

CRISPR *toolbox*: a versatilidade da técnica

Capítulo 3

Caroline Junqueira^{1,a}, Patrícia Renck Nunes^{2,a}, Gabriela Molinari Roberto^{3,b}
e María Sol Brassesco Annichini^{4,b}

¹*Centro de Pesquisas René Rachou/ FIOCRUZ, MG - Brasil*

²*Dept.º de Ciências da Vida, Escola Politécnica Federal de Lausanne - Suíça*

³*Programa de Pós-Graduação em Oncologia Clínica, FMRP USP, SP - Brasil*

⁴*Dept.º de Biologia, FFCLRP USP, SP - Brasil*

^a*Autoria das seções: 1 a 5*

^b*Autoria das seções: 6 e 7*

Estruturação do capítulo

1. Introdução
2. Edição de DNA
 - 2.1 Inativação gênica
 - 2.2 Integração de transgenes
 - 2.3 Substituição alélica
 - 2.4 Deleção de genes inteiros
3. Marcação de DNA
4. Regulação da expressão gênica
 - 4.1 Para indução da expressão gênica
 - 4.2 Para repressão da expressão gênica
5. Clivagem de RNAs
6. Mapeamento de genes
7. Rastreamento de RNAs
8. Conclusões

1. Introdução

Em essência, a versão inicial da técnica de CRISPR promovia uma clivagem sítio dirigida no DNA-alvo, que poderia ser reparado por NHEJ (culminando em mutações pontuais aleatórias na sequência-alvo) ou HDR (resultando em alterações precisas na sequência nucleotídica-alvo).

Nesta seção veremos como algumas alterações no protocolo experimental permitiram tornar a CRISPR mais versátil, expandido seu repertório de atuação além da **edição genética** (*gene editing*) tais como: **marcação de DNA** (*DNA labeling*), **regulação da expressão gênica** (indução e repressão gênicas), **clivagem de RNAs**, **mapeamento de genes** e **rastreamento de RNA** (*RNA tracking*). As aplicações de cada uma delas serão vistas no capítulo seguinte.

2. Edição de DNA

No contexto da técnica de CRISPR, a expressão “**edição genética**” se refere à capacidade de *alterar a sequência nucleotídica do gene-alvo*, em analogia simples à *edição de textos* no computador: alteramos algumas letras (nucleotídeos) ou palavras (genes inteiros) de um documento (material genético), afetando assim seu significado. Quando estas alterações são *muito extensas* (envolvendo regiões cromossômicas) ou *muito numerosas* (englobando múltiplos genes simultaneamente), referimo-nos à “**edição genômica**”.

As diversas formas de edição genética podem ser categorizadas de várias maneiras. Aqui adotaremos uma classificação que é relativamente comum e simples: (i) a inativação gênica, (ii) a integração de transgenes, (iii) a substituição alélica e (iv) a deleção de genes inteiros.

2.1 Inativação gênica

A inativação ou nocauteamento gênico (*gene knockout - KO*) é tipicamente alcançado por meio de mutações que resultam na *perda da atividade do produto final do gene* (seja ele um RNA funcional ou uma proteína) (William S. Klug et al., 2009). A consequência dessas alterações genéticas são alelos **nulos**, *i.e.*, sem função biológica (Benjamin Lewin, 2008). O nocauteamento pode ser alcançado tanto por clivagem por Cas9 seguida por NHEJ quanto pela HDR, sendo esta última mais específica, por induzir alterações precisas.

Como descrito no capítulo anterior, NHEJ é um processo propenso ao erro, promovendo *indels* no sítio de clivagem (figura 1). Esta inserção/deleção de um ou poucos nucleotídeos, quando não múltiplos de três, altera a *fase de leitura* do gene em questão. Isto, por sua vez, quase sempre resulta em um *códon de parada prematuro*, gerando assim proteínas truncadas (figura 1) (Hsu, et al., 2014; Mei et al. 2016).

Apesar de sua vasta aplicação para o nocauteamento de genes, NHEJ não permite um controle da mutação gerada. Eventualmente podem ocorrer inserções/deleções de nucleotídeos em um número múltiplo de três e *em fase* com a matriz de leitura, resultando na simples adição ou deleção de um ou poucos aminoácidos na proteína final. Esta pequena alteração pode ou não ter efeito biológico significativo.

Alternativamente, pode-se obter o nocaute gênico por HDR, que permite modificações precisas no gene-alvo na presença de um DNA doador. Neste cenário, o DNA doador é, em si, a *versão nocauteada* do gene-alvo. Esta versão nocauteada deve ser construída em laboratório, apresentar uma mutação bem definida e será integrada no genoma por recombinação (figura 2). Para tanto, faz-se necessário a cotransfecção de Cas9, gRNA e o DNA doador. Para edições

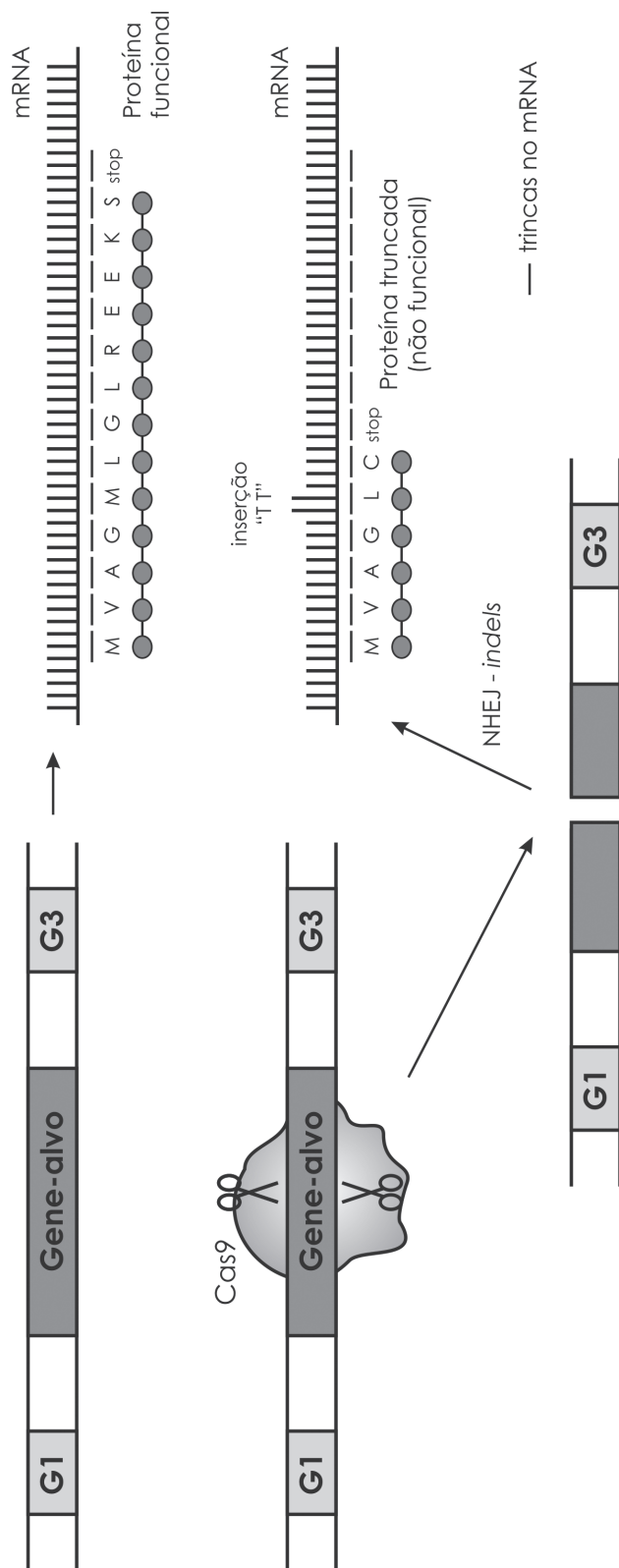


Figura 1. Nocauteamento gênico via CRISPR. A clivagem por NHEJ gera *indels* que, quando não múltiplos de três nucleotídeos, alteram a fase de leitura resultando em códons prematuros de parada. Este polipeptídeo truncado frequentemente não é funcional. G1 e G3: “gene 1” e “gene 3”, para referência. Imagem: Pereira, TC (2016).

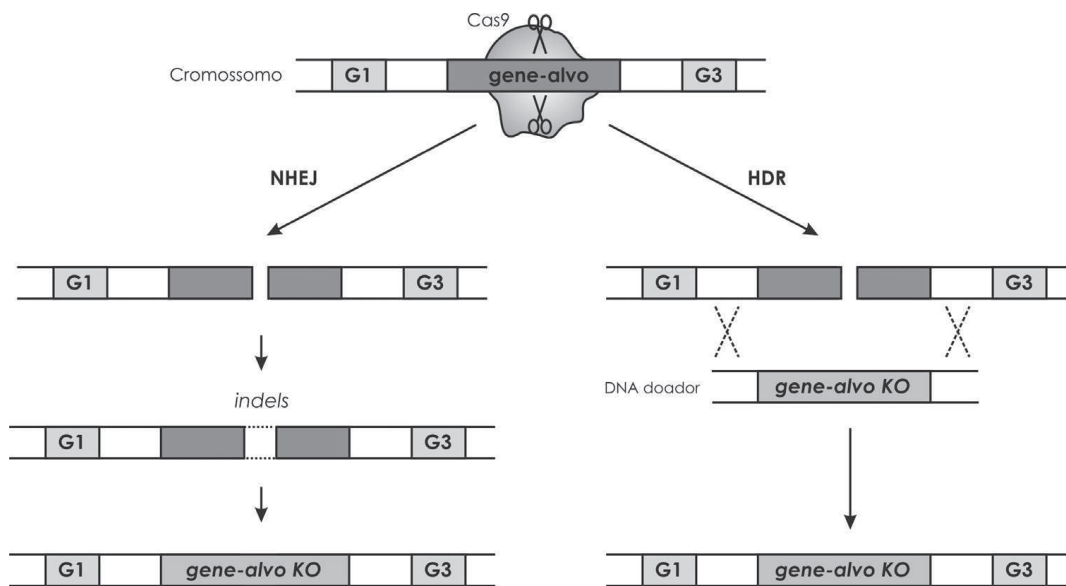


Figura 2. Mecanismos de nocauteamento via CRISPR. Após a clivagem por Cas9, NHEJ repara a dupla fita de DNA, adicionando ou deletando nucleotídeos (*indels*). Esta mutação pontual tem como consequência o aparecimento de um códon prematuro de parada, levando ao nocauteamento do gene (KO). Alternativamente, HDR repara a dupla fita de DNA por recombinação homóloga na presença de um DNA doador com extremidades homólogas ao DNA original (versão nocauteada do gene), usando-o como molde para recombinação. Imagem: Pereira, TC (2016).

curtas, de até 50 pares de bases (bp), oligonucleotídeos de fita simples podem ser utilizados como DNA doador, sendo uma opção simplificada (vide capítulo 9 para detalhes experimentais).

Estratégias mais elaboradas podem ser feitas para o nocauteamento via HDR. Por exemplo, o DNA doador pode ser *um gene marcador de resistência a antibióticos (e.g., Neo^R)* ou *um gene codificador de uma proteína fluorescente (e.g., GFP)* com extremidades homólogas ao gene-alvo. Como resultado da HDR temos uma inativação insercional, *i.e.*, a inserção de Neo^R (ou GFP) *dentro* do gene-alvo resultando em seu nocauteamento. Adicionalmente, a seleção das células modificadas por esta abordagem torna-se um procedimento extremamente fácil.

Contudo, ao contrário de NHEJ, HDR é geralmente ativa apenas em células em divisão. Isto ocorre porque as células possuem mecanismos de inibição de HDR durante as fases iniciais do ciclo celular, a fim de evitar fusão de telômeros. Adicionalmente, a eficácia de HDR pode variar amplamente dependendo do tipo celular, do local de edição e do modelo de reparo. Portanto, apesar de NHEJ não ser o modelo de nocauteamento genético mais preciso, é por muitas vezes a técnica de escolha devido a dificuldades metodológicas de implementação de HDR (RAN et al., 2014).

Em síntese, a inativação gênica (KO) pela técnica de CRISPR pode ser alcançada via NHEJ, gerando *indels* que frequentemente resultam em proteínas truncadas. Alternativamente, o KO pode ser obtido via HDR, utilizando um DNA doador, levando a uma inativação de natureza predeterminada. Todas estas modificações resultam em alelos nulos.

2.2 Integração de transgenes

Apesar de existirem diversas estratégias para a integração de sequências exógenas (*e.g.*, transgenes), geralmente elas não são sítio dirigidas, *i.e.*, o evento ocorre ao acaso no

genoma. Isto pode ser problemático se a inserção do transgene ocorrer *dentro* de um gene endógeno importante, inadvertidamente inativando-o.

A estratégia de *knock-in* (KI) foi desenvolvida para sanar esta problemática, permitindo a integração de transgenes em regiões predeterminadas do genoma, evitando-se assim efeitos indesejados (Tom Strachan e Andrew Read, 2013). Isto é possível porque a KI é um processo baseado em recombinação homóloga (HDR), muito semelhante ao visto para KO. A única diferença é que o DNA doador, neste caso, é um transgene apresentando extremidades com homologia a alguma região intergênica do genoma (*i.e.*, uma região sem genes) (figura 2).

2.3 Substituição alélica

Adicionalmente, a KI permite outro efeito: a substituição alélica, mencionada no capítulo anterior. Este processo visa a troca da sequência nucleotídica do alelo endógeno por outra sequência exógena ligeiramente diferente, portando SNPs ou pequenas mutações (Zhang et al., 2014). O princípio geral é o mesmo: a administração de Cas9, gRNA e um DNA doador (que é, em si, o novo alelo) permite a recombinação homóloga, resultando na substituição alélica.

2.4 Deleção de genes inteiros

Em algumas situações, a *deleção da sequência completa do gene* é mais interessante do que gerar mutações pontuais nele (por NHEJ ou HDR). Isto porque muitas mutações de ponto eventualmente geram alelos **não nulos**, *i.e.*, que permitem uma atividade ainda que parcial do gene-alvo. Entretanto, a deleção completa de um gene resultará inexoravelmente em uma mutação nula. Uma célula cujos dois alelos do “gene A” foram completamente deletados apresenta um genótipo Δ “gene A”/ Δ “gene A” (Δ é símbolo genético para deleção).

A deleção de genes inteiros pela técnica de CRISPR pode ocorrer tanto por *uma única* quebra da dupla fita de DNA, quanto por *duas* quebras da dupla fita (Sander e Joung, 2014; Mei et al. 2016).

Na primeira, o mecanismo de reparo será sempre HDR, sendo similar ao KO e KI induzidos por recombinação homóloga. Neste caso, porém, vale ressaltar que o DNA doador é composto por uma sequência sem o gene de estudo, apenas com as extremidades homólogas ao cromossomo (figura 3A).

Na segunda, diferentemente das edições gênicas anteriormente descritas, a deleção ocorre por intermédio de *duas* quebras na dupla fita de DNA. Para isso, empregam-se duas endonucleases Cas9 e dois gRNAs com sequências distintas. Na ausência de uma sequência doadora ocorrerá NHEJ, via junção de extremidades que não envolvem o gene-alvo (figura 3B). Contudo, na presença de um DNA doador (de mesma natureza que no exemplo A), ocorrerá o HDR.

3. Marcação de DNA

A técnica de *DNA labeling* associada à CRISPR consiste na marcação *in situ* de locos gênicos com fluorocromos, que possibilitam a visualização tridimensional do loco em células fixadas e também vivas (Anton et al., 2014) sem a necessidade de desnaturação do DNA (Ma et. al, 2015). Técnicas que possibilitam a localização espacial de locos são importantes

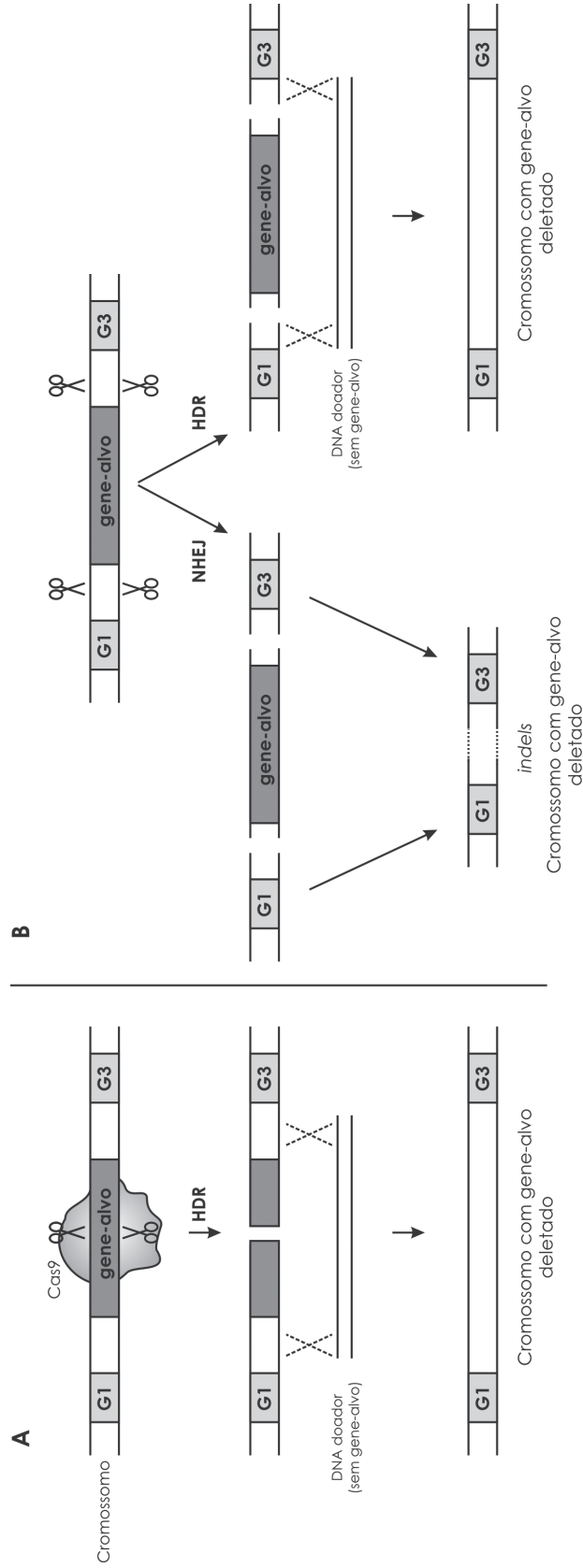


Figura 3. Deleção de genes inteiros. (A) Uma única quebra da dupla fita de DNA, sob o reparo por HDR, leva à deleção do gene inteiro se o DNA doador for composto por uma sequência com homologia com o cromossomo, mas sem o gene-alvo. (B) Quebras duplas na dupla fita de DNA (mediada por dois gRNAs distintos) também podem levar à deleção do gene, por duas maneiras distintas. A esquerda, via NHEJ, a deleção ocorre devido à simples junção de extremidades sem o envolvimento do gene-alvo. A direita, via HDR, ocorre o reparo por recombinação homóloga, tendo um DNA doador (o mesmo do exemplo "A") como molde para a junção. G1 e G3: "gene 1" e "gene 3", para referência. Imagem: Pereira, TC (2016).

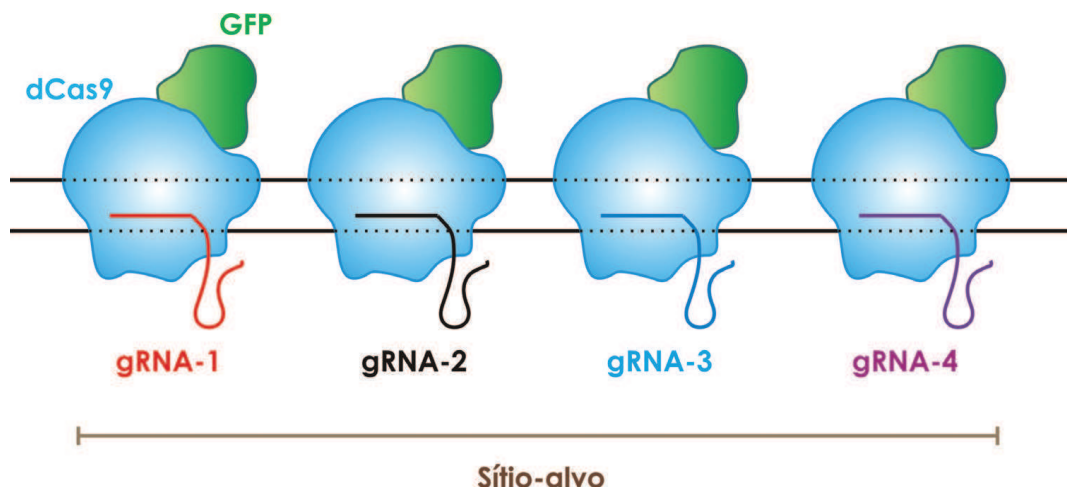


Figura 4. DNA labeling por dCas9. A hibridização *in situ* de locos gênicos ocorre pela ligação de diversas *dead* Cas9 à dupla fita de DNA. gRNA com especificidade para diferentes sítios de um mesmo loco direcionam o local alvo de ligação para dCas9. O acúmulo de dCas9, em associação com proteínas fluorescentes, leva ao aumento pontual de fluorescência, detectável por microscopia de fluorescência de alta resolução. Imagem: Pereira, TC (2016).

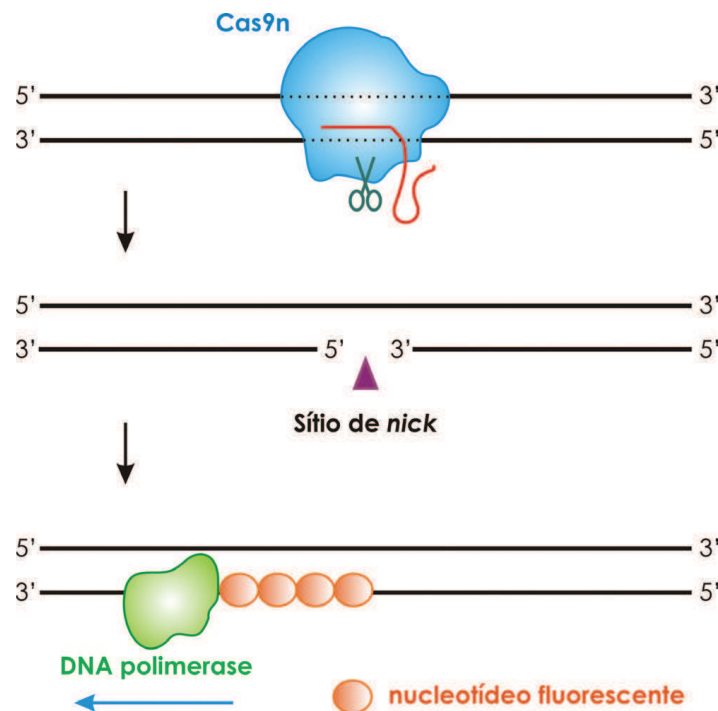


Figura 5. DNA labeling por Cas9n. A marcação de regiões específicas do genoma por nucleotídeos fluorescentes ocorre por intermédio de uma Cas9 modificada (Cas9n), que realiza a quebra apenas na fita 3'-5' do DNA. Em seguida, pela ação da Taq DNA polimerase, nucleotídeos fluorescentes são adicionados adjacentes à quebra da fita dupla, marcando portanto a região específica para o gRNA. Imagem: Pereira, TC (2016).

para o entendimento dos mecanismos de regulação gênica espaço-temporal. Adicionalmente, consiste em uma poderosa ferramenta de diagnóstico genético *in situ*.

Por diversas décadas, técnicas de hibridização *in situ* associadas a ácidos nucleicos fluorescentes (FISH) permitiram a localização de locos gênicos. Porém, devido à necessidade de fixação da célula e desnaturação do DNA, nunca foi possível precisar tais informações em células vivas (Ma et al., 2015).

A presente metodologia consiste na utilização de uma nuclease Cas9 cataliticamente inativa, ou seja, incapaz de clivar a dupla fita de DNA mesmo estando associada gRNA. Devido a esta característica, ela é denominada **dead Cas9 (dCas9)**. Ao gene da dCas9 associa-se uma sequência codificadora de uma proteína de marcação, tal como a proteína verde fluorescente (GFP), proteína vermelha fluorescente (RFP), dentre outras. Portanto, ao ser expressa, a dCas9 apresenta-se indiretamente fluorescente.

Por intermédio da associação de dCas9 a diversos gRNAs específicos para diferentes sítios do mesmo loco, há uma concentração de dCas9 fluorescentes na região de interesse do genoma (figura 4). Conseqüentemente, por meio de microscopia de fluorescência de alta resolução é possível observar pontos correspondentes ao loco gênico.

Com o avanço contínuo das diversas aplicações do sistema CRISPR nas ciências biológicas, hoje já existem variações da técnica de *DNA labeling*. Dentre as mais relevantes estão a fusão de proteínas fluorescentes, tanto com a dCas9 quanto com o gRNA, desta forma disponibilizando um painel de cores diferentes, que devem se sobrepor no momento de uma ligação dCas9/gRNA, assim como reduzir o *background* de ligações inespecíficas. Também como variáveis da técnica ressalta-se tanto a transfecção dos plasmídeos de expressão de dCas9 e gRNA em células vivas, quanto a metodologia de marcação de células fixadas (Deng et al., 2015). Finalmente, uma metodologia inovadora propõe a adição, por meio de uma DNA polimerase, de nucleotídeos fluorescentes adjacentes ao sítio de corte de fita simples pela Cas9 nickase (Cas9n), desta forma amplificando o sinal de fluorescência e possibilitando uma análise com maior cobertura do genoma (McCaffrey et al., 2015) (figura 5).

4. Regulação da expressão gênica

Uma das primeiras aplicações provenientes da criação dCas9 foi a regulação da expressão gênica (Gilbert et al., 2013). Pesquisadores fusionaram a dCas9 a proteínas reguladoras da transcrição gênica, transformando-a em uma *proteína de transporte*, capaz de levar esses elementos para uma sequência de DNA específica graças ao gRNA. Dependendo da classe de proteína reguladora conduzida pela dCas9, pode-se *induzir* ou *reprimir* a expressão do gene, modificando-se ou não a cromatina ao seu redor. Nesta seção vamos apresentar algumas variantes da dCas9 que podem ser usadas com esse propósito, assim como as vantagens e desvantagens de cada abordagem.

4.1 Para indução da expressão gênica

Induzir a expressão de um gene usando a dCas9 é talvez uma das aplicações mais desafiadoras desse sistema. Isso porque o posicionamento incorreto da dCas9 no promotor ou em suas sequências regulatórias (*enhancers*) pode impedir a ligação de fatores de transcrição endógenos e do complexo da própria RNA polimerase ao DNA. Além disso, altos níveis

de compactação da cromatina, fenômeno típico de genes silenciados, pode dificultar o posicionamento do gRNA.

A versão da dCas9 mais utilizada para indução de transcrição é a dCas9-VP64 (Maeder et al., 2013). Nessa variante, a dCas9 é fusionada a um *tetrâmero* da proteína VP16* (figura 6A), um fator de transcrição do *Herpesvirus simplex* tipo 1 usado para transcrição robusta e imediata de genes no começo da infecção. Embora a dCas9-VP64 seja suficiente para indução de transcrição da maior parte dos genes, também é possível utilizar a versão com 10 domínios do VP16 (dCas9-VP160) caso a indução seja insuficiente. Da mesma maneira, a dCas9-VP48 (com 3 domínios do VP16) pode ser usada para uma indução mais modesta do gene de interesse (Cheng et al., 2013).

Outro aspecto a ser observado para indução de transcrição é a quantidade e o posicionamento dos gRNAs. Salvo raras exceções, múltiplos gRNAs devem ser usados simultaneamente para que qualquer aumento de transcrição seja detectado (Maeder et al., 2013). Geralmente, 3 ou 4 gRNAs posicionados na sequência a montante (extremidade 5') do sítio de início de transcrição (do inglês, *transcription start site*, ou TSS) são suficientes para gerar transcrição. Entretanto, em alguns casos 10 gRNAs ou mais podem ser necessários.

Os gRNAs devem ser desenhados de forma que se liguem a sequências próximas ao promotor do gene de interesse. Contudo, os gRNAs devem evitar as regiões utilizadas para a ligação do complexo da RNA polimerase. Os gRNAs mais bem-sucedidos são complementares à fita de DNA que servirá de molde para o RNA, com algumas exceções. Dessa forma, a posição do gRNA vai variar de gene para gene, e cabe ao pesquisador encontrar aqueles adequados para o seu experimento. Por exemplo, para ativar o gene *VEGFA* é preciso colocar os gRNAs entre 450 e 600 pares de bases a montante do TSS. Já para o gene *NFT3*, os melhores gRNAs estão a 135 pares de bases do TSS (Maeder et al., 2013).

Em casos em que a dCas9-VP64 não é suficiente para induzir transcrição, alternativas como a dCas9-SunTag_{10x}-VP64 (figura 6B) podem ser utilizadas (Tanenbaum et al., 2014). O sistema *SunTag* foi criado em 2014 para a amplificação de sinais de fluorescência emitidos por uma proteína. A SunTag é uma sequência de peptídeos que pode ser adicionada à proteína de interesse. Cada peptídeo da SunTag é reconhecido por um fragmento de anticorpo de cadeia simples (scFv, do inglês *single-chain variable fragment*) fusionado a outras proteínas. No caso da dCas9-SunTag_{10x}-VP64, a dCas9 apresenta 10 cópias do peptídeo, que são reconhecidos por 10 anticorpos scFv fundidos a uma VP64 cada. Desta forma, para cada dCas9-SunTag_{10x} presentes no DNA, 10 fatores de transcrição são recrutados. O resultado é um aumento robusto na transcrição do gene de interesse. Em vários casos é possível inclusive induzir um aumento de 3 a 40 vezes na transcrição de um gene com um único gRNA. A desvantagem dessa técnica é que a célula precisa apresentar não somente o gRNA e a dCas9-SunTag_{10x}, mas também o anticorpo scFv contra a sequência de peptídeos. A sua maior vantagem está na possibilidade de se usar somente um gRNA, em vez de 3 ou até 10 gRNAs.

Uma estratégia diferente para indução de transcrição baseia-se em modificar a cromatina onde o TSS se situa. Genes silenciados possuem uma assinatura heterocromática, que inclui metilação e deacetilação de histonas associadas. Por sua vez, genes ativos são caracterizados por pouca metilação e muita acetilação. Por essa razão, pesquisadores começaram a fusionar a dCas9 a enzimas que induzem acetilação de histonas. Essa abordagem é especialmente

* **Nota:** Veja que os multímeros de VP16 são nomeados de acordo com os valores múltiplos correspondentes. Por exemplo, o trímero de VP16 é chamado VP48 (pois $16 \times 3 = 48$); o tetrâmero, VP64; o decâmero, VP160.

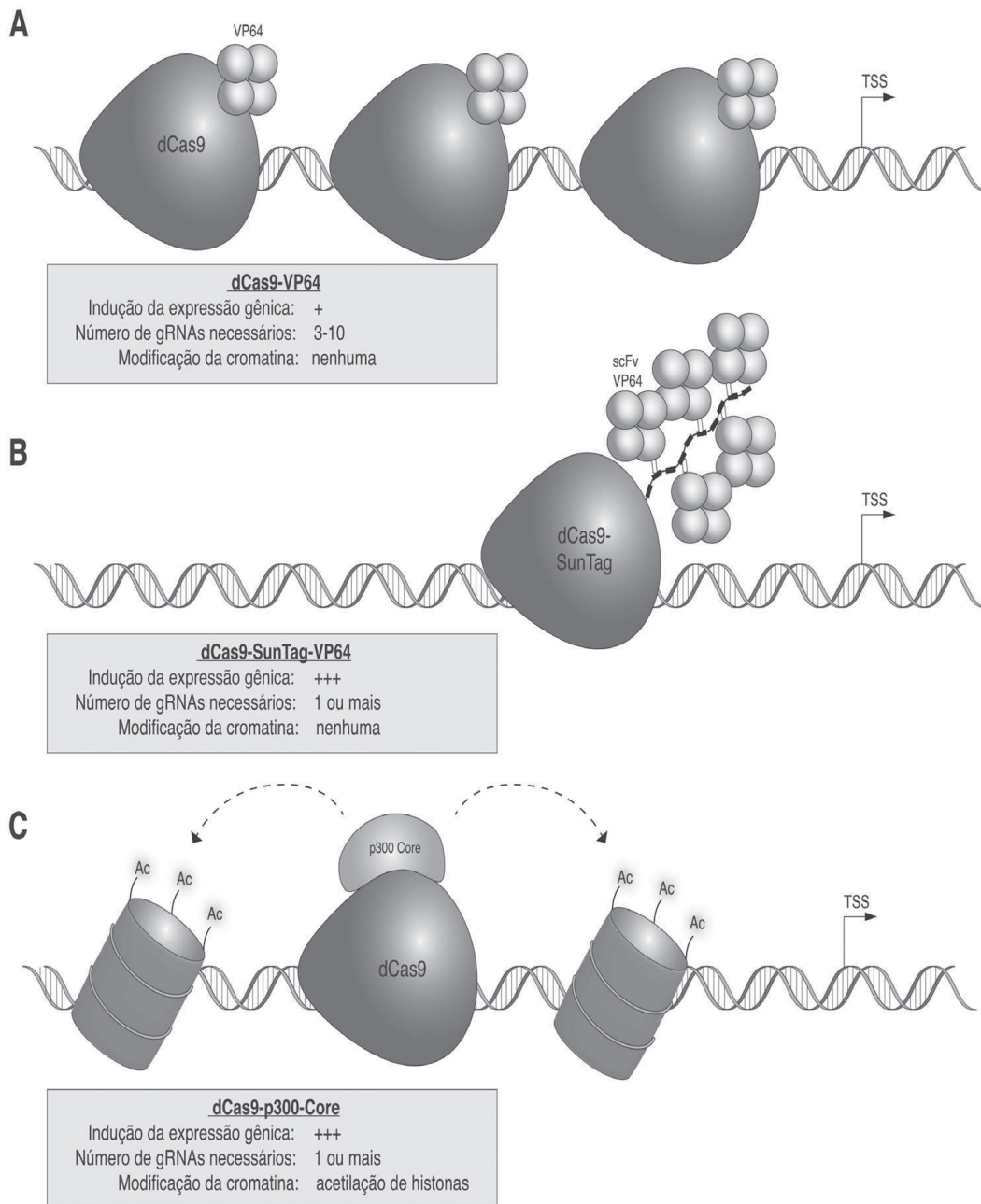


Figura 6. Modelos de uso da dCas9 para indução da expressão gênica. A) dCas9-VP64: a dCas9 é fusionada a 4 cópias do fator de transcrição VP16. Múltiplas cópias da dCas9-VP64 devem ser recrutadas por vários gRNAs às regiões próximas ao promotor do gene de interesse. B) dCas9-SunTag-VP64: a dCas9 é fusionada a uma cadeia de peptídeos (representados como retângulos pretos) e levada ao promotor do gene de interesse por um ou mais gRNAs. A cadeia de peptídeos é reconhecida por anticorpos de cadeia simples fundidos ao VP64. C) dCas9-p300core: variante que, através do núcleo catalítico do p300, induz acetilação de histonas e abertura da cromatina próxima ao TSS (sítio de iniciação da transcrição) do gene de interesse. Imagem: Nunes, PR (2016).

interessante para ativação transcricional de genes presentes em regiões de cromatina muito condensada, onde até a ligação de diversos gRNAs pode ser difícil.

Uma das enzimas acetiltransferases mais conservadas em eucariotos é a p300. O núcleo da p300 (p300^{core}) pode ser fusionado à dCas9 e levada, por meio de um gRNA, até a região regulatória (*enhancer*) do gene de interesse (figura 6C). Uma vez posicionada, a p300^{core} adiciona grupos acetil às histonas vizinhas (principalmente na histona H3, lisina 27) e assim induz aberturas na heterocromatina e o recrutamento de fatores de transcrição (Tsankov et al., 2015). Além disso, a p300^{core} também é capaz de recrutar a RNA polimerase II para formar o complexo de iniciação de transcrição (Cho et al., 1998).

A dCas9-p300^{core} induz transcrição em níveis maiores do que a dCas9-VP64 e, assim como a dCas9-SunTag_{10x}-VP64, pode ser usada com um único gRNA. A vantagem da dCas9-p300^{core} sobre a versão SunTag é que a célula só precisa apresentar a dCas9-p300^{core} e o gRNA, sem necessidade adicional dos anticorpos scFv contra a cadeia de peptídeos.

Continuamente, sistemas cada vez mais complexos e eficientes para indução da expressão gênica podem ser encontrados. A dCas9 foi recentemente fundida a três ativadores de transcrição ao mesmo tempo, formando a dCas9-VP64-p65-Rta. A nova versão promete uma indução 40 a 100 vezes mais forte do que a dCas9-VP64 original (Chavez et al., 2015). Ao mesmo tempo, o próprio gRNA começa a ser modificado para formar estruturas que podem recrutar indiretamente fatores de transcrição. O recrutamento é feito por meio de outras proteínas (fundidas aos fatores de transcrição de interesse) que reconhecem essas estruturas formadas pelo RNA (Konermann et al., 2014).

A indução de transcrição gênica a partir dos próprios promotores endógenos dos genes é uma das aplicações mais poderosas da dCas9. Isto porque diferentes isoformas do transcrito, geradas pelo processo de *splicing*, não podem ser reproduzidas com outras técnicas de superexpressão do DNA complementar (cDNA). Além disso, tornou-se mais fácil e fisiológico induzir a expressão de RNAs longos e não codificantes (lncRNAs) e de RNAs que possuem diferentes TSS. A expressão de múltiplos gRNAs ao mesmo tempo possibilita ainda a indução de vários genes simultaneamente na mesma célula, no processo chamado *multiplexing* (Cheng et al., 2013). Através do uso das variantes da dCas9, também pode-se atingir diferentes níveis de indução de expressão gênica. Cabe ao pesquisador determinar os melhores gRNAs a serem utilizados e a dCas9 mais apropriada para o seu objetivo.

4.2 Para repressão da expressão gênica

Uma das melhores maneiras de se determinar a função de uma proteína ou de um RNA é basicamente eliminá-los da célula e observar quais processos são afetados por sua ausência. Por muitos anos, a técnica de interferência por RNA (RNAi) foi usada com esse propósito. Entretanto, RNAi é um método potencialmente caro e pouco efetivo contra certos alvos, principalmente lncRNAs. Para contornar esse problema, a técnica de *Interferência via CRISPR (CRISPRi)* foi introduzida (figura 7).

Usar a dCas9 para reprimir a expressão de um gene é relativamente mais fácil do que usá-la para induzir transcrição. Isso porque um único gRNA bem posicionado já é suficiente para supressão de até 80% da transcrição (Gilbert et al., 2013). Entretanto, se o uso de somente um gRNA não alcançar a repressão gênica desejada, múltiplos gRNAs podem ser usados.

A dCas9 pode ser utilizada sozinha ou em sua forma fusionada ao repressor de transcrição KRAB (do inglês, *Krüppel-associated box domain*). Em ambos os casos, a dCas9

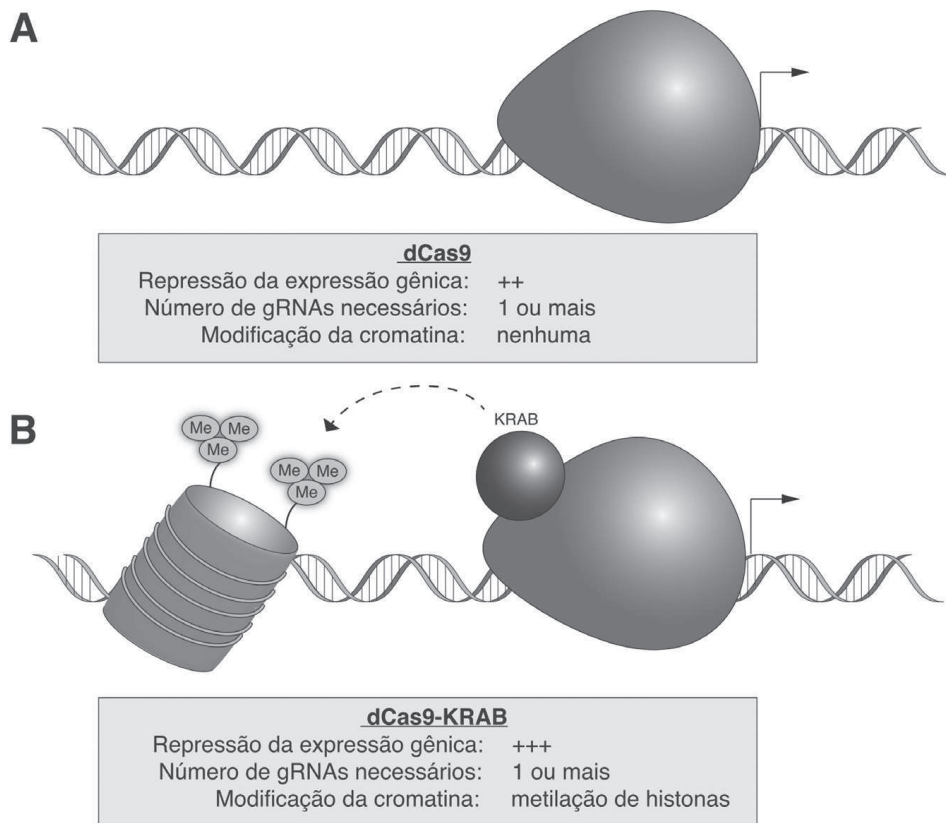


Figura 7. Modelos de uso do dCas9 para repressão da expressão gênica. A) dCas9: esta proteína sozinha pode ser usada para impedir a ligação de fatores de transcrição e do complexo da RNA polimerase II ao DNA. Para isso, deve-se posicioná-la perto do TSS (sítio de iniciação de transcrição) do gene de interesse. B) dCas9-KRAB: essa variante, além de servir de barreira física contra a transcrição, também modifica a cromatina próxima do TSS de maneira a torná-la mais heterocromática e menos acessível à maquinaria de transcrição. Imagem: Nunes, PR (2016).

deve ser direcionada ao promotor de um gene ou ao seu TSS. A dCas9 liga-se com grande afinidade ao DNA através do gRNA e bloqueia a ligação de fatores de transcrição e do complexo da própria RNA polimerase, impedindo a transcrição (Gilbert et al., 2013).

Já a dCas9-KRAB, além de impedir fisicamente a ligação de outras proteínas ao DNA, também promove a formação de heterocromatina ao seu redor. O domínio KRAB recruta indiretamente proteínas da família da HP1 (do inglês, *heterochromatin protein 1*) que, por sua vez, interagem com histonas metiltransferases, induzindo metilação da histona H3, por exemplo.

O local de ancoragem (posicionamento) da dCas9 (com ou sem KRAB) também é um aspecto essencial para repressão eficiente da transcrição. O gRNA deve ser dirigido à região do TSS, ou até 100 pares de bases antes ou depois do mesmo (Gilbert et al., 2013). Em geral, quanto mais longe do TSS estiver a dCas9, menos efetiva a repressão será. A exceção acontece quando a dCas9-KRAB é posicionada de forma a silenciar *enhancers* distantes (a dezenas de milhares de pares de bases do TSS) (Thakore et al., 2015). Tal estratégia pode ser usada para mapear *enhancers* de genes específicos, mas a repressão em si é menos robusta do que aquela obtida com posicionamento da dCas9-KRAB próxima ao TSS.

Assim como na indução da expressão gênica, múltiplos gRNAs podem ser usados simultaneamente em combinação com a dCas9 ou dCas9-KRAB de forma a reprimir a expressão de vários genes ao mesmo tempo (*multiplexing*), o que não é facilmente feito com técnicas de RNAi. A CRISPRi também possibilita a repressão da expressão de proteínas na fonte, por interferir diretamente com a produção do mRNA (*i.e.*, silenciamento gênico transcricional). Da mesma forma, tornou-se mais fácil controlar os níveis de lncRNAs e de outros RNAs contra os quais técnicas tradicionais de RNAi não são eficientes. Embora seja uma técnica simples, rápida e robusta, o pesquisador deve determinar a melhor posição da dCas9 ou da dCas9-KRAB e se mais de um gRNA é necessário.

5. Clivagem de RNAs

O domínio catalítico da Cas9 que corta a fita de DNA complementar ao gRNA (domínio chamado HNH) tem a mesma estrutura de alguns domínios que cortam RNA. A partir dessa observação, a possibilidade de se usar a Cas9 para clivar RNAs começou a ser estudada. O maior problema encontrado é que a Cas9 se liga com pouca afinidade ao RNA de fita simples (ssRNA), e não se liga a níveis detectáveis ao RNA de fita dupla (dsRNA). Para induzir a associação da Cas9 com o RNA é necessário prover a célula com sequências chamadas **PAMmers** (figura 8): cadeias de DNA em fita simples (ssDNA) formadas por 10 ou até 20 nucleotídeos e cujas extremidades 5' contêm a sequência PAM (5'-NGG-3', ou somente 5'-GG-3') (O'Connell et al., 2014). A função dos PAMmers é se parear com o RNA-alvo e apresentar a sequência PAM à Cas9. Isso induz uma mudança de conformação da Cas9 e permite que seus sítios catalíticos clivem o RNA.

Essa estratégia leva à clivagem eficiente do RNA de interesse, mas também pode levar à associação do complexo Cas9/gRNA com outros RNAs contendo PAMmers. Para reduzir a associação da Cas9/gRNA com alvos indesejados, os PAMmers devem ser estendidos em até 8 bases na extremidade 5' (figura 8B). Nesse caso, parte do PAMmer fica alinhado à sequência reconhecida pelo gRNA, impedindo a ligação do gRNA ao RNA sem que haja dissociação de parte do complexo PAMmer-RNA. Essa dissociação necessita da energia que é normalmente liberada pelo pareamento de bases entre o gRNA e o RNA-alvo. Assim, a ligação do gRNA a um RNA que não tem uma sequência compatível com o gRNA torna-se energeticamente desfavorável, reduzindo a afinidade do complexo Cas9/gRNA por outros RNAs.

Outro problema em potencial é a provável clivagem do DNA que deu origem ao RNA-alvo pela Cas9, já que ambos possuem a mesma sequência. Para contornar este problema, os PAMmers devem se direcionar à parte do RNA onde não haja uma sequência PAM presente, criando um desalinhamento de três pares de base (figura 8C). A sequência PAM presente no PAMmer, mesmo não complementar ao RNA-alvo, é suficiente para recrutar e ativar a Cas9, enquanto a ausência da PAM no genoma impede que ele seja clivado pela Cas9 (O'Connell et al., 2014). Tal estratégia também permite a clivagem de RNAs que eventualmente não possuam uma sequência PAM.

Em resumo, PAMmers simples (figura 8A) são os mais eficientes em induzir clivagem do RNA-alvo, porém existe a possibilidade de clivagem de outros RNAs contendo PAMmers e do DNA. PAMmers com extensão na extremidade 5' (figura 8B) reduzem significativamente a clivagem de outros RNAs, mas também podem diminuir em até 10 vezes a clivagem do RNA-alvo. PAMmers com sequência PAM não complementar ao RNA-alvo (figura 8C) eliminam a

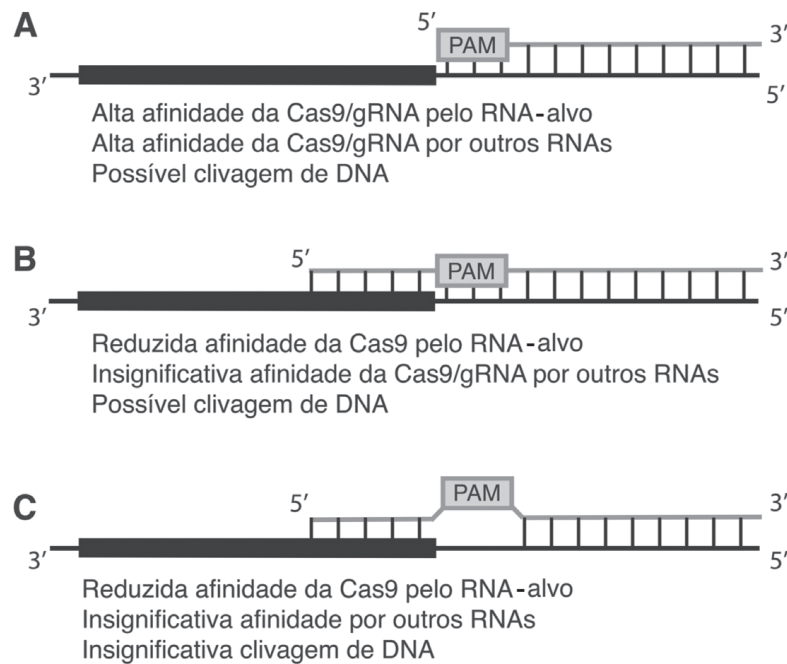


Figura 8. Clivagem de RNA pela Cas9 e os diferentes PAMmers que podem ser utilizados. O RNA é mostrado em preto (linha inferior) e o retângulo preto representa a sequência do RNA reconhecida pelo gRNA. Os PAMmers (ssDNA) são mostrados em cinza (linha superior). A) PAMmer simples, com a sequência PAM no seu final 5'. Este PAMmer tem alta afinidade pelo RNA-alvo, mas também se liga a outros RNAs e ao DNA. B) O PAMmer com uma extensão na sua extremidade 5' tem maior afinidade pelo RNA-alvo e menor afinidade por outros RNAs. Entretanto, esses PAMmers dificultam a interação do complexo Cas9/gRNA com o RNA, já que a ligação entre o PAMmer e o RNA-alvo deve ser desfeita antes do alinhamento do gRNA com o RNA de interesse. C) O PAMmer com uma sequência PAM não complementar ao RNA-alvo tem menor afinidade pelo mesmo, mas não leva à clivagem do DNA que deu origem ao RNA de interesse. Imagem: Nunes, PR (2016).

clivagem do DNA. Todas as modificações do PAMmers têm vantagens e desvantagens. Cabe ao pesquisador escolher a abordagem de acordo com a eficiência e especificidade desejadas.

A mesma técnica pode ser usada também para precipitação de RNAs específicos. Para isso, deve-se usar a *dead* Cas9 em vez da Cas9 cataliticamente ativa e utilizar PAMmers feitos de oligômeros modificados com 2' O-Metil, o que impede que RNases degradem a duplec RNA:DNA (O'Connell et al., 2014). A dCas9 nesse caso pode ser precipitada com um anticorpo contra a própria dCas9 ou contra uma *tag* adicionada a ela.

O uso da Cas9/dCas9 para clivar ou precipitar RNAs específicos pode transformar o modo como estudamos estas moléculas. A técnica promete a clivagem específica dos RNAs selecionados e a possibilidade de se isolar um RNA sem que uma *tag* tenha sido inserida nele previamente. Além disso, seu baixo custo e facilidade interessa toda a comunidade científica.

6. Mapeamento de genes

Um dos principais objetivos da genética moderna consiste na localização e identificação de sequências de DNA responsáveis por determinadas variações fenotípicas. Em qual(ais) cromossomo(s) estão os genes responsáveis pela cor dos olhos? E pelo tamanho das asas em

moscas de fruta? E por certas deficiências como hemofilia ou talassemia? Algumas destas perguntas puderam ser respondidas através dos estudos iniciais de Morgan e de seu aluno Sturtevant, como a herança de genes localizados no mesmo cromossomo. Nesses estudos, a taxa de recombinação gênica meiótica entres dois genes era correspondente à distância destes no cromossomo (Snustad e Simmons, 2012).

A partir disso, o mapeamento genético vem sendo realizado através do estudo da cosegregação (herança conjunta) de determinados marcadores genéticos com o fenótipo de interesse em um amplo grupo de indivíduos, para assim determinar as regiões específicas do genoma que influenciam este fenótipo (Snustad e Simmons, 2012). Entretanto, esta forma de localização depende da taxa de recombinação meiótica na região entre os diferentes marcadores genéticos que, em geral, é muito baixa. Dessa forma, na prática, a localização específica de genes ou variantes internas de cada gene, como polimorfismos, torna-se inviável, visto que exigem construções de painéis imensos de indivíduos ou estudos de várias gerações.

Um novo método também baseado em eventos de recombinação direcionados a regiões de interesse vem sendo desenvolvido através do estudo de recombinações mitóticas, permitindo o aumento da resolução do mapeamento (Yin e Petes, 2013). A recombinação mitótica ocorre de forma rara durante o reparo de quebras de fita dupla por recombinação homóloga, gerando um cromossomo recombinante. Supondo uma célula heterozigótica, este processo, conhecido por perda da heterozigosidade, resultaria em células filhas cujo genótipo é completamente homozigoto do ponto de quebra e reparo até o telômero; no restante do cromossomo, ele permanece semelhante à célula-mãe – heterozigoto. Assim, para construção do mapa genômico utilizam-se indivíduos com perda da heterozigosidade em diferentes pontos do genoma associados com a variação do fenótipo estudado (Henson et al., 1991). Na prática, estes estudos também são limitados devido à baixa frequência em que esses eventos de recombinação mitótica ocorrem naturalmente.

É neste contexto que o sistema CRISPR-Cas9 pode ser aplicado no processo de mapeamento genético. Esse sistema facilita a construção de painéis de indivíduos com diferentes níveis de perda de heterozigosidade através da indução de eventos de recombinação mitótica em locais específicos (Sadhu et al., 2016) (figura 9). Como já elucidado no capítulo 2, a endonuclease Cas9 é capaz de causar quebra dupla na fita de DNA em uma sequência específica de acordo com o RNA-guia ligado a ela. No entanto, para que Cas9 se ligue e clive o DNA com sucesso é necessária uma sequência PAM localizada na fita complementar à sequência-alvo (Doudna e Charpentier, 2014).

Para a ocorrência da perda da heterozigosidade em indivíduos heterozigotos como explicado acima, é necessário a quebra de dupla fita em apenas um dos cromossomos homólogos. Mas como garantir que CRISPR atue em apenas um dos cromossomos? Este objetivo é alcançado com a utilização de sítios PAM polimórficos, garantindo que apenas um dos cromossomos seja afetado pela quebra da fita dupla (Sadhu et al., 2016).

Recentemente, a aplicação de CRISPR permitiu mapear a sensibilidade a manganês a um polimorfismo único no transportador Pmr1 (Sadhu et al., 2016), mostrando alta resolução da técnica de mapeamento. Para tanto, inicialmente foram desenvolvidos 95 RNAs-guias que direcionaram Cas9 a sítios distribuídos ao longo de todo o braço esquerdo do cromossomo 7 da levedura *Saccharomyces cerevisiae*. A cepa utilizada no estudo era diploide heterozigótica originária do cruzamento de uma cepa de laboratório (BY) e um cepa de vinhedo (RM). Depois do processo de quebra e reparo, as células que sofreram reparo foram separadas por citometria de fluxo de acordo com a perda do gene GFP próximo ao telômero. O sequenciamento

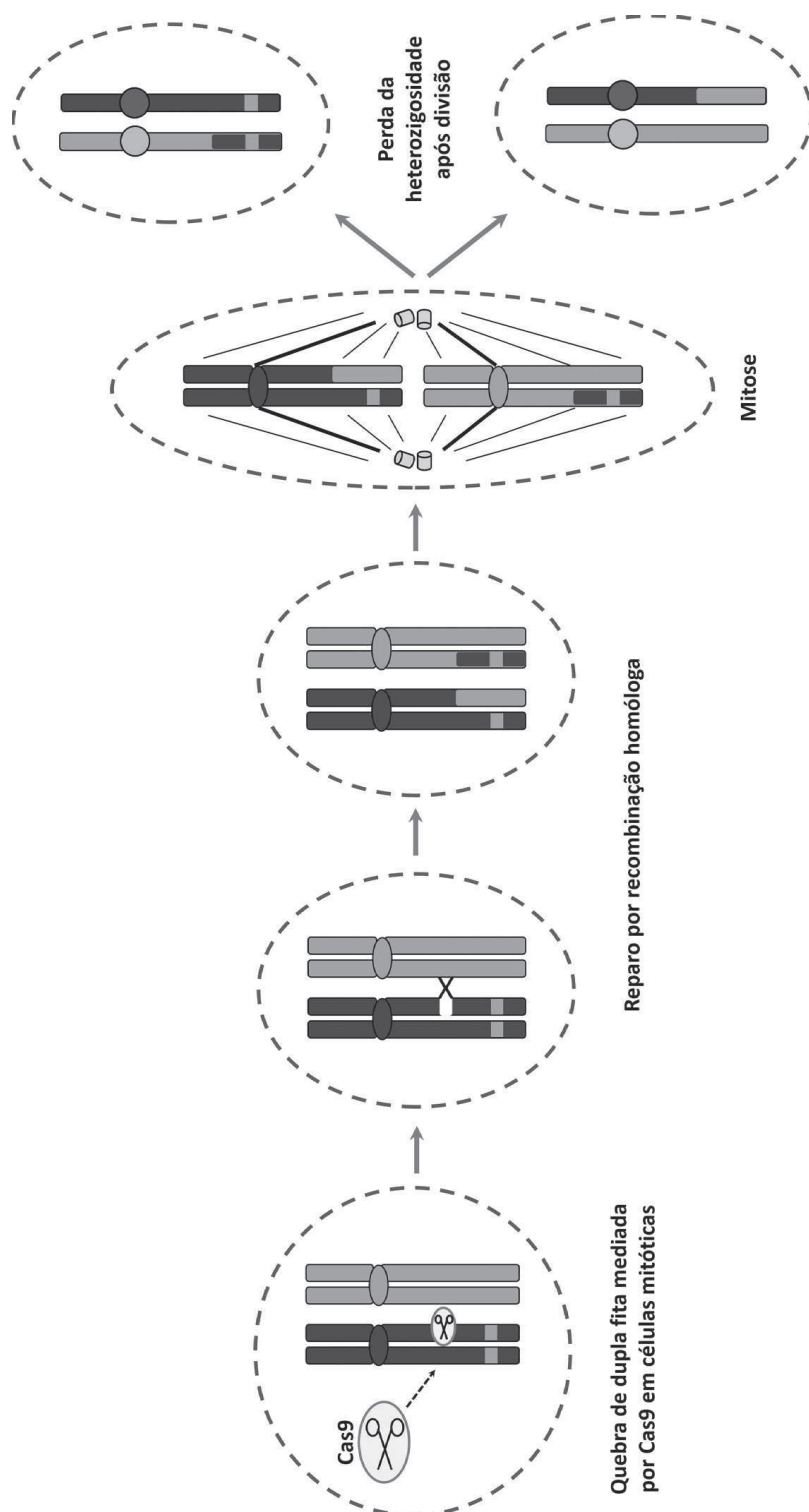


Figura 9. Perda de heteroziguidade induzida por sistema CRISPR-Cas9. Os indivíduos com perda de heteroziguidade (LOH) são isolados pela ausência do marcador dominante (banda preta). O processo pode ser repetido várias vezes e assim, pelo seguimento de marcadores, é possível mapear as variantes gênicas de uma determinada região cromossômica. Adaptado de Sadhu et al., 2016. Imagem: Roberto, GM e Annichini, MSB (2016).

posterior de 4 linhagens originárias de cada RNA-guia mostrou que 75% dos eventos de recombinação ocorreram de forma bem-sucedida em torno de 20 kb da sequência-alvo. O crescimento dessas linhagens foi observado em 12 condições diferentes que já haviam sido relacionadas ao cromossomo 7. A análise estatística mostrou que a perda da heterozigose afetava a sensibilidade ao manganês das linhagens com recombinação após a posição 180-200 kb. Uma análise mais fina avaliando um maior número de cepas cuja ligação e quebra via Cas9 ocorreram especificamente nesta região permitiu uma identificação mais precisa do intervalo em que as variantes se separam, e assim a descoberta de um polimorfismo responsável por aumentar a sensibilidade nas cepas BY ao manganês. Foi visto que esta variante codifica uma fenilalanina em BY e uma leucina de RM no gene *Pmr1*, um transportador de manganês.

O mapeamento por perda de heterozigose já se mostrava interessante, pois o restante do genoma permanece intacto, aumentando a confiabilidade do resultado. Com a introdução de CRISPR à técnica, a especificidade desta aumenta, pois torna-se possível gerar alta frequência de eventos de recombinação e acessar qualquer região do genoma, tudo isso em pouco tempo. Dessa forma, essa metodologia facilita o entendimento da genética por trás das variações fenotípicas (Sadhu et al., 2016).

Além disso, essa metodologia pode ser aplicada não somente para mapeamento mas também no caso de certos alelos relacionados a doenças cujo acesso direto por CRISPR é difícil devido à complexidade da sequência. Estes alelos podem ser reparados pela indução de DSB nas proximidades do alelo patogênico e sua substituição pelo alelo sadio através de perda de heterozigosidade (Sadhu et al., 2016).

7. Rastreamento de RNAs

Embora atualmente existam diversas e bem estabelecidas maneiras de reduzir a expressão gênica através do emprego de RNAs de interferência ou de sequências nucleotídicas antissenso, um estudo mais profundo relacionado à regulação pós-transcricional dos RNAs mensageiros como o tráfico intracelular, formas de *splicing* e poliadenilação, bem como os parâmetros espaço-temporais envolvidos na tradução, ainda carecem de uma metodologia mais robusta para avaliação.

A localização do RNA na célula, bem como o momento e a forma pela qual ele chega a um determinado compartimento subcelular, podem influenciar se as proteínas correspondentes estão sendo produzidas no local ou na hora correta. Diversas doenças são decorrentes diretamente de alterações no RNA; defeitos no transporte de mRNAs ou nos mecanismos de *splicing* estão relacionados a quadros como autismo ou câncer, por exemplo (Srebrow e Kornblihtt, 2006; Liu-Yesucevitz et al., 2011). Dessa forma, a elucidação destes processos se beneficiaria imensamente de técnicas para rastreamento de RNAs em células vivas, permitindo observar desde sua localização até suas variações ao longo do eixo transcrição-tradução.

Algumas técnicas têm sido empregadas para rastrear moléculas de RNA utilizando proteínas, oligonucleotídeos complementares ao alvo ou uma combinação dos dois (Lu et al., 2009; Sokol et al., 1998; Fouts et al., 1997; Bertrand et al., 1998). Entretanto, todas elas necessitam de sofisticada manipulação genética e não servem para reconhecimento arbitrário de sequências de RNA, além de potencialmente influenciarem o papel do RNA endógeno.

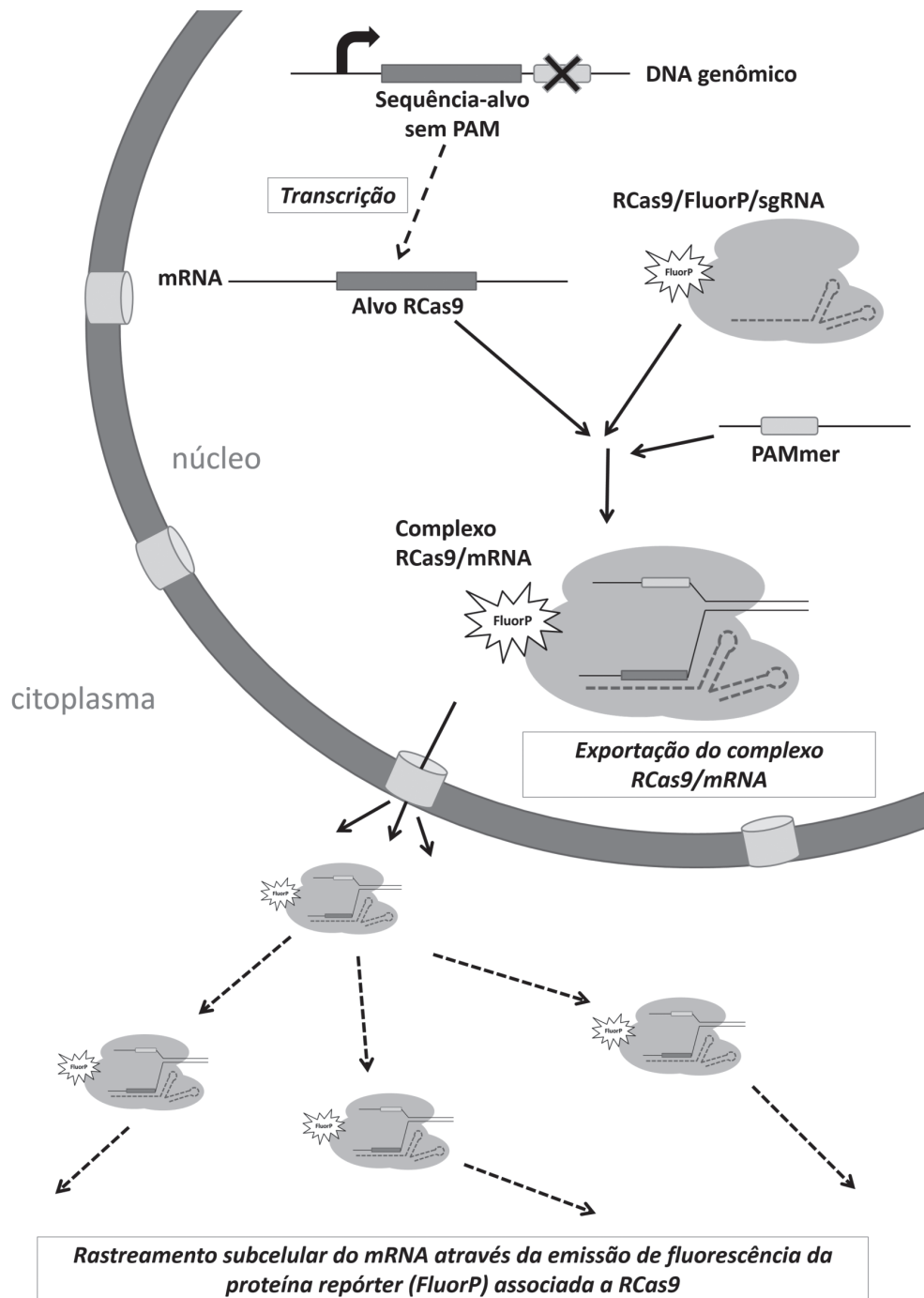


Figura 10. Rastreamento de RNA em células vivas através de sistema CRISPR/RCas9 associado a uma proteína fluorescente e sequência PAMmer. Imagem: Roberto, GM e Annichini, MSB (2016) adaptado de Nelles et al. (2016).

Nelles e colaboradores (2016) mostraram que a ligação de dCas9:GFP, associada a um gRNA, é capaz de se ligar à sequência 3' UTR do mRNA-alvo ainda no núcleo (sistema denominado de RCas9) sem interferir com sua exportação para o citoplasma, sua abundância ou a quantidade de proteína produzida a partir dele (Nelles et al., 2016). Esta abordagem

apresenta certas vantagens em relação às outras técnicas de rastreamento de RNAs, visto que o reconhecimento se baseia apenas na especificidade do gRNA, sem a necessidade de manipulação genética e outras proteínas ligantes de RNA, permitindo expandir as habilidades dos pesquisadores para rastrear os transcritos.

Dentre as futuras aplicações desta técnica pode-se citar: (i) a avaliação ou alteração do *splicing* por meio do recrutamento de diferentes proteínas e (ii) o rastreamento de RNAs podendo levar à descoberta de diferenças no padrão de expressão gênica entre indivíduos saudáveis e doentes. A longo prazo torna-se possível a modulação de muitas características e reprogramação do comportamento celular via alterações no processamento do RNA, na expressão gênica ou na concatenação de enzimas no RNA-alvo visando aplicação na biologia sintética e no tratamento de doenças.

8. Conclusões

É admirável observar como, a partir de apenas dois elementos – Cas9 e gRNA –, os pesquisadores têm sido capazes de expandir a caixa de ferramentas moleculares baseadas em CRISPR (figura 11). Apenas poucos anos após o lançamento dos alicerces da técnica, diversas estratégias derivadas de CRISPR foram desenvolvidas e provavelmente muitas outras virão nos próximos anos.

	Ação de Cas9	Alvo	Variação da técnica	Objetivo
CRISPR	Clivagem	DNA	Edição genética (KO; KI; SA e DG)	Modificação
			Mapeamento gênico	Localização
	Ligação	RNA	Clivagem de RNA	Expressão
			Regulação da expressão gênica	
		DNA	Marcação do DNA	Localização
			RNA	

Figura 11. Relação entre as formas de ação de Cas9 sobre diferentes ácidos nucleicos e as técnicas derivadas. Imagem: Pereira, TC (2016).

Referências

- Anton T, Bultmann S, Leonhardt H, Markaki Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*. 2014;5(2):163-72.
- Bertrand E, Chartrand P, Schaefer M, Sheno SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. *Mol Cell*. 1998;2:437-45.
- Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle MPR, Iyer E, et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods*. 2015;12:326-8.
- Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res*. 2013;23:1163-71.

- Cho H, Orphanides G, Sun X, Yang X-J, Ogryzko V, Lees E, et al. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol*. 1998;18:5355-63.
- Deng W, Shi X, Tjian R, Lionnet T, Singer RH. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc Natl Acad Sci USA*. 2015;112(38):11870-5.
- Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346:1258096. <http://dx.doi.org/10.1126/science.1258096>.
- Fouts DE, True HL, Celander DW. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. *Nucleic Acids Res*. 1997;25:4464-73.
- Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*. 2013;154:442-51.
- Henson V, Palmer L, Banks S, Nadeau JH, Carlson GA. Loss of heterozygosity and mitotic linkage maps in the mouse. *Proc Natl Acad Sci USA*. 1991;88:6486-90.
- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157(6):1262-78.
- Klug WS, Cummings MR, Spencer CA, Palladino MA. *Conceitos de Genética*. 9. ed. Artmed; 2010.
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2014;517:583-8.
- Lewin, B. *Genes IX*. 9. ed. Artmed; 2008.
- Liu-Yesucevitz L, Bassell GJ, Gitler AD, Hart AC, Klann E, Richter JD, et al. Local RNA translation at the synapse and in disease. *J Neurosci*. 2011;31(45):16086-93.
- Lu G, Dolgner SJ, Tanaka TM. Understanding and engineering RNA sequence specificity of PUF proteins. *Hall Curr Opin Struct Biol*. 2009;19(1):110-5.
- Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci USA*. 2015;112(10):3002-7.
- Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013;10:977-9.
- McCaffrey J, Sibert J, Zhang B, Zhang Y, Hu W, Riethman H, et al. CRISPR-CAS9 D10A nickase target-specific fluorescent labeling of double strand DNA for whole genome mapping and structural variation analysis. *Nucleic Acids Res*. 2016;29;44(2):e11.
- Mei Y, Wang Y, Chen H, Sun ZS, Ju XD. Recent Progress in CRISPR/Cas9 Technology. *J Genet Genomics*. 2016;43(2):63-75.
- Meru J, Sadhu JS, Bloom LD, Kruglyak L. Crispr-directed mitotic recombination enables genetic mapping without crosses. *Science*. 2016. <http://dx.doi.org/10.1126/science.aaf5124>.
- Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, et al. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*. 2016;165(2):488-96. <http://dx.doi.org/10.1016/j.cell.2016.02.054>.
- O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*. 2014;516:263-6.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-308.
- Sadhu MJ, Bloom JS, Day L, Kruglyak L. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science*. 2016 May;352(6289):1113-6.
- Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32(4):347-55.
- Snustad P, Simmons M. *Principles Of Genetics*. Hoboken: Wiley; 2012.
- Sokol DL, Zhang X, Lu P, Gewirtz AM. Real time detection of DNA. RNA hybridization in living cells. *Proc Natl Acad Sci USA*. 1998;95:11538-43.
- Srebrow A, Kornbliht AR. The connection between splicing and cancer. *J Cell Sci*. 2006;119(Pt 13):2635-41.
- Strachan T, Read A. *Genética Molecular Humana*. 4. ed. Artmed; 2013.
- Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. A protein tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*. 2014;159:635-46.
- Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*. 2015;12:1143-9.
- Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, et al. Transcription factor binding dynamics during human ES cell differentiation. *Nature*. 2015;518:344-9.
- Yin Y, Petes TD. Genome-wide high-resolution mapping of UV-induced mitotic recombination events in *Saccharomyces cerevisiae*. *PLOS Genet*. 2013. <http://dx.doi.org/10.1371/journal.pgen.1003894>.
- Zhang C, Xiao B, Jiang Y, Zhao Y, Li Z, Gao H, et al. Efficient editing of malaria parasite genome using the CRISPR/Cas9 system. *MBio*. 2014 Jul;5(4):e01414-14.