



A revolucionária técnica de edição genética “CRISPR”

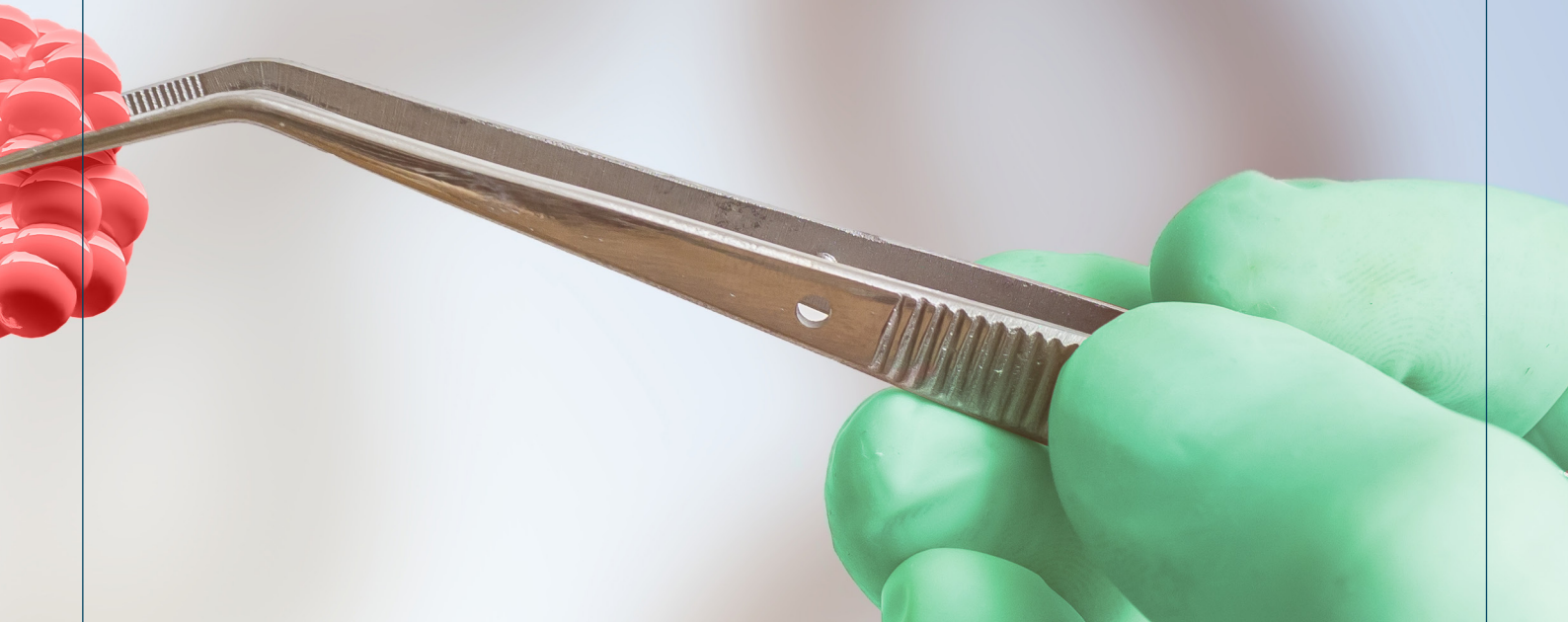
Gabriel José de Carli¹, Tiago Alves Jorge de Souza¹, Tiago Campos Pereira^{1,2}

¹ Programa de Pós-Graduação em Genética, FMRP, Universidade de São Paulo, Ribeirão Preto, SP

² Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo

Autor para correspondência: tiagocampospereira@ffclrp.usp.br

Palavras-chave: sgRNA, Cas9, RNA guia, PAM, endonuclease, NHEJ



A técnica conhecida como “CRISPR” tem impactado de maneira singular o campo da Biologia Molecular por permitir a edição genética de várias espécies de forma simples, rápida e a baixo custo. Essa ferramenta foi desenvolvida a partir da descoberta do locus CRISPR em procariontes, quase trinta anos atrás. Diversos estudos evidenciaram que este locus age como um sistema imune adaptativo de bactérias e arqueias, reconhecendo e degradando o DNA de patógenos, além de armazenar estas informações em seus genomas, protegendo-os assim de eventuais reinfecções. A partir desse conhecimento foi possível desenvolver a tecnologia de edição genética “CRISPR”, composta pela endonuclease Cas9, que funciona como uma tesoura molecular, e por um pequeno RNA guia que direciona os cortes dessa tesoura. A utilização dessa ferramenta tem permitido a inativação de genes (*no caute*), inserção de elementos genéticos (*knock-in*), silenciamento gênico, mapeamento genômico, entre outras aplicações. CRISPR é um exemplo de conexão bem-sucedida entre ciência básica e aplicada, demonstrando que os esforços despendidos na compreensão de mecanismos moleculares presentes em espécies simples podem proporcionar enormes avanços na medicina e biotecnologia.

O HISTÓRICO DA EDIÇÃO GENÉTICA

A possibilidade de alterar de maneira controlada e precisa a informação genética das espécies para fins de pesquisa científica, produção de fármacos, geração de plantas resistentes a estresses bióticos e abióticos – entre muitos outros, como a terapia gênica – tem sido uma busca constante da ciência nas últimas décadas.

Algumas formas de se modificar o DNA já existiam desde a década de 1980, mas elas eram inespecíficas (não se sabia o local exato onde ocorreria a alteração) e apresentavam baixa taxa de sucesso. O primeiro grande passo foi dado em 1985 com a estratégia de *gene targeting* que naquela época se baseava apenas em **homologia** de sequências (Figura 1A). Neste processo, insere-se na célula um fragmento de DNA (denominado DNA doador) apresentando homologia à sequência que se objetiva alterar. Em termos práticos, o DNA doador é em si a nova versão do gene. Por exemplo, suponha que a célula possui um alelo mutante, causador de uma doença genética. Na estratégia de *gene targeting*, administra-se na célula um DNA doador que é correspondente ao alelo selvagem. Assim, por haver elevada homologia entre ambos os alelos, a célula realiza um processo de recombinação, substituindo o mutante pelo selvagem. Apesar do *gene targeting* prover especificidade à modificação genética, sua taxa de eficiência ainda era baixa.

O grande salto seguinte deu-se com o uso de **endonucleases**, a partir de 1989, as quais realizam cortes específicos no genoma (Figura 1B). Essas quebras no DNA são letais para a célula, que então ativa sistemas de reparo para corrigir o dano. Os pesquisadores notaram que, ao ativar estes sistemas de reparo, a eficiência de recombinação com o DNA doador aumentava drasticamente. Exemplos dessas enzimas que agem como “tesouras moleculares” incluem as **meganucleases**, as nucleases dedos de zinco (ZFNs) e as nucleases fusionadas ao efetor semelhante ao ativador de transcrição (TALENs). Apesar de tornarem o processo mais eficiente, elas pecavam por serem muito complexas e caras.

Todas estas abordagens descritas anteriormente são formas (menos ou mais sofisticadas) de **edição genética**, que pode ser comparada a uma edição de textos. Por exemplo, podemos inserir uma letra ou retirar uma vírgula de um manuscrito – equivalente à inserção ou deleção de apenas um nucleotídeo no DNA. Alternativamente, podemos adicionar ou subtrair palavras inteiras do texto, evento análogo ao acréscimo ou remoção de genes inteiros do genoma. Diversas outras formas de edição são possíveis, tais como a *inversão* de uma sentença no texto, assim como múltiplas alterações simultâneas, em quaisquer partes específicas do documento.

Foi então no final do ano de 2012 que um grupo de pesquisadores apresentou o passo mais recente nesta jornada: a técnica de edição genética CRISPR/**Cas9** - ou simplesmente CRISPR - uma estratégia baseada em uma endonuclease guiada por RNA (Figura 1C). Ela apresenta todas as vantagens das abordagens anteriores sendo, contudo, mais simples e barata. Isto ocorre porque as ferramentas anteriores baseavam-se no reconhecimento entre proteína-DNA, cuja interação é complexa; adicionalmente a síntese (de ZFNs e TALENs) envolve tecnologias sofisticadas e, portanto, custosas. Já a CRISPR baseia-se no simples reconhecimento por emparelhamento de sequências nucleotídicas, cuja interação é simples (adenina com uracila; guanina com citosina). Adicionalmente, a síntese (de Cas9 e sgrRNA) abrange técnicas mais tradicionais e, assim, mais acessíveis. Como diversas outras grandes descobertas na ciência, ela começou a partir de uma pesquisa básica, sem grandes pretensões.

A DESCOBERTA DO LÔCUS CRISPR

A história que resultou no desenvolvimento da inovadora técnica de edição genética CRISPR começou no final da década de 1980. Naquele período, pesquisadores japoneses identificaram um locus genômico da bactéria *Escherichia coli* composto por *sequências repetidas* contendo 29 nucleotídeos, intercaladas com **sequências espaçadoras** de 32 nucleotídeos (Figura 2).

Edição genética - conjunto de estratégias que permitem modificações precisas na sequência nucleotídica.

Homologia - similaridade na sequência nucleotídica ou de aminoácidos que reflete uma origem evolutiva comum.

Cas9 (CRISPR-associated protein 9) - uma endonuclease codificada por um gene localizado próximo ao locus CRISPR. Ela é dirigida por um RNA guia até a sequência alvo, na qual realiza cortes nas duas cadeias de DNA.

Endonucleases - enzimas que realizam cortes internos na cadeia de DNA. Por sua vez, as endonucleases de restrição são capazes de clivar, isto é, cortar, em locais precisos no genoma.

Meganuclease - é um tipo de endonuclease de restrição que, devido às suas características, realiza poucos cortes no genoma, tornando-a, assim, tecnicamente mais específica.

Sequências espaçadoras - são pequenas sequências nucleotídicas derivadas de vírus e que são inseridas no locus CRISPR, servindo como uma memória de invasores.

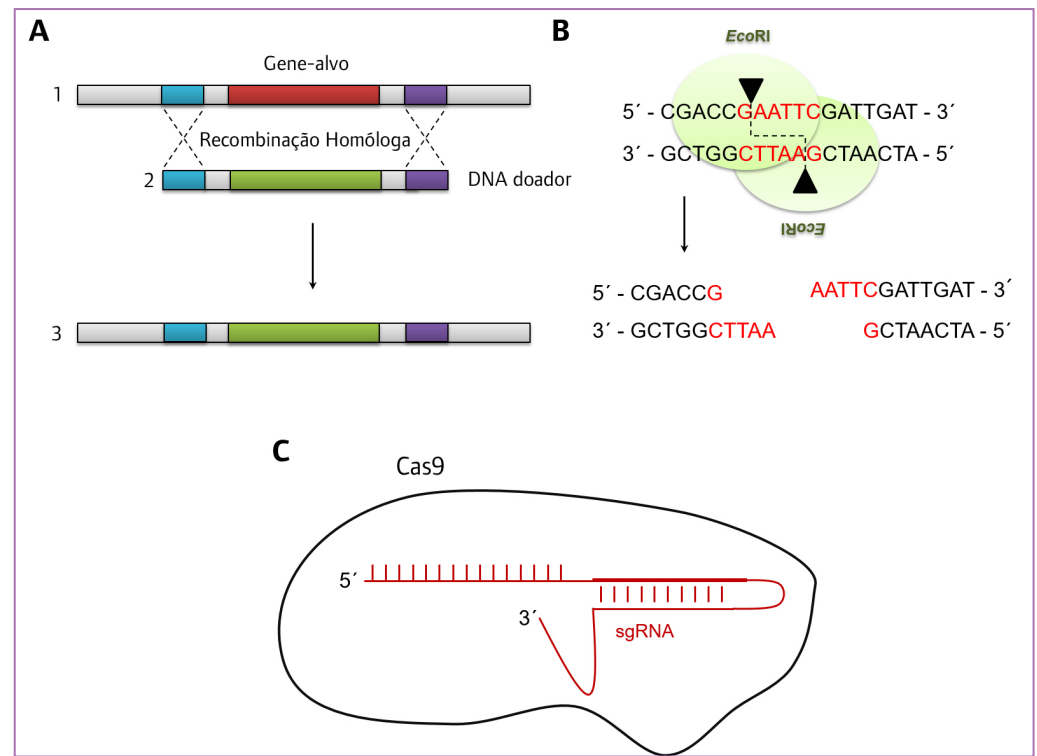


Figura 1.

A. Esquema ilustrativo da estratégia de *Gene targeting* na qual, por meio da recombinação de sequências homólogas (retângulos azuis e roxos), substitui-se (ou se insere) um fragmento de interesse no genoma do organismo. **B.** Mecanismo de ação de uma endonuclease ilustrativa – *EcoRI*, que reconhece e corta uma sequência nucleotídica específica (indicada em letras vermelhas). Os triângulos indicam os pontos de corte da enzima. O resultado da clivagem é apresentado logo abaixo. **C.** Ferramenta de edição genética CRISPR (Cas9 + sgRNA).

Arqueia - um dos três grandes domínios da vida, sendo os outros dois: eukarya e bactéria. O domínio arqueia é representado por organismos unicelulares que tipicamente vivem em regiões consideradas extremas, como fontes hidrotermais e lagos de elevada salinidade.

Acrônimo - uma palavra formada a partir das letras iniciais de uma série de palavras. Por exemplo, ABIN – Agência Brasileira de Inteligência.

Repetições Palindrômicas - são sequências nucleotídicas cuja informação é a mesma quando lidas em um sentido ou no outro, por exemplo: ACGTTGCA.

Helicase - enzima capaz de separar as duas cadeias de DNA (ou de RNA).

Patógeno - qualquer agente causador de uma doença, como certos vírus, bactérias e fungos.

Esta descoberta, assim como diversas outras na história da ciência, permaneceu negligenciada pela comunidade científica por mais de uma década, até o começo do século XXI, quando vários grupos de pesquisa identificaram regiões semelhantes em diversas outras espécies de bactérias e **arqueias**. A partir dessas observações foi cunhado o **acrônimo** CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats* ou **Repetições Palindrômicas** Curtas Agrupadas e Regularmente Interespaçadas) para designar essa região genética evolutivamente conservada em procariontes e de arquitetura peculiar. A partir de então, os cientistas começaram a dar mais atenção a este locus, visando identificar seu papel biológico nos procariontes.

A FUNÇÃO BIOLÓGICA DO LÓCUS CRISPR

A compreensão do papel biológico do locus CRISPR demandou aproximadamente uma década. Inicialmente, pensava-se que ele estivesse relacionado com o sistema de reparo a danos no DNA em procariontes, mas análises de bioinformática indicavam a presença de grandes quantidades de material exógeno

(oriundo de vírus, por exemplo) no locus – algo inesperado. Ao se analisar as regiões adjacentes ao locus CRISPR foram encontrados genes codificadores de nucleases, polimerases e **helicases**, os quais foram nomeados *Cas* e numerados de 1 em diante (*Cas1*, *Cas2*, ... *Cas9*, etc) (MALI *et al.*, 2013).

Curiosamente, a elucidação da real função de CRISPR começou por meio de experimentos de pesquisadores que trabalhavam na indústria de laticínios Danisco. Eles perceberam que bactérias da espécie *Streptococcus thermophilus* (usada na produção de leite e queijo) expostas a certos vírus apresentavam, posteriormente, sequências nucleotídicas destes vírus no locus CRISPR (BARRANGOU *et al.* 2007).

Dessa forma, levantou-se a hipótese (posteriormente comprovada) de que quando invadidas por um vírus, as bactérias e arqueias seriam capazes de incorporar um fragmento do DNA do **patógeno** como novas regiões espaçadoras do locus CRISPR, o qual agiria assim como um sistema de registro molecular de invasores anteriores, ou em outras palavras, um sistema imune adaptativo baseado em ácidos nucleicos. Uma vez contendo

a informação genética do invasor, a bactéria poderia utilizá-la para combater o patógeno. Isso é possível porque a Cas9 – uma das proteínas associadas ao locus CRISPR – atua como uma endonuclease guiada por RNA, ou seja, ela é capaz de clivar moléculas de DNA que sejam complementares ao RNA nela presente.

Neste momento é importante lembrar que o locus CRISPR possui dentro de si um pequeno fragmento do DNA do invasor (o novo espaçador) (Figura 2). Esta sequência será transcrita, gerando o RNA que guiará Cas9, levando ao combate específico do patógeno, sem afetar as sequências genéticas da

própria bactéria. Além deste primeiro RNA direcionador, tecnicamente denominado de crRNA, outra molécula de RNA atua conjuntamente (conhecido como tracrRNA) com papel apenas de ativar a Cas9.

Logo após a elucidação do funcionamento do locus CRISPR, ficou claro que o mesmo poderia ser redirecionado para engenharia genética de procariontes assim como de eucariontes. Devido à sua simplicidade e características, a Cas9 encontrada na bactéria *Streptococcus pyogenes* (*S. pyogenes*) foi escolhida como modelo para o desenvolvimento da técnica de CRISPR. Contudo, há variações utilizando outras nucleases.



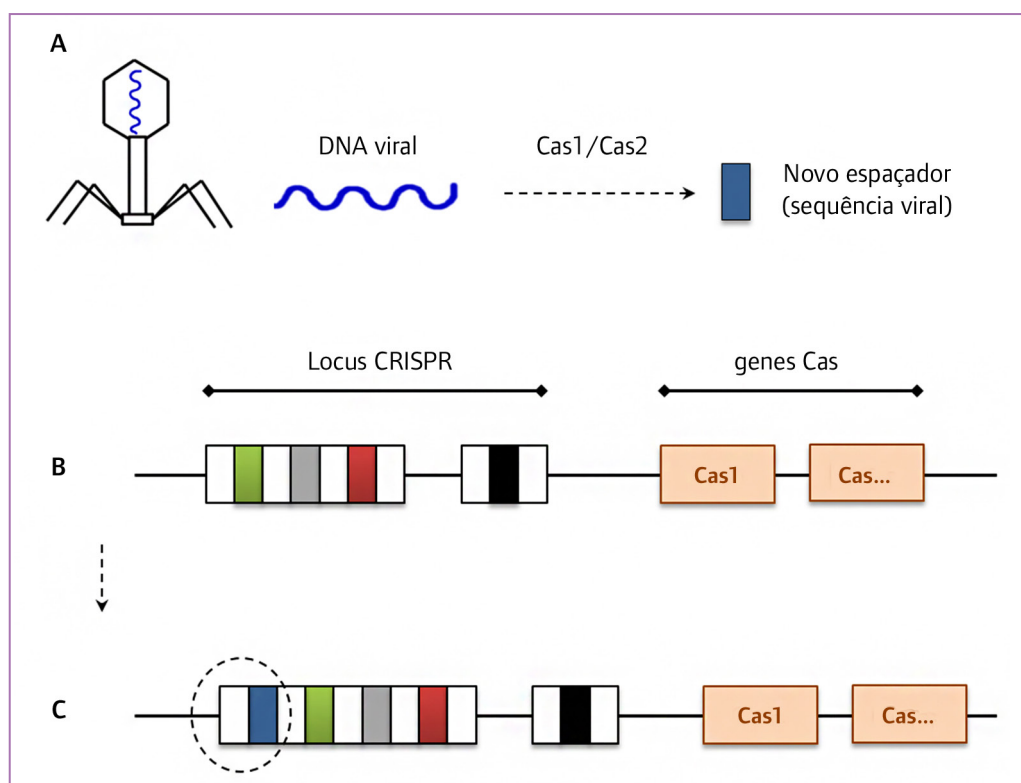


Figura 2. Estrutura do locus CRISPR de *Escherichia coli*. (A) Enzimas Cas clivam o genoma de bacteriófagos invasores. (B e C) Fragmentos do DNA viral são incorporados como novos espaçadores no locus CRISPR, representados pelos pequenos retângulos verticais coloridos em azul, verde, cinza, vermelho e preto. As sequências repetidas são indicadas pelos pequenos retângulos verticais em branco. O círculo tracejado indica a integração do fragmento do vírus.

Sequência PAM - uma curta sequência nucleotídica (e.g., NGG) próxima à sequência alvo, porém localizada na cadeia oposta de DNA. Ela é essencial para a interação de Cas9 com o DNA alvo. NGG é um motivo PAM específico para a Cas9 de algumas bactérias, incluindo a espécie *Streptococcus pyogenes*. Contudo, endonucleases Cas de outras espécies podem requerer variações deste motivo. Por exemplo, Cas9 de *Neisseria meningitidis* requer o domínio NNNNGATT.

OS PRINCÍPIOS DA TÉCNICA DE CRISPR

A técnica de CRISPR é muito similar ao sistema imune presente em *S. pyogenes* sendo que a principal diferença está na molécula de RNA utilizada como guia. Como visto anteriormente, o sistema imune bacteriano baseia-se em dois RNAs (tracrRNA e crRNA) que formam um complexo para ativar e guiar Cas9, respectivamente. Na técnica de CRISPR esses dois RNAs são fusionados em apenas um, denominado sgRNA (sigla do inglês para *single guide RNA*, ou RNA guia único) (VIEIRA *et al.*, 2016). Dessa forma, a técnica de CRISPR depende apenas de dois elementos: o sgRNA e a endonuclease Cas9 (RAN *et al.*, 2013). É importante ressaltar que o sgRNA pode ser desenhado para direcionar Cas9 para qualquer local no genoma.

Assim, uma vez que se administre o complexo Cas9/sgRNA na célula (ou no orga-

nismo), a endonuclease é guiada até certa região no genoma complementar ao sgRNA nela presente e que possua uma **sequência PAM** flanqueando a região. Caso haja total emparelhamento entre a sequência guia e o DNA alvo este é clivado (Figura 3), resultando em uma quebra de fita dupla de extremidades abruptas (RAN *et al.*, 2013). Este tipo de dano no DNA é letal para a célula, que tentará remediar via mecanismos endógenos de reparo, tais como o NHEJ e o HDR.

O processo de reparo NHEJ (*Non-Homologous End Joining* ou junção não-homóloga das pontas) é o mais comum e propenso a erro, *i.e.*, normalmente insere ou deleta alguns pares de bases na região de ligação das extremidades (Figura 4A). Já no processo de HDR (*Homology Directed Repair* ou reparo direcionado por homologia) a restauração da quebra se baseará na existência de uma outra molécula de DNA semelhante à quebrada (*i.e.*, homólogo – o DNA doador), que

servirá como molde para o reparo (Figura 4B). Portanto, enquanto o NHEJ resulta em um **reparo intrinsecamente imperfeito**, HDR promove um reparo perfeito de acordo com o molde (WYMAN; KANAAR, 2006).

Baseando-se nessas características de clivagem da Cas9 e dos sistemas de reparo do DNA, pesquisadores converteram o sistema imune CRISPR, cuja função original era de combate a patógenos, em uma ferramenta de edição genética de mesmo acrônimo.

Reparo intrinsecamente imperfeito - é um tipo de reparo que leva a inserção e deleção de pares de bases, denominados *indels*.

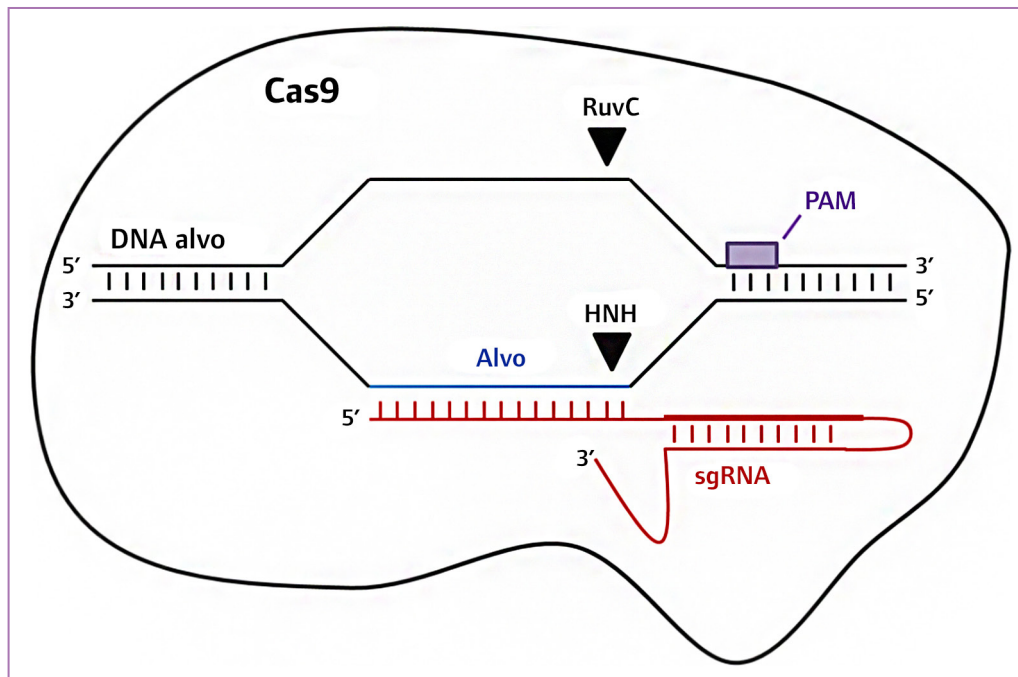


Figura 3. Esquema ilustrativo do funcionamento da técnica de CRISPR. Evidenciando Cas9, sgRNA e PAM no DNA-alvo. Os triângulos invertidos indicam os sítios de quebra do DNA por ação dos domínios catalíticos RuvC e HNH da Cas9.

EDIÇÃO GENÉTICA VIA CRISPR

Lembre-se de que edição genética envolve dois pilares: (i) em uma região específica do genoma (*e.g.*, um gene ou uma pequena sequência nucleotídica) ocorrerá (ii) uma alteração específica (*e.g.*, troca de uma citosina por uma adenina). O sgRNA é o elemento que guiará a Cas9 para o ponto específico do genoma (o primeiro pilar) ao passo que o DNA doador permitirá a alteração desejada (o segundo pilar). Portanto, o DNA doador deve ter homologia com a sequência no genoma que se quer editar, além de conter a nova informação nucleotídica (a edição em si).

Existem diferentes formas de se editar o DNA de acordo com o interesse. A primeira delas é o *knock-in* (KI), empregado para

se introduzir uma nova sequência de interesse no genoma do organismo, tal como um transgene que torne plantas resistentes a determinada praga. Outra possibilidade é realizar uma substituição alélica (que é uma forma de KI), na qual é introduzida uma sequência muito semelhante ao gene, mas com pequenas alterações, modificando-se assim o alelo. Por meio dessa abordagem é possível permutar um alelo mutante (associado à determinada doença genética) por um alelo selvagem – ou seja, uma estratégia de terapia gênica. Em ambos os casos utilizam-se um DNA doador e reparo via HDR.

Curiosamente, apesar de modificações genéticas precisas serem possíveis apenas com o uso de um DNA doador (via HDR), o uso mais frequente de CRISPR é para o uso de modificações genéticas menos preci-

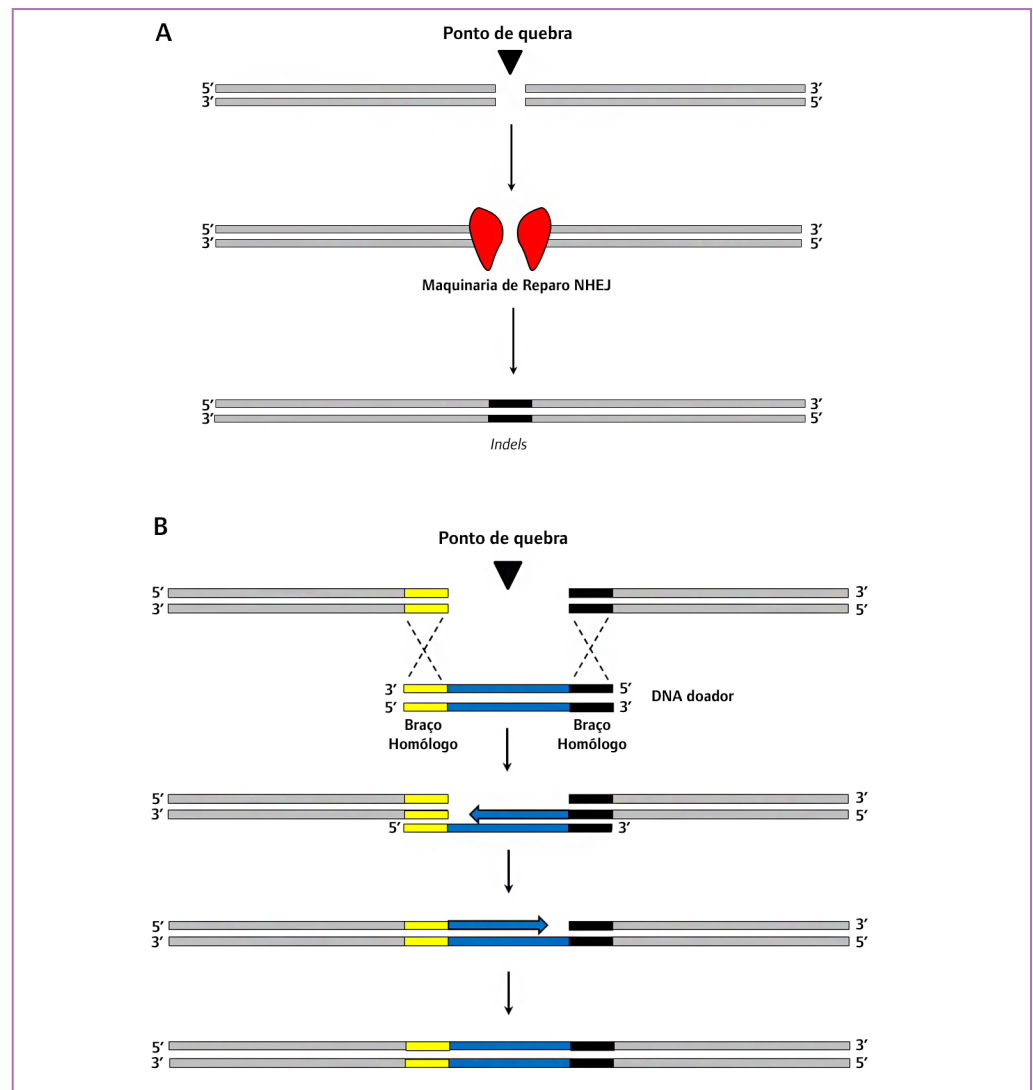


Figura 4. Esquema ilustrativo dos mecanismos de reparo (A) NHEJ (*Non-Homologous End Joining* ou junção não-homóloga das pontas) e (B) HDR (*Homology Directed Repair* ou reparo direcionado por homologia).

Nocaute - inativado. Refere-se a qualquer gene que tenha sido modificado, de tal forma a não mais expressar seu produto funcional final correspondente.

Deleção gênica - processo molecular que envolve a remoção de um gene de dentro do genoma.

sas (sem DNA doador, via NHEJ). Lembre-se de que NHEJ promove a inserção ou deleção de um ou poucos nucleotídeos no sítio clivado, resultando, frequentemente, na inativação deste gene, caso haja mudança de fase no quadro de leitura. Este tipo de edição genética gera **nocautes** (KO) e constitui o maior interesse dos geneticistas pois permite estudar a função dos inúmeros genes identificados a cada dia, isto porque a alteração fenotípica resultante da inativação de certo gene é indicativa da função biológica do mesmo. Se, por exemplo, ao se inativar o fictício gene *Slim* em camundongos observa-se que os mesmos permanecem

magros mesmo quando submetidos a dietas hipercalóricas – pode-se concluir que este gene está associado ao metabolismo de gorduras.

Outro exemplo de edição via NHEJ é a **deleção gênica**, que pode se dar por meio de duas clivagens flanqueando determinada sequência genética que se deseja deletar do genoma (Figura 5). Um exemplo deste tipo de edição é a remoção de sequências do HIV integradas no genoma humano durante a infecção. Note que mesmo que haja *indels* no sítio de religação, isto é irrelevante pois o objetivo maior - a excisão do genoma do HIV - foi alcançado.

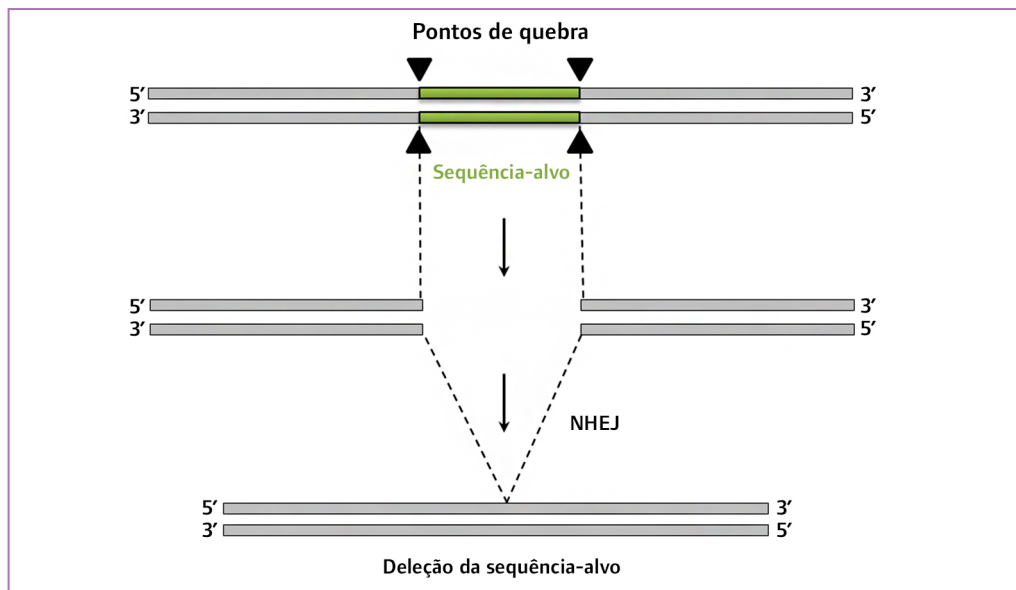


Figura 5. Esquema ilustrativo da deleção gênica. A sequência-alvo em verde (por exemplo, um gene de vírus que se integrou no genoma humano) pode ser removido via duas quebras mediadas por Cas9 e reparo por NHEJ.

VARIAÇÕES DA TÉCNICA

Diversas outras aplicações podem ser alcançadas utilizando variantes da endonuclease Cas9 e do sgRNA. Por exemplo, a dCas9 (*dead Cas9*) é uma variante com **domínios catalíticos** inativos, capaz de ser direcionada ao DNA mas não de clivá-lo. Portanto, ao se fusionar a **proteína fluorescente GFP** à dCas9 (denominada dCas9::GFP) complexada a um sgRNA, é possível visualizar a localização de genes no genoma devido à fluorescência de GFP – esta técnica é denominada de marcação de DNA (Figura 6A). Alternativamente, pode-se utilizar de artifícios que permitem que dCas9::GFP e o sgRNA se liguem a um RNA mensageiro, sendo possível assim rastrear o movimento de qualquer mRNA dentro da célula – tal estratégia é denominada de rastreamento de RNA (Figura 6B).

Um grande potencial emerge a partir da fusão de dCas9 a determinadas proteínas capazes de alterar o nível de compactação da cromatina (abrindo-a ou fechando-a) levando assim ao aumento ou redução, respectivamente, da expressão de genes localizados naquela região do DNA. Assim, dCas9::VP64 com um sgRNA permite a ativação transcricional de genes-alvo (Figura 6C). Este tipo de regulação da expressão gênica é interessante em

situações nas quais se objetiva induzir genes hiporexpressos em determinadas patologias (assim como silenciar genes hiperexpressos em outras doenças).

Outras possibilidades são: clivagem do RNA e o mapeamento gênico. Neste último, a Cas9 realiza quebras no DNA, facilitando assim o processo *crossing over* e, conseqüentemente, a identificação de células com genótipos recombinantes. Baseando-se na frequência destes recombinantes, é possível calcular a distância entre os genes, que configura o mapeamento genético. A todo esse conjunto de variações da técnica denomina-se a caixa de ferramentas da CRISPR (CRISPR *toolbox*), que promete subsidiar uma revolução gigantesca na genética.

CONCLUSÕES

Tecnologias do tipo tesoura molecular para edição gênica já existiam antes do desenvolvimento da técnica de CRISPR, entretanto, o alto custo e a dificuldade envolvida na elaboração/uso dessas nucleases limitavam a acessibilidade e as aplicações dessas ferramentas. Nesse contexto, CRISPR apresenta uma grande vantagem devido à simplicidade, facilidade e ao baixo custo envolvidos na obtenção da enzima Cas9 e no desenho e síntese do sgRNA.

Domínio catalítico - é a região dentro da enzima responsável por realizar a catálise, isto é, a atividade enzimática.

Proteína fluorescente GFP - uma proteína codificada por um gene encontrado em espécies de água-viva como a *Aequorea victoria* e que é capaz de emitir fluorescência verde quando excitada por luz ultravioleta.

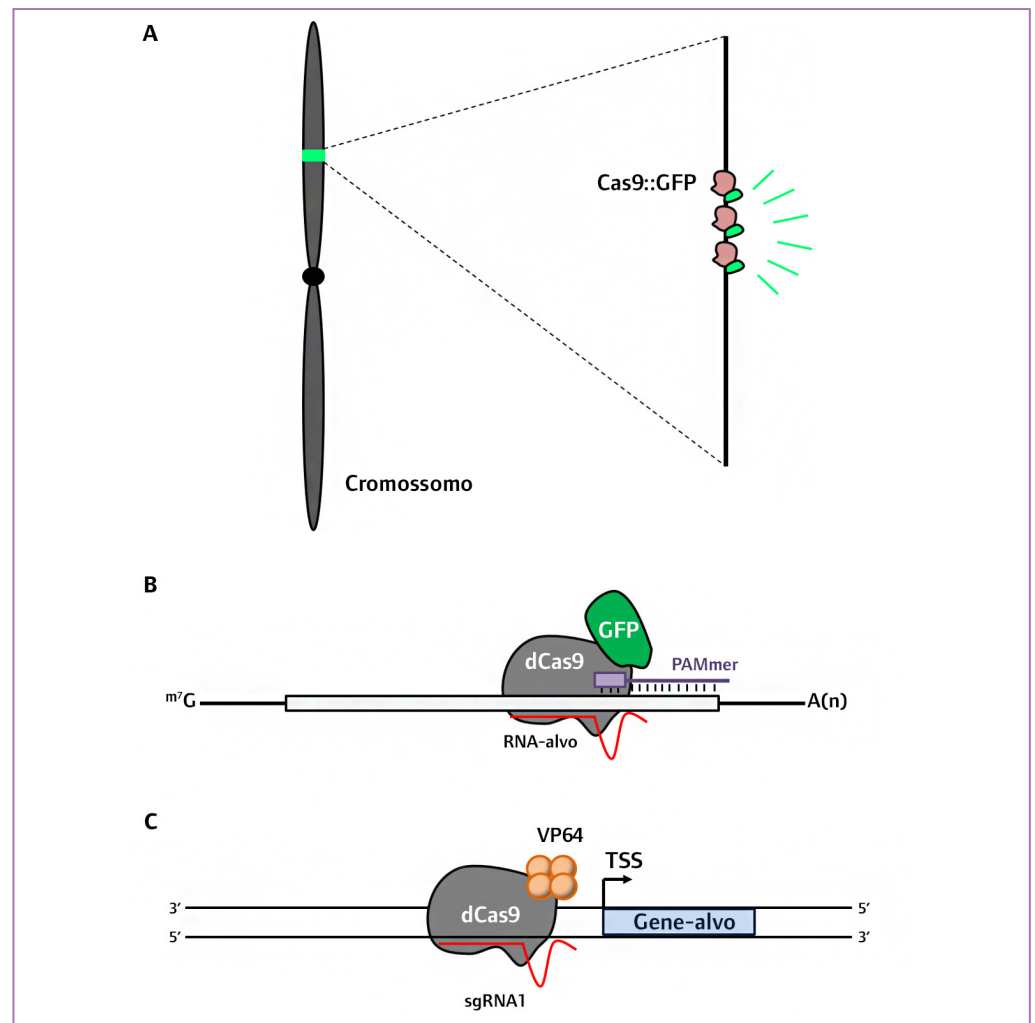


Figura 6.

(A) Uso de dCas9 fusionada à GFP para marcação de DNA, possibilitando assim a identificação da posição de determinado gene no genoma ou (B) para rastreamento de RNA, permitindo assim a acompanhar a movimentação desta molécula ao longo de compartimentos subcelulares. (C) Alternativamente, dCas9 pode ser fusionada a peptídeos ativadores de transcrição (exemplificado pelo tetrâmero VP64) ou a enzimas capazes de alterar o nível de compactação da cromatina, visando a regular a expressão gênica. TSS: sítio de início da transcrição do gene. PAMmer: uma pequena molécula de DNA que age como um artifício utilizado para “enganar” Cas9 e direcioná-la a RNAs.

Surpreendentemente, as aplicações de CRISPR aqui descritas representam uma parcela ínfima de seus usos na pesquisa básica, medicina, agricultura, veterinária, indústria e muitos outros. Devido à sua imensa versatilidade, CRISPR está literalmente reeditando o presente e futuro das ciências da vida.

REFERÊNCIAS

BARRANGOU, R.; FREMAUX, C.; DEVEAU, H.; RICHARDS, M.; BOYAVALL, P.; MOINEAU, S.; ROMERO, D. A.; HORVATH P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, v. 23, n. 315(5819), p. 1709-12, 2007.

MALI, P.; AACH, J.; STRANGES, P. B.; ESVELT, K. M.; MOOSBURNER, M.; KOSURI, S.; YANG, L.; CHURCH, G. M.

CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, v. 31, n. 9, p. 833-8, 2013.

RAN, F. A.; HSU, P. D.; WRIGHT, J.; AGARWALA, V.; SCOTT, D. A.; ZHANG, F. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, v. 8, p. 2281–2308, 2013.

VIEIRA, G. V.; CECÍLIO, N. T.; ARRUDA, L. M.; SALES, K. K. Visão geral do mecanismo básico de ação. In: Pereira, T. C. *Introdução à técnica de CRISPR*. São Carlos: Editora Cubo, 2016. p. 39-47.

WYMAN C.; KANAAR R. DNA double-strand break repair: all's well that ends well. *Annual Review Genetics*, v. 40, p. 363-83, 2006.