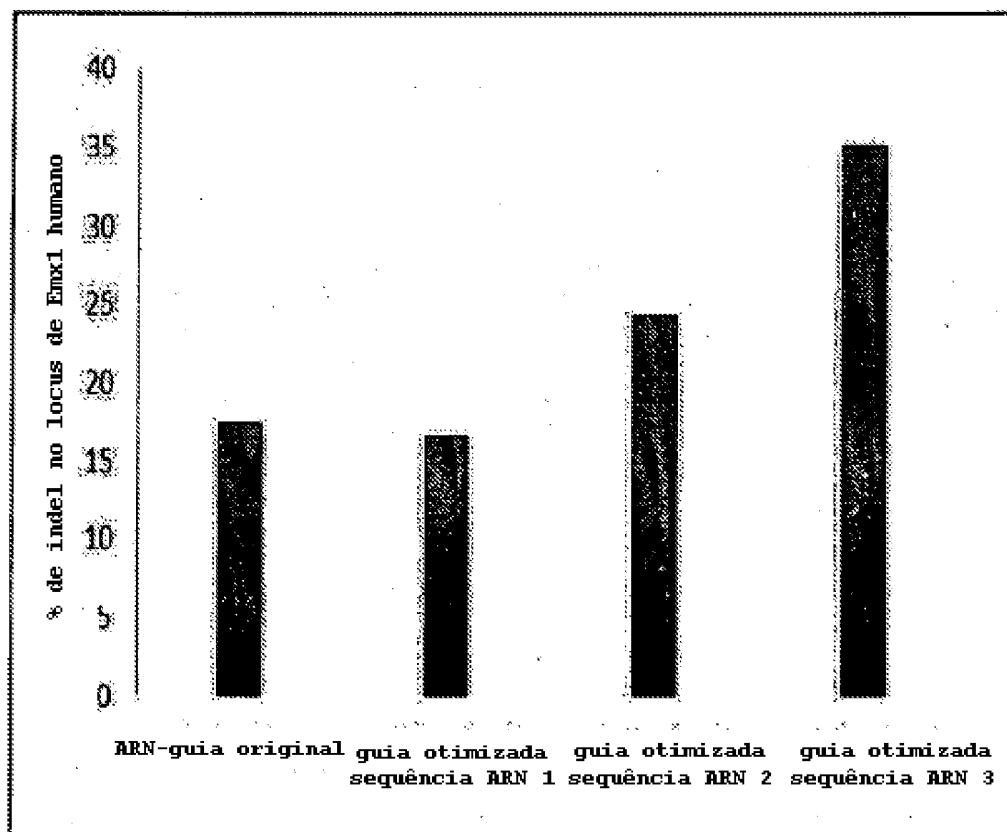


FIG. 14F

**FIG. 15**

conceção de RAN-guia curricular

[illegible]

FIG. 16

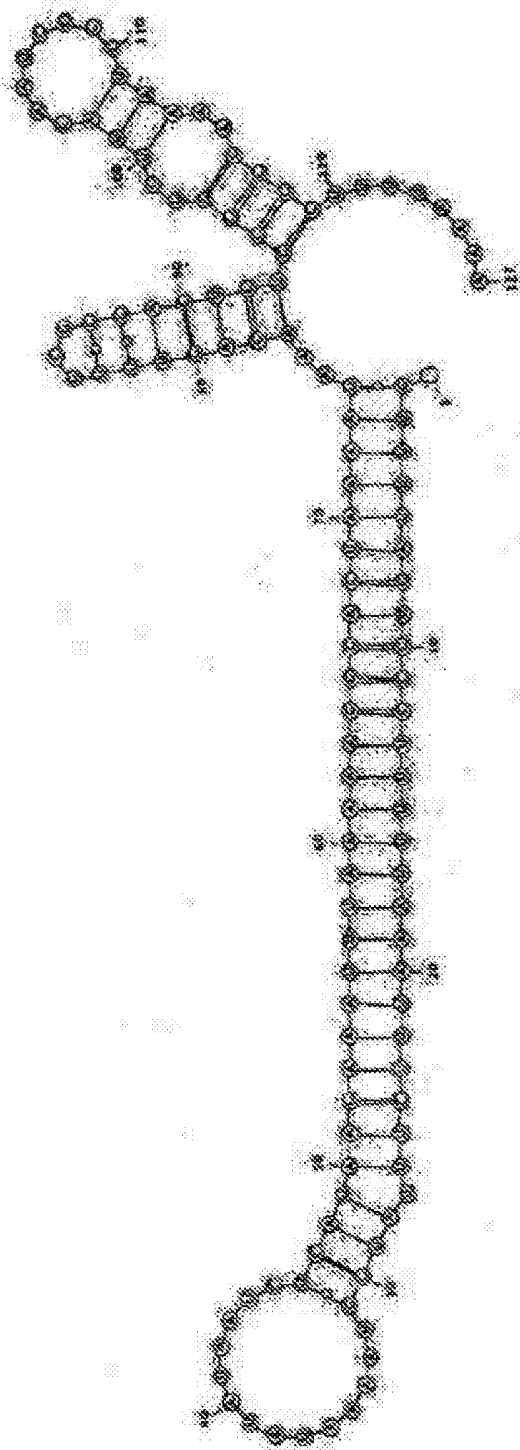
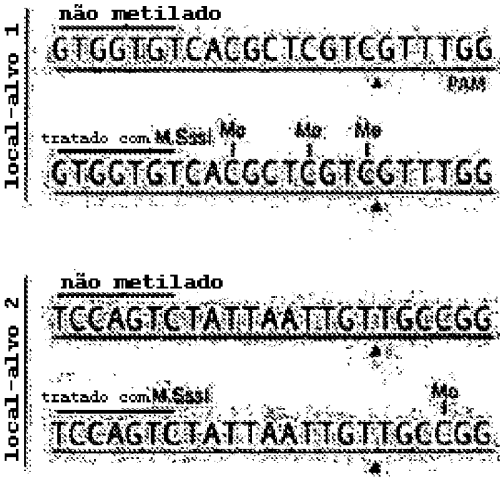


FIG. 17

a



b

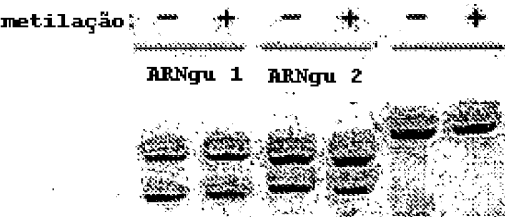


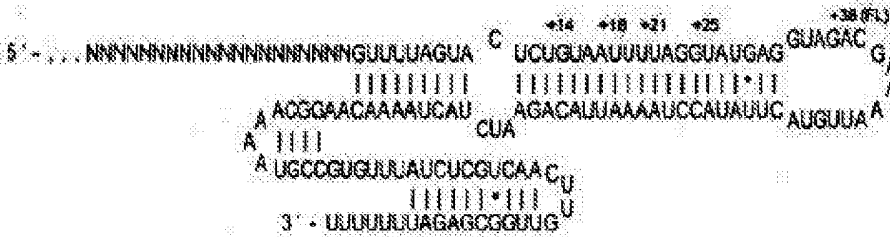
FIG. 19A-B







A



B

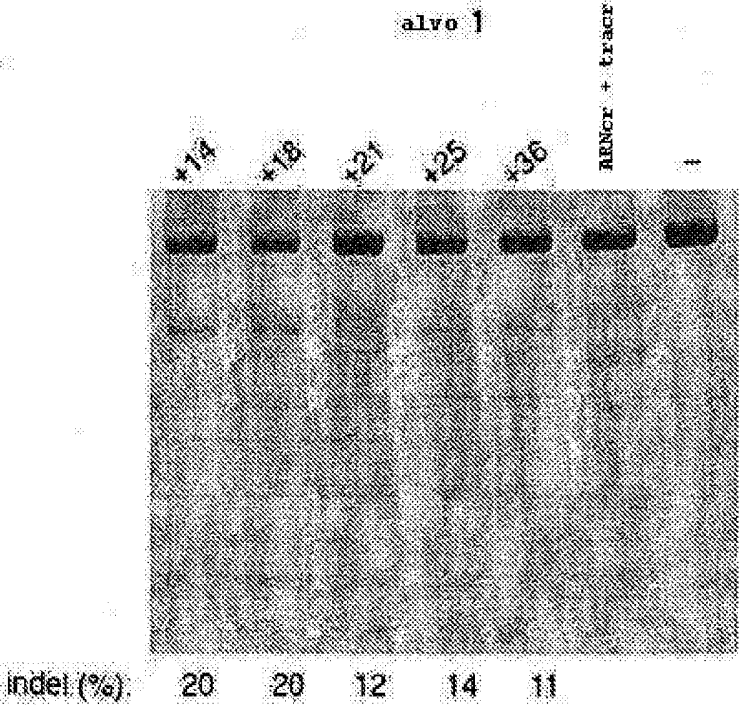


FIG. 23A-B