

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Programa de Pós-Graduação em Genética e Biologia Molecular

OTIMIZANDO A EDIÇÃO GÊNICA: DO SISTEMA IMUNE AOS *OFF-TARGETS*

Martiel Vaz de Freitas

(VERSÃO PARCIAL)

Orientadora: Dra. Ursula da Silveira Matte

PORTO ALEGRE

Maior 2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Programa de Pós-Graduação em Genética e Biologia Molecular

OTIMIZANDO A EDIÇÃO GÊNICA: DO SISTEMA IMUNE AOS *OFF-TARGETS*

Martiel Vaz de Freitas

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutora em Ciências (Genética e Biologia Molecular).

Orientadora: Dra. Ursula da Silveira Matte

PORTO ALEGRE

Maio/ 2022

Este trabalho foi realizado no Laboratório de Células, Tecidos e Genes - CTG, no Centro de Pesquisa Experimental do Hospital e Clínicas de Porto Alegre, e vinculado ao Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul e no Núcleo de Bioinformática do Hospital de Clínicas de Porto Alegre.

Apoio financeiro:

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

Fundo de Incentivo à Pesquisa e Eventos do Hospital de Clínicas de Porto Alegre (FIPE)

Dedicado à Elaine Maurícia (*in memoriam*),
a melhor bruxa madrinha que o mundo já viu.

Agradecimentos

Já olhou para trás pra ver se você está onde queria?

Aos 6, eu queria ser professora. Minha mãe me incentivava através dos livros e de um quadro negro na área da entrada de nossa casa, onde eu ensinava meus três irmãos mais novos a ler. Obrigada **Mãe** por me mostrar desde muito, muito cedo que o caminho que leva a resolução de todos os problemas é calçado com palavras. Obrigada aos **meus irmãos** pelas primeiras lições de vida sobre compartilhamento e coragem, e por partilharem comigo dos mundos imaginários vindos dos livros dispostos na estante no final do corredor.

Aos 8 mudei de ideia e decidi que seria “engenheira de aviões”. Meu pai me incentivava sempre que me via desmontando os mais variados eletrodomésticos perguntando: “Entendeu como funciona?” Ao meu sim, replicava: “Então, agora monta de volta.” Obrigada **Pai** pela tua paciência. Não muito comprida é verdade, mas que eu sempre achava sem fim quando se tratava de me deixar descobrir as coisas por mim mesma.

Aos 12 decidi que seria pesquisadora. “De que?” “De algo que eu ainda vou descobrir.” Não mudei mais de ideia. Agradeço a todas as minhas professoras na escola que não se importavam com as perguntas e respostas mais impertinentes que meu eu criança era capaz de formular. Até hoje meu deus é o Tempo: tudo pode, tudo sabe e tudo vê.

Ao longo do meu tortuoso caminho acadêmico, existiram momentos em que me afastei da estrada que me levaria a ser uma pesquisadora. Hoje, sei que isso aconteceu por que eu ainda não tinha descoberto onde eu queria estar. Durante o processo de percorrer, sair e voltar à essa estrada, tenho a sorte de dizer que encontrei não somente um tema que me move e diverte ao explorar, mas também muitos incentivadores do meu desenvolvimento. De modo que quaisquer palavras postas aqui me deixam a sensação de serem insuficientes. É preciso tentar, entretanto.

Começo então, agradecendo à minha amiga, companheira, esposa, suporte emocional e outra mãe dos meus pets: **Carolina**. Tenho certeza de que se o caminho até aqui foi mais leve, isso se deve a tua presença. Obrigada por confiar em nós, esperar por nós, cuidar de nós e investir em nós. Sou, diariamente, a pessoa mais feliz do mundo simplesmente porque tu estás nele. Seguimos na média. ;)

Agradeço a todos os **meus amigos**. Não me encontro disposta a correr o risco de esquecer algum; assim, não cito nomes. Foram todos essenciais. Agradeço os cafés, almoços e jantares, pelas risadas e pelos abraços; agradeço as dicas de leitura e encontros no parque; as viagens, os jogos e churrascos; os pedidos de ajuda que me faziam estudar mais; agradeço pelas discussões, científicas ou não.

Agradeço a todos os professores do Laboratório, bem como a todos os colegas e amigos: **Ágnis, Ana, Cristal, Lariane Laura, Luís, Lucca, Marina, Mariana, Pâmella e Paola. Grazi, Luisa, Hallana, Yorran...** Enfim, todos! Ter convivido de maneira mais próxima com cada um de vocês foi incrível, ainda que tenha sido só pela metade do período completo do doutorado.

Agradeço à minha orientadora, **Ursula Matte**. Primeiro, por me receber no CTG; segundo, por sempre achar um tempinho para avaliar minhas ideias. Obrigada pela confiança e pela liberdade dada para que eu fosse incluindo cada vez mais imuno aos meus trabalhos. Obrigada também pela compreensão, conselhos, dicas, críticas, elogios, e pelas piadinhas de qualidade duvidosa. Sentirei falta!

Agradeço aos **demais amigos, colegas e professores** com os quais convivi no Hospital de Clínicas de Porto Alegre, no Núcleo de Bioinformática e no PPGBM; agradeço ao **Elmo** especialmente, por ser o melhor amigo colorado que uma gremista poderia ter.

Em tempo agradeço: ao CNPq, à UFRGS e ao HCPA; ao relator e à banca, pelo tempo dispendido na leitura e avaliação deste conjunto de trabalhos.

“Uma única pergunta pode ser mais explosiva do que mil respostas.”

(Jostein Gaarder)

Sumário

Resumo	9
Abstract	10
Introdução	12
1 <i>CRISPR</i>	12
1.1 Edição gênica	12
1.2 A ação dos sistemas CRISPR/Cas	14
1.3 Desafios éticos à edição gênica	17
1.4 Melhorias nas proteínas Cas	19
2 <i>Resposta imune</i>	21
2.1 Patógenos e sistema imune	21
2.2 Resposta contra vetores e outros produtos	22
2.3 O impacto do sistema imune em terapia gênica	22
2.4 A diversidade dos HLAs e edição gênica	24
3 <i>Efeitos off-target</i>	25
3.1 Predição de <i>off-targets</i>	25
Objetivos	27
2.1 <i>Objetivo Geral</i>	27
2.2 <i>Objetivos específicos</i>	27
Capítulo 2	29
Protection is not always a good thing: the immune system's impact on gene therapy	29
<i>Abstract</i>	30
Capítulo 3	33
Small enemies: how HLA variation across populations may interfere in Cas9-mediated gene editing	33
<i>Abstract</i>	33
Capítulo 4	36
Online tools for structural analysis of Cas9 off-targets occurrence	36
Capítulo 5	39
Off-target discovering tools aren't all the same: understanding commonly used predictive tools using experimental data	39
<i>Abstract</i>	39
Discussão Geral	42
Referências adicionais	52

A edição gênica por CRISPR/Cas9 tem potencial para mudar o tratamento de doenças. Entretanto um aspecto importante a ser considerado é o sistema imunológico, que pode afetar a eficiência da terapia gênica de diferentes formas. Neste trabalho revisamos, de forma geral, produtos de terapias aprovados e os pontos onde a resposta imune pode ser elicitada, principalmente em edição gênica. A saber, esses pontos incluem respostas contra o vetor ou seu conteúdo após a entrega e contra o produto do gene corrigido. Apresentamos também as estratégias desenvolvidas para transpor estes obstáculos. Ainda, apresentamos análises a respeito da necessidade de avaliação da variabilidade. Os resultados mostram que dentre os epitopos que podem ser gerados, uma grande quantidade seria capaz de induzir uma resposta imune quando apresentados por alelos HLA específicos comuns a três grupos étnicos. Ao mesmo tempo que a facilidade da aplicação de CRISPR/Cas9 nos mais variados estudos traz benefícios, o mesmo sistema tem o potencial de causar efeitos colaterais perigosos, especialmente quando a edição ocorre fora do sítio alvo, gerando potenciais distúrbios no bom funcionamento dos mecanismos da célula. A maioria dos trabalhos sobre *off-targets*, no entanto, leva em conta apenas as sequências de DNA genômico e sgRNA, esperando que a semelhança nessas sequências auxilie na diminuição de sua ocorrência. Acreditamos, entretanto que outros aspectos deveriam ser levados em consideração. Assim, não só avaliamos os resultados de preditores como também descrevemos um pipeline baseado em ferramentas online que mostra a importância de considerar os recursos estruturais na análise de *off-targets*.

Gene editing by CRISPR/Cas9 has the potential to change the treatment of disease. However, an important aspect to be considered is the immune system, which can affect the efficiency of gene therapy in different ways. In this work, we generally review approved therapy products and the points where the immune response can be elicited, mainly in gene editing. Namely, these points include responses against the vector or its contents after delivery and against the corrected gene product. We also present the strategies developed to overcome these obstacles. Furthermore, we present analyses regarding the need to evaluate the variability. The results show that among the epitopes that can be generated, a large amount would be able to induce an immune response when presented by specific HLA alleles common to three ethnic groups. While the ease of application of CRISPR/Cas9 in the most varied studies brings benefits, the same system has the potential to cause dangerous side effects, especially when editing occurs outside the target site, generating potential disturbances in the proper functioning of the mechanisms of the cell. Most of the work on off-targets, however, takes into account only the DNA and sgRNA sequences, hoping that the similarity in these sequences will help to reduce their occurrence. We believe, however, that other aspects should be taken into account. Thus, we not only evaluate predictor results but also describe a pipeline based on online tools that shows the importance of considering structural features in off-target analysis.

Capítulo 1

Introdução e Objetivos

1 CRISPR

1.1 Edição gênica

Compreender a história da edição gênica parece ser extremamente importante para entender o atual estado deste campo de pesquisa. Avanços significativos em biologia molecular geralmente são construídos após décadas de estudos e união de esforços de grandes cientistas da área. Isso não é diferente na área de edição gênica. Foram eventos importantes como a descoberta da dupla hélice do DNA (Watson and Crick, 1953) e do DNA recombinante (Jackson, Symons and Berg, 1972) que deram suporte para o surgimento e crescimento deste campo.

Dos primeiros estudos até a consolidação do uso de CRISPR em edição gênica, detalhados na Figura 1, anos de descobertas, desenvolvimento e aprimoramento de técnicas resultaram em sistemas de alta precisão.

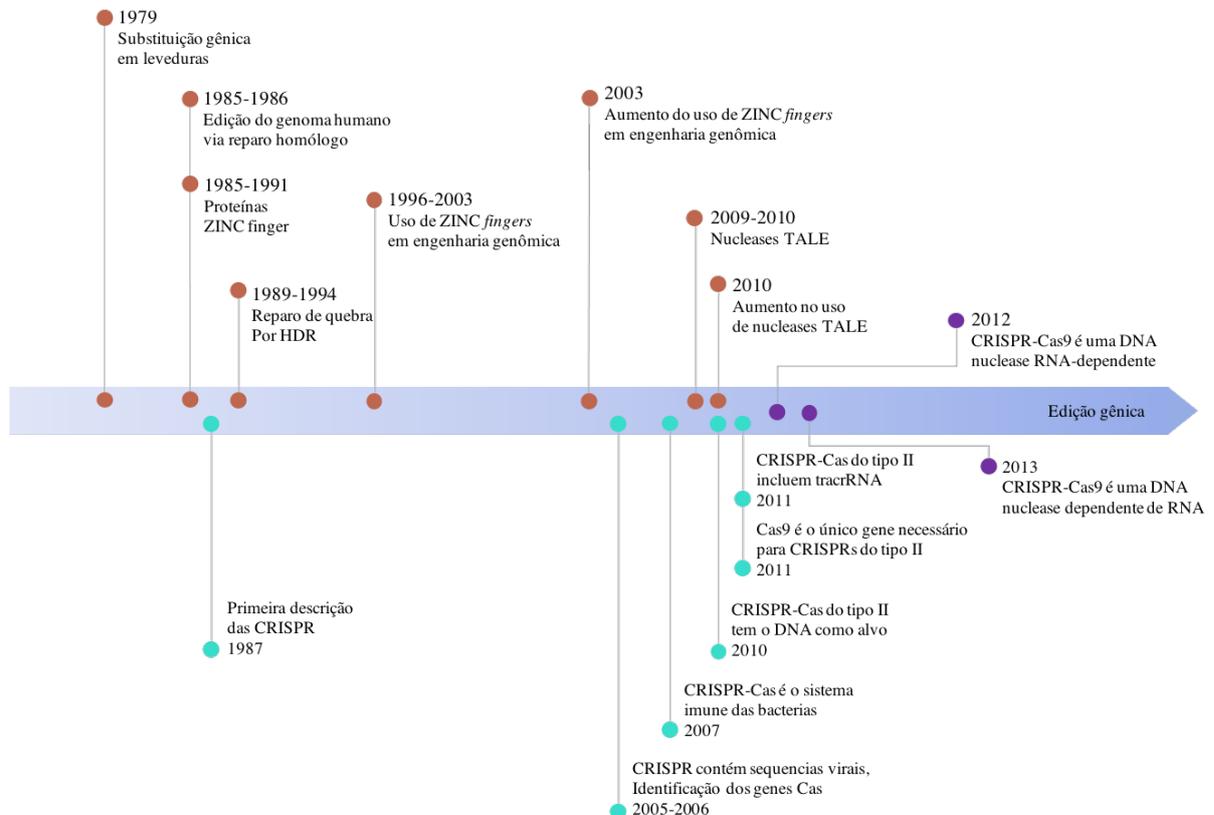


Figura 1. **Linha do tempo até o estabelecimento da edição gênica em 2013.** Cor marrom representa achados importantes relacionados a ZNF e TALEN; cor verde indica pontos importantes no desenvolvimento das técnicas CRISPR; cor roxa representa achados que culminaram no estabelecimento da edição gênica com CRISPR/Cas.

A descoberta de nucleases de dedo de zinco (ZFN), possivelmente a primeira técnica para engenharia de genoma altamente direcionada, melhorou a eficácia do direcionamento de genes de várias maneiras. Cada ZFN se liga especificamente a um conjunto de três pares de bases, e pode ser combinado para reconhecer sequências de DNA mais longas (Gupta and Musunuru, 2014). Os pesquisadores fundiram a enzima de corte de DNA *FokI* ao domínio de ligação do DNA ZFN para criar 'tesouras' capazes de clivar o DNA em um local específico (Eid and Mahfouz, 2016), gerando uma quebra de fita dupla (DSB).

Se um molde de reparo é co-transfectado na célula, então uma fração delas sofre recombinação homóloga no local da clivagem, que resulta na incorporação do molde de interesse (Zaboikin *et al.*, 2017). E, graças a métodos como as ZFNs, somos capazes de, conhecendo uma mutação patogênica em um gene específico, modificá-la de forma a restaurar a função gênica normal.

Anos mais tarde foi a vez das TALEN. Essas nucleases seguiram os passos das ZFNs e consistem de fusões de um efetor semelhante ao ativador de transcrição (TALE) e o domínio catalítico da endonuclease de restrição *FokI* (Gupta & Musunuru, 2014). Esta foi a primeira ferramenta de edição gênica que pode ser projetada e construída com relativa facilidade para atingir qualquer *locus* genômico de interesse, com alta precisão e alta eficiência (Becker and Boch, 2021). Posteriormente, o TALEN foi rapidamente aplicado para edição de genoma em culturas de células. Entretanto, foi amplamente substituído pelas tecnologias CRISPR, que são um pouco mais fáceis de construir, e mais fáceis de usar nos mais vários estudos (Becker and Boch, 2021).

Quase ao mesmo tempo em que se descobriu as ZFN, a primeira descrição da região CRISPR foi publicada. Em 1987, uma equipe japonesa de cientistas da Universidade de Osaka notou um estranho padrão de sequências de DNA em um gene pertencente à *Escherichia coli* (Ishino *et al.*, 1987). Esse gene consistia de cinco pequenos segmentos repetidos de DNA separados por sequências curtas de um DNA 'espaçador' que não se repetiam.

Os cinco fragmentos repetidos possuíam sequências exatamente iguais, compostas de 29 nucleotídeos. Em contraste, cada uma das sequências 'espaçadoras' tinha sua própria sequência, composta por 32 bases. Apenas no final dos anos 1990, compreendeu-se que esse era um padrão prevalente também em outras bactérias (Mojica *et al.*, 2000).

As repetições presentes eram tão comuns entre os organismos que receberam o nome de 'repetições palindrômicas curtas agrupadas regularmente interespaçadas', ou CRISPR, dado por uma equipe de cientistas holandeses em 2002 (Jansen *et al.*, 2002). Neste mesmo ano, o

grupo observou outro conjunto de sequências que sempre acompanhava a sequência CRISPR. Este segundo conjunto de sequências recebeu o nome de ‘Cas’ (do inglês, *CRISPR associated protein*). E esses genes pareciam codificar enzimas capazes de clivar o DNA (Jansen *et al.*, 2002).

Em 2005, três equipes científicas descobriram de forma independente que as sequências 'espaçadoras' entre as sequências CRISPR compartilham semelhanças com o DNA viral e levantaram a hipótese de que poderia se tratar de um mecanismo de defesa das bactérias (Bolotin, Quinquis e Sorokin, 2005), o que só se confirmaria em 2007 (Barrangou *et al.*, 2007). Foi em 2008 que ocorreu a primeira demonstração de como o mecanismo CRISPR/Cas funcionava (Brouns *et al.*, 2008). O trabalho relatou que quando uma bactéria enfrenta um invasor, ela copia e incorpora pequenos segmentos de DNA do invasor ao seu próprio genoma. Essas sequências são os ‘espaçadores’ entre as repetições curtas de DNA. E são esses segmentos que, quando transcritos em pequenos RNAs denominados CRISPR RNAs (crRNAs), permitem que a bactéria reconheça qualquer DNA proveniente de um invasor e ajudam a proteína Cas no processo de clivagem para desativação do patógeno.

Ficou claro assim, que o sistema CRISPR/Cas tinha como alvo molecular para clivagem não apenas RNA (Hale *et al.*, 2009), mas também sequências de DNA (Marraffini and Sontheimer, 2008). Até então considerado um paralelo aos mecanismos de silenciamento de RNAi eucarióticos, que têm como alvo o RNA, essa descoberta surpreendeu os cientistas. Marraffini e Sontheimer explicitam em seu trabalho que esse sistema poderia ser uma ferramenta poderosa se pudesse ser transferido para sistemas não bacterianos, o que ocorreu no final de 2012 (Jinek *et al.*, 2012).

1.2 A ação dos sistemas CRISPR/Cas

A ação do sistema CRISPR/Cas em procariotos é dividida em três etapas (Makarova and Koonin, 2005), descritas abaixo e esquematizadas na figura 2:

- i. adaptação ou integração do espaçador;
- ii. processamento do transcrito primário do locus CRISPR (pré-crRNA) e maturação do crRNA;
- iii. interferência de DNA (ou RNA).

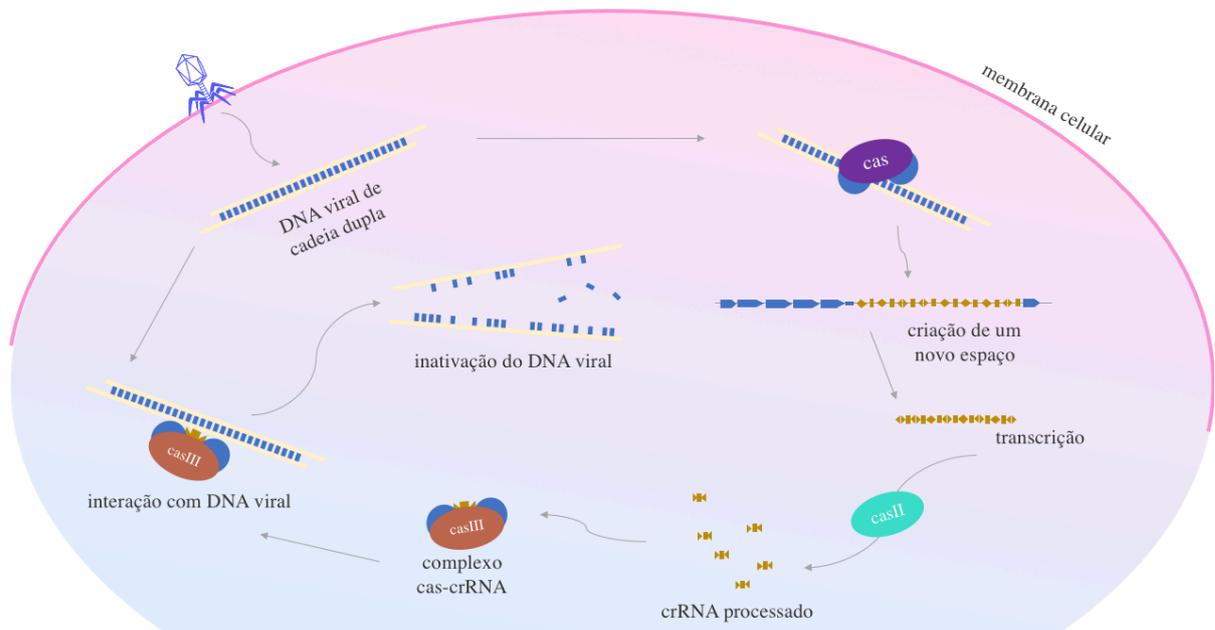


Figura 2. **Esquema de funcionamento do sistema CRISPR.** Quando o DNA viral é transfectado para o interior da bactéria o sistema mediado pelas proteínas Cas trabalha na identificação, no processamento e na inativação do DNA invasor.

A primeira etapa é mediada por duas proteínas: Cas1 e Cas2. Essas proteínas estão presentes na maioria dos sistemas CRISPR/Cas conhecidos e são suficientes para a inserção de espaçadores nos cassetes CRISPR (Yosef, Goren and Qimron, 2012). Quando unidas, formam o complexo necessário para o processo de adaptação do eucarioto, e são quase autônomas do resto do sistema (Makarova and Koonin, 2005). A atividade endonuclease Cas1 é necessária para a integração do espaçador ao genoma bacteriano, enquanto a Cas2 desempenha uma função não enzimática (Nuñez *et al.*, 2014).

A segunda etapa, o processamento de pré-crRNA em crRNAs guia, é realizada por um complexo de RNA endonuclease dedicado ou por meio de um mecanismo alternativo que envolve RNase III bacteriana e um RNA adicional (Deltcheva *et al.*, 2011). O crRNA maduro é ligado por uma (tipo II) ou várias (tipos I e III) proteínas Cas que formam o complexo efetor que tem como alvo o DNA. O complexo efetor dos sistemas tipo I é conhecido como Cascade (*CRISPR-associated complex for antiviral defense*) (Brouns *et al.*, 2008).

Em 2009, os motivos adjacentes protoespaçadores (PAM) foram descobertos durante a análise computacional de sequências conservadas próximas a esses motivos (Mojica *et al.*, 2009). Essa sequência curta e conservada de 2 a 5 pb está localizada próxima ao DNA alvo e é necessária para que o maquinário bacteriano seja capaz de discriminar entre ‘próprio’ e ‘não-próprio’. Proteínas Cas específicas reconhecem e se ligam à sequência PAM, desenrolando a hélice de DNA adjacente.

Assim, o DNA fica disponível para hibridização com o crRNA, produzindo uma estrutura R-loop de fita tripla. As sequências *seed* próximas a esses elementos PAM são interrogadas quanto à complementaridade com o espaçador de crRNA para induzir o emparelhamento completo de bases e a subsequente clivagem (Sternberg *et al.*, 2014).

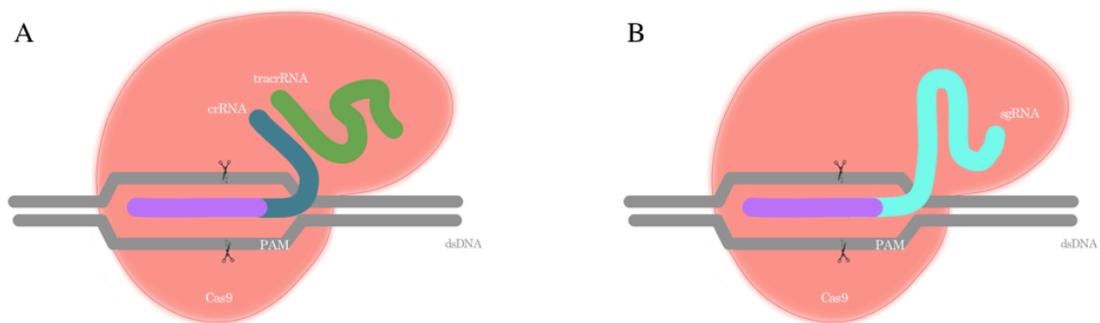


Figura 3. Esquema de anelamento entre o sgRNA, o DNA alvo e a proteína Cas9. A) Esquema encontrado na maioria das bactérias composto de crRNA e tracrRNA. B) Esquema simplificado, contendo apenas um sgRNA, apresentado por Doudna e Charpentier (Jinek *et al.*, 2012).

Por questões de simplicidade, pesquisadores envolvidos com edição gênica, passaram a estudar um sistema CRISPR mais simples, proveniente de *Streptococcus pyogenes*. Esse sistema, de tipo II, se baseia unicamente na proteína Cas9. A endonuclease Cas9 é um sistema de quatro componentes que inclui duas moléculas pequenas: crRNA e RNA CRISPR de transativação (tracrRNA). Jennifer Doudna e Emmanuelle Charpentier reprojaram a endonuclease Cas9 para que fosse um sistema mais gerenciável, com apenas dois componentes. Isso foi feito através da fusão das duas moléculas de RNA tradicionais em um “RNA guia único” (sgRNA) que, quando combinado com Cas9, poderia encontrar e promover a quebra do DNA no local especificado pelo guia.

O trabalho de Feng Zhang em demonstrar a utilidade do CRISPR para manipulação genética de células eucarióticas foi extremamente significativo. Zhang, que já havia trabalhado em outros sistemas de edição de genoma, como TALENs, adaptou com sucesso o sistema CRISPR/Cas9 para edição de genoma em células eucarióticas (Cong *et al.*, 2013). Zhang e sua equipe projetaram dois ortólogos Cas9 diferentes (*S. pyogenes* e *S. thermophilus*) e demonstraram a clivagem do genoma direcionado em células humanas e de camundongos. Neste trabalho, mostraram ainda que o sistema (i) poderia ser programado para atingir vários *loci* genômicos e (ii) poderia conduzir o reparo dirigido por homologia (Mali *et al.*, 2013),

conforme mostrado na figura 4. Assim o sistema CRISPR/Cas9 expandiu ainda mais suas aplicações.

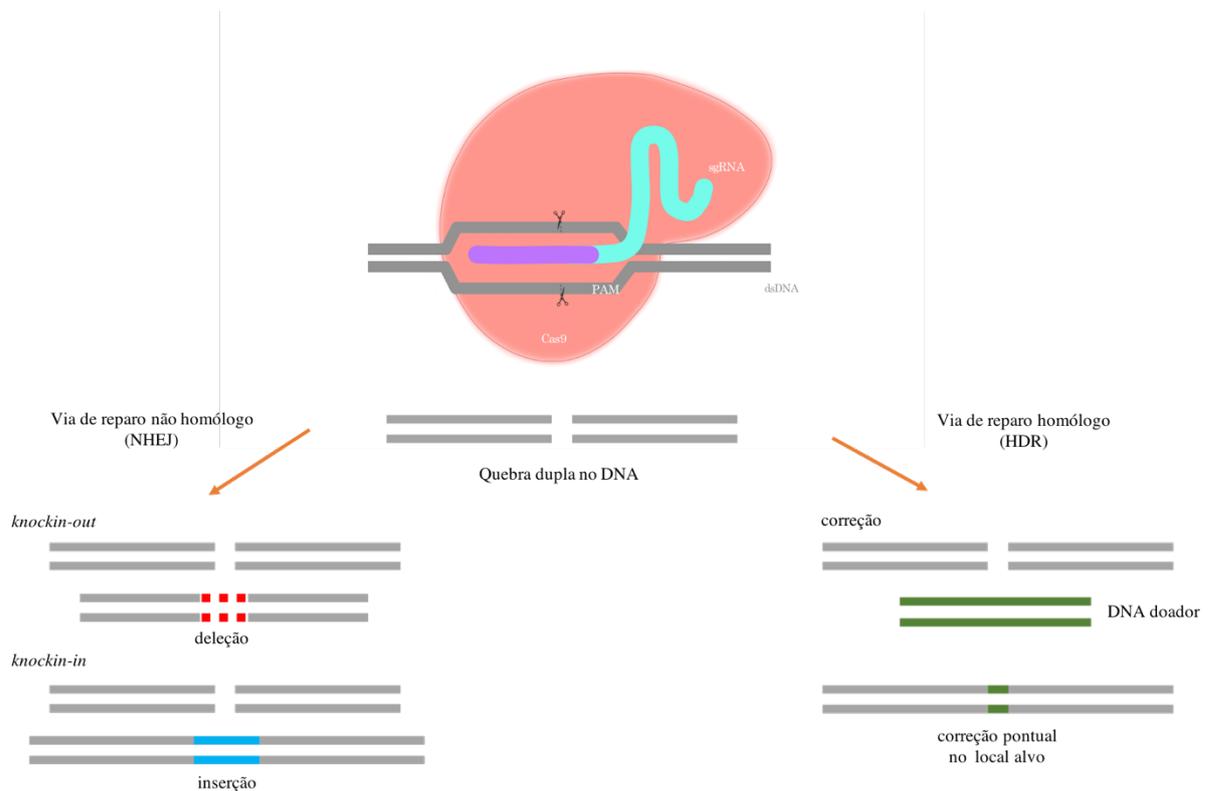


Figura 4. Possíveis formas de reparo do DNA após a quebra dupla promovida pelo sistema CRISPR/Cas9 e seus produtos. A quebra da dupla fita realizada pela nuclese Cas9 pode sofrer reparo não homólogo (NHEJ), que resultaria na inserção ou deleção de tamanhos variáveis no local onde o reparo ocorre; ou via reparo homólogo (HDR), que parte do pressuposto da presença de uma sequência molde (DNA doador) que fornece a correção a ser introduzida através de mutações de ponto ou inserção de novas sequências no DNA da célula. Adaptado de Sander and Joung (2014).

1.3 Desafios éticos à edição gênica

Seja para correção de mutações no DNA, criação de modelos animais (Zarei *et al.*, 2019) ou inibição/ativação de genes específicos (Truong *et al.*, 2019), CRISPR ganhou notoriedade. E não apenas dentro dos laboratórios de universidades e empresas, mas também com o potencial de revolucionar a prática médica. Através dela não só é possível tratar, como é possível prevenir doenças.

Sendo assim, não é surpresa que tamanho poder tenha acendido alguns alertas no decorrer do seu rápido desenvolvimento e amplo emprego. Em 2015, por exemplo, a Sociedade Japonesa de Terapia Gênica (JSGT) manifestou-se, em conjunto com a Sociedade Americana de Terapia Genética e Celular (ASGCT), considerando a manipulação genética humana inadequada até que questões éticas e científicas fossem resolvidas (Friedman *et al.*, 2015).

Seguindo essa mesma linha de pensamento, em dezembro deste mesmo ano, Jennifer Doudna e outros pesquisadores (Sanders, 2015) disseram acreditar que os trabalhos com CRISPR deveriam continuar, mas que tentativas de alteração de genes em óvulos humanos, espermatozoides ou embriões deveriam ser interrompidas até que pesquisadores, médicos e o público leigo soubessem das implicações dessas modificações não apenas em uma pessoa, mas em toda a sua linhagem para dali em diante.

Mais tarde, em 2016, a JSGT em conjunto com outras instituições científicas japonesas, defendeu uma proposta para que a aplicação clínica da edição gênica em linhagens germinativas fosse proibida. Além disso, solicitou ao governo do país que fossem definidas diretrizes nacionais apropriadas voltadas para a pesquisa básica com esse tipo de ferramenta (Ormand *et al.*, 2017).

A regulamentação surgiu enquanto o uso da técnica aumentava, com um panorama regulatório variado ao redor do mundo. Por exemplo, a Alemanha tem experimentos com linhagem germinativa humana limitados pela Lei de Proteção de Embriões de 1990, que proíbe a geração e o uso de embriões para pesquisa básica. Na Coreia do Sul, a proibição de experimentos genéticos de modificação de embriões humanos fica a cargo da Lei de Bioética e de Biossegurança, enquanto a Lei de Notificação proíbe qualquer alteração genética (Kim, 2017).

Na China, em 2003 o Ministério da Saúde Chinês (NHFPC) proibiu a manipulação de embriões humanos através da emissão de duas regulamentações. Todavia, esses documentos não têm o status de lei, mas de diretrizes ministeriais (Song & Joly, 2021). E foi assim que, em 2018, o pesquisador chinês He Jiankui anunciou a edição de embriões, na polêmica tentativa de modificar as duas cópias do gene codificador da proteína CCR5, na alegada tentativa de conferir resistência contra infecções por HIV. Esta alteração resultaria em indivíduos resistentes ao HIV. Dos embriões editados, dois nasceram em outubro daquele ano (Greely, 2019).

Após o evento, a Organização Mundial da Saúde (OMS) reuniu um Comitê Consultivo de Especialistas multidisciplinares em Edição do Genoma Humano com a função de examinar os desafios científicos, éticos, sociais e legais associados à edição genômica em humanos, fosse em células somáticas ou em células germinativas (WHO, 2019). Além disso, o comitê foi instruído a aconselhar e fazer recomendações sobre as diretrizes apropriadas ao tema.

Assim como as questões éticas seguem sendo debatidas, as questões científicas acerca do uso de sistemas CRISPR tampouco arrefecem. A técnica mais empregada para edição gênica atingiu rapidamente seus primeiros "*milestones*". Entretanto, apesar de seus inúmeros sucessos,

não são apenas os pontos positivos que vêm sendo destacados ao longo dos anos. Com o uso, novas questões a respeito de sua segurança vêm sendo levantadas, como por exemplo:

- i) qual a melhor proteína Cas para ser associada à CRISPR, em função da especificidade da PAM;
- ii) como evitar resposta imune contra os vetores;
- iii) como evitar resposta imune contra a célula já editada;
- iv) como prever e mitigar a ocorrência de efeitos *off-targets*.

1.4 Melhorias nas proteínas Cas

Questões associadas à escolha da proteína visando uma melhor performance da técnica vêm sendo discutidas desde que a ferramenta se mostrou útil à edição gênica. Em 2011, Makarova *et al.* apresentaram a primeira classificação das proteínas com base na filogenia dos genes Cas mais comuns em *archeas* e bactérias, bem como considerando a sequência e a organização das repetições CRISPR, além da arquitetura do *loci* CRISPR/Cas.

Dada a diversidade de proteínas Cas passíveis de serem utilizadas para edição, notou-se o importante papel da PAM no processo. O motivo adjacente do protoespaçador (PAM) é uma sequência curta de DNA (geralmente 2-5 pares de bases de comprimento). O sítio de clivagem da nuclease Cas9 situa-se a três nucleotídeos de distância dessa região. Em 2015, uma série de modificações para seleção positiva com o intuito de localizar mutantes de *SpCas9* capazes de clivar sequências de DNA localizadas a montante de motivos PAM NGA ou NGC (Kleinstiver *et al.*, 2015) identificou quatro novas variantes de *SpCas9* com especificidade de PAM alterada (Tabela 1).

Tabela 1. Variantes de *SpCas9* e possíveis sequências PAM descritas por Kleinstiver *et al.* (2015).

Variante de <i>SpCas9</i>	Mutações realizadas	Sequências PAM observadas
D11351E	D1135E	NGG
VQR	D1135V, R1335Q e T1337R	NGAN ou NGNG
EQR	D1135E, R1335Q e T1337R	NGAG
VRER	D1135V, G1218R, R1335E e T1337R	NGCG

Ainda que a nuclease mais conhecida e utilizada atualmente seja a Cas9 isolada de *Streptococcus pyogenes* (*SpCas9*), com um sgRNA associado, muitas outras foram descobertas em função dessa necessidade de reconhecimento de outras PAMs. Assim passou-se a olhar para ocorrências naturais de variantes de Cas9 em outros organismos (Collias & Beisel, 2021).

A Cas9 proveniente de *Staphylococcus aureus* (*SaCas9*) é a segunda nuclease mais popular em edição gênica, muito graças a seu tamanho reduzido. São 1053 aminoácidos, resultando em um quase 1 kb a menos do que a *SpCas9*. Isso a torna mais fácil de empacotar nos vetores virais utilizados em aplicações clínicas (Kim *et al.*, 2017).

A habilidade de editar tecidos ou tipos celulares específicos é de extrema importância em terapia gênica. Edição utilizando *SaCas9* mostrou-se bastante eficiente em editar células neuronais específicas em cérebros de ratos (Sun *et al.*, 2020). Além de ter se mostrado eficiente em inibir a replicação do vírus da hepatite B (HBV) *in vitro* e *in vivo*, quando entregues via AAV8 (vírus adenoassociado sorotipo 8) com promotores específicos de células hepáticas (Yan *et al.*, 2021).

Não são apenas esses os organismos estudados, entretanto. Trabalhos que utilizam variantes de Cas9 de *Streptococcus thermophilus* (*StCas9*) (Deveau *et al.* 2008), *Francisella novicida* (*FnCas9*) (Hirano *et al.*, 2016) e outras (Shmakov *et al.*, 2015) também têm sido publicados (Tabela 2).

Tabela 2: Organismos e especificidade de motivos PAM identificados para cada organismo.

Nuclease	Organismo	PAM observada	Publicação
<i>StCas9</i>	<i>Streptococcus thermophilus</i>	NNAGAAW	Deveau <i>et al.</i> 2008
<i>NmCas9</i>	<i>Neisseria meningitidis</i>	NNNNGATT	Zhang <i>et al.</i> , 2013
<i>SaCas9</i>	<i>Staphylococcus aureus</i>	NGRRT ou NGRRN	Friedland, 2015
<i>LbCpf1</i>	<i>Lachnospiraceae bacterium</i>	TTTV	Zetche <i>et al.</i> , 2015
<i>AacCas12b</i>	<i>Alicyclobacillus acidiphilus</i>	TTN	Shmakov <i>et al.</i> , 2015
<i>FnCas9</i>	<i>Francisella novicida</i>	NNGRRT	Hirano <i>et al.</i> , 2016
<i>AsCpf1</i>	<i>Acidaminococcus</i> sp.	TTTV	Zetche <i>et al.</i> , 2016
<i>CjCas9</i>	<i>Campylobacter jejuni</i>	NNNNRYAC	Kim <i>et al.</i> , 2017
Cas14	Arquea não cultivada	TTTA para dsDNA	Harrington <i>et al.</i> , 2018
<i>BhCas12b</i>	<i>Bacillus hisashii</i>	ATTN, TTTN e GTTN	Strecker <i>et al.</i> , 2019

O fato é que, quanto maior a especificidade da nuclease, menor a chance de que ocorram hibridizações RNA/DNA fora do sítio de interesse. Um bom exemplo disso é a *SaCas9*

projetada por Tran *et al.* (2020) chamada de microABE1744. Essa versão da SaCas9 possui uma especificidade muito maior quanto ao *on-target*, reduzindo significativamente os efeitos *off-target*. Além disso é uma das menores Cas9-ABE existentes, facilitando sua entrega através de vetores virais.

2 Resposta imune

2.1 Patógenos e sistema imune

Patógenos são fatores chave na ecologia e na evolução da vida como as conhecemos. Vírus em especial, devido a sua capacidade de agir como predadores ou facilitadores genéticos (DeFelippes & Villarreal, 2000). De modo que não é surpresa que organismos em geral tenham aprendido a conviver ou tenham desenvolvido formas de se defender de patógenos. Porém, vetores virais e proteínas que medeiam a edição gênica são provenientes de patógenos. Assim, é razoável que, quando apresentado a esses elementos, um organismo desenvolva uma reação imune contra esse "invasor".

Os conceitos a respeito de imunidade datam do século V a.C. através dos relatos de Thucydides sobre indivíduos que, após terem contato com uma doença, recuperaram-se e não a contraíram uma segunda vez (Kernbach *et al.*, 2018). Porém, foi apenas na China, no século XVI, que o primeiro relato de uma indução intencional de imunidade contra um vírus (varíola - endêmica na região à época) pode ser encontrado (Boylston, 2012).

O processo de expor pessoas saudáveis ao patógeno causador da varíola se espalhou pelo território asiático e ficou conhecido como "variolação" (inoculação) (Louten, 2016). A exposição era feita utilizando o conteúdo das lesões causadas pela doença sob a pele ou mais frequentemente, inserindo um pó de pústulas no nariz do indivíduo saudável (insuflação) (Boylston, 2012). Surgem assim os primeiros conceitos sobre desenvolvimento de imunidade.

Apesar disso, foi no século XIX que os conceitos de imunologia foram realmente inseridos no contexto da teoria científica, quando o estudo dos componentes moleculares que compõem o sistema imunológico se tornou central (Kaufmann, 2019). A partir desses estudos o sistema imune foi dividido em sistema imune inato, do qual faz parte o sistema complemento, tido como um sistema de resposta contra patógenos muito mais primitivo, por tratar-se de uma resposta rápida e inespecífica, que seu complementar em vertebrados, o sistema imune adaptativo ou adquirido (Cruvinel *et al.*, 2010).

Isso é importante, pois a maioria dos organismos utilizados como vetores virais para terapias gênicas ou mediadores de edição gênica são provenientes de patógenos conhecidos dos seres humanos. Desse modo, uma reação imune contra esses vetores e produtos não é totalmente inesperada. Assim, a maior questão aqui não é se uma reação imune vai ocorrer, mas onde, quando e em qual intensidade.

2.2 Resposta contra vetores e outros produtos

Os vetores virais utilizados para entrega de produtos gênicos necessários a terapias são, em geral, provenientes de patógenos humanos conhecidos: os adenovírus (AdV), que usualmente se instalam nas vias respiratórias humanas causando resfriados, bronquites e até mesmo pneumonias (Kajon & Lynch, 2016); ou os retrovírus, que além do HIV podem ser responsáveis por doenças hematológicas, ou neurodegenerativas (Weiss, 1987; Jeziorski *et al.*, 2016).

As proteínas utilizadas para edição gênica estão associadas às mesmas questões. O *Streptococcus pyogenes* é considerado altamente infeccioso e costuma ser o responsável por faringites (Moreno-Tores *et al.*, 2020). Já as infecções por *Staphylococcus aureus* costumam ser encontradas em nossa pele ou mucosa nasal, mesmo em pessoas saudáveis e estão associadas a infecções de pele menores (Blicharz, Rudnicka and Samochocki, 2019).

Os únicos que parecem fugir desse perfil são os vírus adenoassociados (AAV), que até então permanecem sem um perfil patológico definido, causando infecções assintomáticas em indivíduos imunocompetentes. E os *Streptococcus thermophilus*, que aparecem muito mais associados aos benefícios nas alergias alimentares e à saúde gastrointestinal (Markakiou *et al.*, 2020).

2.3 O impacto do sistema imune em terapia gênica

Os primeiros estudos com adenovírus humano (AdV) são de 1953 e foram caracterizados por dois grupos de pesquisa independentes. Já os vírus adenoassociados (AAV) foram descobertos como contaminante de uma preparação de adenovírus símio nos anos 1960 (Atchison *et al.*, 1965). E o primeiro retrovírus humano (HTLV-I) foi reportado em 1979 e confirmado em 1980 (Gallo, 2005). De lá para cá, diversas cepas desses tipos virais vêm sendo estudadas, especialmente para aplicações em terapia gênica.

Os chamados parasitas intracelulares obrigatórios, só são capazes de replicar seu próprio genoma quando encontram e infectam uma célula hospedeira (Summers, 2009). É exatamente essa característica que os torna tão interessantes para aplicação em terapias gênicas. Sua composição, baseada em proteínas organizadas em um capsídeo cuja função é proteger seu genoma, de RNA ou DNA (Summers, 2009), é tanto simples quanto atraente do ponto de vista bioquímico.

É a partir da infecção do hospedeiro que começa a duplicação do seu material genético e a síntese das suas proteínas através do uso do maquinário proteico da célula invadida (Actor, 2012). Adenovírus (AdV), Vírus Adenoassociados (AAV) e Retrovírus possuem composição genética diferente, mas seus mecanismos de replicação variam de forma bastante sutil (Bulcha *et al.*, 2021).

Os AAV pertencem ao gênero *Dependoparvovirus*, da família *Parvoviridae*. São pequenos, desprovidos de envelope, com uma única fita de DNA. Seu genoma possui cerca de 4.7 Kb. Seu capsídeo tem formato icosaédrico e seu genoma possui apenas três proteínas virais, com ao menos 12 serotipos naturais já descobertos (Asokan *et al.*, 2012). Essa diversidade os torna capazes de infectar uma gama de animais, nos quais se incluem os humanos.

As infecções por AAVs podem permanecer por tanto tempo quanto viver o organismo infectado (Berns and Muzyczka, 2017). Entretanto, como o próprio nome deixa claro, os adenoassociados são capazes de se replicar apenas na presença de fatores que os auxiliem, como outros vírus (Meier *et al.*, 2020). É o caso do HSV-1 e do citomegalovírus humano, HCMV. Sem a ajuda de outros fatores virais, os AAV entregam o seu genoma na célula infectada, que em pouco tempo é capaz de eliminá-lo (Wang, Tai and Gao, 2020). Entretanto, é sabido que alguns AAV mantêm sua expressão de forma mais duradoura nas células infectadas, associado ao fato de seu conteúdo genético assumir uma forma circular, conhecida como concatâmero (Arbuckle and Medveczky, 2011).

Os AdV são um grupo de vírus que, usualmente, estão associados a doenças respiratórias, oculares ou do trato intestinal (Lynch and Kajon, 2016). Há mais de 80 sorotipos de adenovírus capazes de infectar vertebrados, pássaros, peixes, répteis e anfíbios (Li *et al.* 2010), dos quais aproximadamente 50 são capazes de infectar humanos (Crystal, 2014). Com um formato icosaédrico conservado, o AdV tem seu genoma organizado em uma dupla fita de DNA, com tamanho de cerca de aproximadamente 8 kb. Ao contrário dos AAV, os AdV estão associados a conhecidos efeitos imunogênicos e toxicidade celular (Wang, Tai and Gao, 2019).

Tanto os vírus adenoassociados quanto os adenovírus, são incapazes de integrar seu próprio conteúdo genético no genoma do hospedeiro (Naso *et al.*, 2017), o que faz com que sua

expressão seja transiente. Os retrovírus, por outro lado, possuem essa habilidade para que possam iniciar o processo de replicação de seu conteúdo genético, organizado em uma fita simples de DNA, com cerca de 10 kb de tamanho (Coffin *et al.*, 1997). Por estarem inseridos no genoma, os retrovírus aproveitam-se de forma muito mais eficiente da maquinaria celular do hospedeiro, fazendo com que sua expressão permaneça estável ao longo do tempo (Ciuffi A, 2016).

Estes mecanismos são interessantes quando pensamos em tratamentos para doenças que dependem da expressão de um produto gênico que as próprias células do paciente são incapazes de produzir. Entretanto, não é esperado que as células do hospedeiro produzam proteínas diferentes das habituais, sejam elas expressas de maneira permanente ou transitória. Além disso, qualquer vetor viral pode ser entendido pelo sistema imune como algo a ser combatido, independente da razão pela qual o vetor foi administrado.

Mantendo isso em mente, se faz importante entender onde estão os pontos críticos à terapia gênica com relação ao sistema imune, uma vez que passamos a inserir componentes associados a patógenos no organismo humano. Assim, o primeiro artigo (segundo capítulo) desta tese faz uma revisão que tem como principal objetivo reunir informações a respeito das terapias gênicas existentes e os pontos onde o sistema imune pode atuar.

2.4 A diversidade dos HLAs e edição gênica

Para manter a integridade do nosso organismo é essencial que nosso sistema imune seja capaz de distinguir entre aquilo que faz parte de nosso corpo (próprio) e aquilo que pode, eventualmente, fazer parte de um agente externo (não-próprio). No início da evolução, organismos multicelulares simples desenvolveram um sistema de defesa ativado pela detecção de moléculas típicas e estruturas químicas associadas a patógenos.

Entretanto, o sistema imune faz muito mais do que apenas a diferenciação entre o próprio e o não-próprio (Jiang *et al.*, 2009). Ele protege nosso corpo contra certos patógenos, ao mesmo tempo em que aprende a viver em simbiose com outros, de forma a nos beneficiar (vide a microbiota) (Zheng, Liwinski e Elinav, 2020). Também nos defende diariamente contra células cujos genes apresentam perfil de expressão diferenciado (Orrù *et al.*, 2013). Cada uma dessas habilidades culmina no controle da infecção através de uma resposta inflamatória.

Todavia, essas respostas não acontecem todas da mesma maneira em qualquer indivíduo. Ainda que nosso sistema imune seja relativamente estável ao longo da nossa vida,

ele é bastante diverso entre indivíduos. Essas variações vêm não apenas das influências patogênicas ou simbióticas dos microrganismos e processos de adaptação, como são consequência de influências hereditárias. E quando pensamos nessas variações do ponto de vista genético, a maior fonte apontada são os genes *HLA* (*human leukocyte antigen*).

A região gênica do antígeno leucocitário humano (HLA) é considerada uma das mais polimórficas do genoma (Bravo-Egana *et al.*, 2021). A maioria dos genes nesta região é responsável pela codificação de genes envolvidos nas vias de apresentação de antígeno (Robinson *et al.*, 2013). Em função de seu importante papel para o sistema imune, os HLAs vêm sendo utilizados para o entendimento da diversidade genética humana e da genética de populações (Meyer *et al.*, 2018). Já em 2005, Prugnolle *et al.* demonstravam a importância da compreensão da diversidade genética dos HLAs no entendimento das pressões seletivas relacionadas à resposta contra patógenos.

Mantendo esse ponto de observação, se faz importante entender não apenas como funciona a relação evolutiva entre HLAs e patógenos naturais, mas também como se dá essa evolução quando passamos a inserir componentes associados a patógenos no organismo humano como forma de tratamento de doenças. Assim, o artigo correspondente ao terceiro capítulo deste trabalho, é um primeiro esforço para compreender como os diferentes perfis de alelos de HLA na população afetariam a edição gênica mediada pela proteína Cas9 sob a ótica da resposta imune.

3 Efeitos *off-target*

3.1 Predição de *off-targets*

São chamados de efeitos *off-target* toda a edição gênica que ocorre fora do DNA alvo. Muito embora se acredite que a especificidade da Cas9 seja controlada pelo sgRNA, a clivagem de sequências muito similares pode ocorrer com até 5 incompatibilidades distais a PAM (Mali *et al.*, 2013; Hsu *et al.*, 2013; Fu *et al.*, 2013). Assim a detecção de locais com sequências muito semelhantes aos guias sgRNA se torna um desafio no campo de edição gênica.

Um dos métodos experimentais para a detecção de *off-targets* é um ensaio de endonuclease T7, uma enzima seletiva de estrutura capaz de detectar deformidades estruturais na dupla fita de DNA (Sentmanat *et al.*, 2018). Infelizmente, este ensaio apresenta baixa sensibilidade (não detecta alvos com frequência menor que 0,01) (Cho *et al.*, 2013), além de

não ser um ensaio custo-efetivo em larga escala (Kim *et al.*, 2009). Outro método experimental para estudo de *off-targets* envolve NGS (sequenciamento de nova geração), onde são medidas as mutações geradas por edição fora do alvo, em frequências de 0,01 a 0,1% (Cho et al, 2014). Ensaio de ChIP-seq (sequenciamento por imunoprecipitação de cromatina) também são uma forma de identificar sítios de ligação fora do alvo para sgRNAs complexados com Cas9 inativo (dCas9) (Singh *et al.*, 2015). Além disso, como método de prevenção de *off-targets* em aplicações de bancada, muitas ferramentas de predição *in silico* têm sido desenvolvidas.

Há consenso sobre a necessidade de desenvolver mecanismos capazes de lidar com essa questão, pois os *off-targets* gerados pelo sistema CRISPR/Cas9 estão associados a efeitos deletérios que podem ofuscar a aplicação segura de mecanismos de edição (Liang, 2015). Por causa disso, o desenvolvimento de ferramentas de bioinformática aumentou em uma tentativa de resolver essa questão. No entanto, cada programa disponível considera um conjunto diferente de parâmetros e retorna diferentes resultados para uma mesma sequência avaliada. Além disso, programas web têm uma limitação inata, uma vez que assumem que as sequências *off-target* estão intimamente relacionadas com o sítio alvo, o que acarreta perder sítios fora do alvo prejudiciais com menos similaridade de sequência.

Outro ponto a ser observado é que a ocorrência de *off-targets* não se trata meramente de um pareamento perfeito entre as sequências do sgRNA e o DNA. Esse é um tema que depende também das bases estruturais do pareamento e estabilidade de interação com a proteína Cas9. O trabalho de Bravo *et al.* (2022) descreve essa dinâmica, onde *mismatches* distais a PAM são estabilizados devido a uma reorganização de um loop no domínio RuvC da proteína que impediria a clivagem do DNA. Entretanto, quando os *mismatches* ocorrem nas posições 18 a 20, a proteína ignora os *mismatches* prosseguindo com a clivagem do DNA.

Desse modo os trabalhos que compõe o quarto e quinto capítulos desta tese mostram como é possível estudar o comportamento estrutural da proteína Cas9 empregando apenas ferramentas disponíveis online, no caso do primeiro; e no segundo, a necessidade de uma análise mais profunda dos programas para predição de *off-targets* em edição gênica com CRISPR/Cas9. Estes dois trabalhos, ainda se encontram em desenvolvimento e têm como objetivo final o desenvolvimento de ferramentas em ambiente Google Colaboratory, para comparação de resultados decorrente dos preditores testados e avaliação dos resultados provenientes de estudos estruturais desses *off-targets*, com a intenção de auxiliar pesquisadores na tomada de decisão sobre as guias a serem utilizadas em seus experimentos.

2.1 Objetivo Geral

O objetivo deste trabalho foi avaliar dois aspectos em aberto com relação à terapia e edição gênica: a resposta imune e os efeitos *off-target*.

2.2 Objetivos específicos

- Apresentar o estado da arte em relação ao impacto da resposta imune em terapia e edição gênica.
- Compreender como as variações de moléculas HLA em diferentes grupos populacionais afetam a apresentação de antígenos de Cas9.
- Expor as diferenças existentes no pareamento da Cas9 com o DNA genômico em regiões *on* e *off-target* através de análises estruturais do complexo proteína/DNA-RNA.
- Comparar preditores de efeitos *off-target* quanto à sua capacidade de acerto frente a um conjunto de dados experimentais.

Capítulo 2

“Protection is not always a good thing: the immune system’s impact on gene therapy”

Artigo *in extenso* aceito pela Revista *Genetics and Molecular Biology*

Protection is not always a good thing: the immune system's impact on gene therapy

Martiel Vaz de Freitas^{1,2,4}[0000-0003-4178-6604]; Lariane Frâncio^{1,2}[0000-0002-4283-3137]; Laura Haleva³[0000-0003-2428-264X]; Ursula da Silveira Matte^{1,2,4,5}[0000-0003-4977-6662].

1 Laboratório Células Tecidos e Genes, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

2 Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

3 Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

4 Núcleo de Bioinformática Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil.

5 Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Corresponding Author: Ursula da Silveira Hospital de Clínicas de Porto Alegre, Laboratório de células, tecidos e genes, Rua Ramiro Barcelos, 2350, 90035-903, Porto Alegre, RS, Brazil.

E-mail: umatte@hcpa.edu.br.

Short running title: Gene therapy and immune system

Abstract

There are many clinical trials underway for the development of gene therapies, and some have resulted in gene therapy products being commercially approved already. Significant progress was made to develop safer and more effective strategies to deliver and regulate genetic products. An unsolved aspect is the immune system, which can affect the efficiency of gene therapy in different ways. Here we present an overview of approved gene therapy products and the immune response elicited by gene delivery systems. These include responses against the vector or its content after delivery and against the product of the corrected gene. Strategies to overcome the hurdles include hiding the vector or/and the transgene product from the immune system and hiding the immune system from the vector/transgene product. Combining different strategies, such as patient screening and intelligent vector design, gene therapy is set to make a difference in the life of patients with severe genetic diseases.

Keywords: gene delivering, gene therapy, immune response

O intervalo das páginas 32 a 74 não está disponível para leitura.

Capítulo 3

“Small enemies: how HLA variation across populations may interfere in Cas9-mediated gene editing”

Artigo (*in extenso*) formatado para Molecular Therapy-Methods & Clinical Development

Small enemies: how HLA variation across populations may interfere in Cas9-mediated gene editing

Martiele Vaz de Freitas^{1,2,3}; Lariane Frâncio^{1,2}; Ursula Matte^{1,2,3,4}.

1 Laboratório Células Tecidos e Genes, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

2 Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

3 Núcleo de Bioinformática Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil.

4 Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Running Title: HLA variation interfere in gene editing

Corresponding author: umatte@hcpa.edu.br

Keywords: CRISPR/Cas9, Gene editing, HLA, Immune response, Populational analysis.

Abstract

Most studies about immune tolerance to Cas9 proteins focus on antibody response. Few works cover the possibility of the immune response through peptide presentation by class I HLA alleles found in different populations. Using well-known tools for epitope prediction, we found 178 epitopes of the Cas9 protein from the three most used organisms (*S. pyogenes*, *S. aureus*, and *S. thermophilus*) in gene editing with the CRISPR systems, considering only the highly probably presented peptides. Our results show that several epitopes can be generated and would predictably elicit an immune response when presented by specific HLA alleles common to the three chosen ethnic groups.

O intervalo das páginas 77 a 99 não está disponível para leitura.

Capítulo 4

“Online tools for structural analysis of Cas9 off-targets occurrence”

Trabalho em andamento.

Online tools for structural analysis of Cas9 off-targets occurrence

Authors: Freitas, M. V.^{1,2,3}; Borges, P.⁴; Carneiro, P. B.^{1,2,3}; Matte, U.^{1,2,3,5}

1 Núcleo de Bioinformática, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

2 Centro de Terapia Gênica, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

3 Post-Graduation Program on Genetics and Molecular Biology, UFRGS, 91501-970 Porto Alegre, RS, Brazil.

4

5 Department of Genetics, UFRGS, 91501-970 Porto Alegre, RS, Brazil.

Abstract

The CRISPR RNAs and the Cas9 proteins have the potential to change the treatment of diseases. However, the same system can cause dangerous side effects, especially off-targets. These cuts occur off the target DNA, generating disturbances in the proper functioning of the cell's mechanisms, even leading to cell death. However, most works about off-targets take into account only the DNA and sgRNA sequences, expecting that similarity in these sequences answer how many off-targets a given sgRNA sequence may be paired with. Structural features, however, must also be taken into account. Here we describe a pipeline using online tools that any user might use and shows the importance of considering structural features in off-target analysis.

O intervalo das páginas 102 a 110 não está disponível para leitura.

Capítulo 5

“Off-target discovering tools aren't all the same: understanding commonly used predictive tools using experimental data”

Trabalho em andamento.

Off-target discovering tools aren't all the same: understanding commonly used predictive tools using experimental data

Authors: Freitas, M. V.^{1,2,3}; Barcelos, P. C.^{1,2,3}; Matte, U.^{1,2,3,4}

1 Núcleo de Bioinformática, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

2 Centro de Terapia Gênica, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

3 Post-Graduation Program on Genetics and Molecular Biology, UFRGS, 91501-970 Porto Alegre, RS, Brazil.

4 Department of Genetics, UFRGS, 91501-970 Porto Alegre, RS, Brazil.

Corresponding Author: Ursula Matte. Hospital de Clínicas de Porto Alegre, Laboratório de células, tecidos e genes, Rua Ramiro Barcelos, 2350, 90035-903, Porto Alegre, RS, Brazil.

E-mail: umatte@hcpa.edu.br

ORCID ID: 0000-0003-4977-6662

Abstract

Several tools for off-target prediction in gene editing with CRISPR are available. However, they rely on different alignment techniques, and some consider additional information such as chromatin accessibility, GC content, or if the target is intronic or exonic. There are also various forms to compute scores and return the results. So, how to choose the best one? This work describes an attempt to understand which tool is more accurate against experimentally-validated off-targets.

O intervalo das páginas 113 a 131 não está disponível para leitura.

Capítulo 6

Discussão

O principal objetivo deste trabalho foi avaliar dois aspectos importantes relacionados à terapia gênica abertos à proposição de melhores descrições de suas bases moleculares. Ainda que intimamente relacionado com terapia gênicas, aqui, o principal alvo de investigação está relacionado à edição gênica. Assim, buscamos compreender as limitações, sugerir pontos onde mais trabalhos são necessários, além de descrever métodos de análise factíveis e de fácil acesso.

Na vanguarda da ciência médica, atualmente as terapias gênicas oferecem a possibilidade de tratamento de doenças variadas. Onde antes os tratamentos apenas controlavam sintomas na expectativa de tornar a doença algo manejável, atualmente falamos de pesquisas delineadas para oferecer maiores chances de cura em poucas ou apenas uma aplicação. Para que tamanha revolução acontecesse, pesquisadores tiraram proveito das habilidades dos vírus em entrar na célula com alta especificidade, tornando-os poderosos veículos para a entrega de genes.

O primeiro ensaio clínico de terapia gênica foi realizado em crianças portadoras de imunodeficiência combinada grave (SCID). O tratamento consistiu na transferência do gene da adenosina deaminase para linfócitos isolados desses pacientes através de um retrovírus (Blaese *et al.*, 1995). Outro ensaio clínico na mesma época foi realizado com melanoma avançado, utilizando linfócitos infiltrantes de tumor modificados por transdução retroviral (Rosenberg *et al.*, 1990).

Foi devido ao grande aumento de ensaios clínicos envolvendo vetores virais e os relatos de eventos adversos graves nesses ensaios que, em janeiro de 2020, o FDA (órgão de regulamentação americano equivalente a ANVISA) publicou uma revisão das diretrizes regulatórias aplicadas à terapia celular e gênica nos EUA. Nesse documento, explicitam alterações relacionadas à análise de vetores virais, incluindo testes específicos para impurezas, titulação, replicação, capacidade de infecção. O controle da fabricação de vetores virais tem como intuito quantificar a heterogeneidade do produto desenvolvido, para que sua eficácia e sua segurança ante a aplicação em seres humanos sejam garantidas.

Assim como o próprio campo das terapias gênicas evoluiu, tudo que está relacionado a melhora e manutenção da segurança evoluiu também. Métodos moleculares como PCR e NGS são empregados para confirmar a identidade dos vetores utilizados nas terapias. Ensaio ELISA, Western blot ou espectrometria de massas são utilizados na caracterização das proteínas dos

capsídeos virais (Gardner *et al.*, 2022). Métodos relativamente recentes como AAV-ID, descrito por Pacouret *et al.* (2017), emprega termoestabilidade diferenciada para distinguir entre os sorotipos de AAVs empregando HILIC (cromatografia de interação hidrofílica).

O desenvolvimento de vetores virais seguros e eficientes na entrega dos produtos terapêuticos é um dos maiores desafios da área, uma vez que o método ideal deve garantir, não apenas a entrega do transgene, como deve ser preciso o bastante para afetar apenas as células alvo. Além disso, não deve ser, por si só, um causador de efeitos colaterais, inflamatórios ou citotóxicos. E ambos os aspectos estão relacionados a uma questão persistente na terapia gênica: a resposta imune.

Como não ser visto? Essa é a pergunta que direciona diversos pesquisadores na área de terapia gênica. O primeiro trabalho desta tese revisou todos os pontos onde o sistema imune merece atenção por ser visto como um fator limitante no sucesso das terapias gênicas. E, como esconder produtos terapêuticos de um sistema sempre presente é um desafio tão grande que, por vezes, desfechos inesperados ocorrem.

O primeiro conjunto de terapias gênicas iniciadas no princípio da década de 1990, teve seu andamento interrompido após a morte de Jesse Gelsinger em 1999 após uma intensa reação imune contra o vetor que transportava o produto terapêutico (Sibbald, 2001). Além disso, quatro outros pacientes desenvolveram leucemia linfoblástica aguda após a infusão da terapia (Wilson, 2010). Somente após esses eventos a comunidade científica percebeu a intensidade dos riscos relacionados à imunogenicidade dos vetores virais e de seus padrões de integração.

Buscou-se compreender mais sobre os adenovírus utilizados nos ensaios clínicos, em busca de informações a respeito de sua toxicidade. Capazes de infectar uma ampla gama de células, como por exemplo, epiteliais, respiratórias, hepatócitos, células da glia e até mesmo macrófagos, essa diversidade de alvos explica por que cerca de 55% dos indivíduos testados nos EUA possuem imunoglobulinas contra adenovírus tipo 5 (AdV5), além dos sorotipos 2, 7 e 17 de adenovírus humano (Chirmule *et al.*, 1999; Zabner *et al.*, 1999).

Ao longo desses estudos, outros já haviam sido iniciados com intuito de inserir modificações nos adenovírus para que eles se tornassem mais apropriados à terapia gênica. Primeiro, os genes E1A e E1B foram excluídos do genoma viral, na chamada primeira geração de vetores adenovirais recombinantes deficientes em replicação (Zhang, 1999). Devido à persistência de respostas imune celular e humoral, surgiu a segunda geração desses vetores, caracterizada por modificações ou deleções adicionais a E1, nos genes E2A, E2B e/ou E4 (Wang & Finer, 1996). E finalmente, chegou-se aos chamados "gutless vectors", com deleções de todos os produtos gênicos (Chen *et al.*, 1997).

Entretanto, apesar de todas essas modificações, os vírus adenoassociados (AAV) se tornaram muito mais interessantes sob a óptica dos efeitos adversos, uma vez que são conhecidos por não causar nenhuma doença em humanos. Isso, entretanto, não os torna livres de questões sensíveis em terapia gênica, como apresentado no primeiro artigo deste trabalho.

Enquanto os AdV e AAV já haviam percorrido boa parte do início do caminho em terapia gênica, o primeiro estudo demonstrando que retrovírus são independentes de sequência ao selecionarem seus sítios genômicos foi publicado em 2002 (Schröder *et al.*, 2002). Divididos em sete gêneros, os retrovírus têm seus perfis de integração bem caracterizados para aqueles organismos dos gêneros lentivirus, spuma-retrovirus, alpha-retrovírus e gama-retrovírus (Poletti and Mavilio, 2017). Dentre os retrovírus podemos citar o HIV e o MLV como os mais relevantes em aplicações clínicas, sendo o HIV frequentemente utilizado como vetores em terapias gênicas que envolvem a necessidade de integração em células quiescentes.

De qualquer modo, por se tratarem de organismos externos ao corpo humano, a chance de que um vetor viral seja eliminado ao se deparar com as células do sistema imune é muito alta. E, apesar de ser uma barreira importante, é apenas ao tentar transpô-la, que somos capazes de reconhecer quais métodos são eficientes.

Nesse sentido, torna-se importante mencionar os vetores não-virais. Estes apresentam vantagens não apenas com relação a sua baixa imunogenicidade, mas também devido a sua composição bioquímica diminuindo efeitos citotóxicos. Usualmente compostos de lipídios, são muito melhor tolerados pelo organismo. Dentre os possíveis métodos não-virais podemos citar o DNA nu, os polímeros e lipídios catiônicos e os complexos conjugados (Sung and Kim, 2019). Uma vez que não há uma forte relação entre o sistema imune e esse vetores, o maior desafio está na eficiência da transfecção, que em muitas vezes ocorre via endocitose, como no caso daqueles baseados em lipossomas.

Todos esses vetores podem também ser empregados na entrega dos sistemas CRISPR para edição gênica. Como apresentado no segundo capítulo desta tese, a proteína responsável por promover a clivagem do DNA pode, por si só, desencadear uma resposta imune contra células já editadas. Chama nossa atenção o fato de a consolidação da edição gênica ter se dado por volta de 2013 (Ran *et al.*, 2013) e apenas cerca de seis anos após, o sistema imune chamar algum tipo de atenção por parte dos pesquisadores, apesar do histórico já existente entre os agentes externos (virais) das terapias gênicas e resposta.

Assim, uma vez que estes estudos estão no seus primeiros passos, faz sentido questionar se edição gênica pode ser uma ferramenta tão generalista quanto gostaríamos, ou se deveríamos compreender um pouco mais os organismos/indivíduos nos quais pretendemos

aplicar esta técnica. Em especial, como o sistema imune de diferentes pessoas responderia quando se deparasse com os subprodutos da proteína.

Reconhecer que a diversidade de alelos HLA pode interferir no sucesso da edição, parece algo difícil. Entretanto, é sabido que a seleção dirigida por patógeno implica que determinados alelos HLA específicos são mais eficientes na apresentação de antígenos de alguns patógenos em comparação a outros (Brandt *et al.*, 2018).

Uma evidência disso são as variantes de HLA-B associadas à progressão para AIDS após a infecção por HIV (*The International HIV Controllers Study* 2010). Ou o estudo de Hu e colaboradores (2013), onde demonstram que variações em HLA-C, HLA-DP e HLA-DQ estão associados ao *clearance* do vírus da hepatite B em alguns indivíduos.

Assim, não é estranho que a predição de peptídeos, apresentada nesse trabalho levando em consideração a origem étnica da população, mostra uma maior diferenciação entre os três grupos apresentados. Não apenas em diversidade de HLAs, mas também no perfil de peptídeos de Cas9 possivelmente imunogênicos, de acordo com esses HLAs.

Entretanto, uma importante limitação se encontra na disponibilidade dos dados nos bancos públicos. Apesar de muitos dados a respeito da diversidade dos HLAs nas populações estarem disponíveis em bancos como o *Allele Frequency Net Database* (AFND) e *the international ImMunoGeneTics information system* (IMGT), apenas os mais recentes apresentam qualidade suficiente para que se confie no dado considerando dois ou mais campos de resolução (HLA-A*:02-01 por exemplo; onde 02 é um campo e 01 é o segundo campo e assim por diante). Uma maior confiabilidade permitiria uma maior diferenciação entre subgrupos específicos das populações incluídas nas análises, uma vez que é possível que dentro de uma mesma população alguns grupos tenham características distintas.

Ao longo dos anos, as predições em ligação e afinidade de peptídeos tornou-se mais confiável em função do uso de dados experimentais gerados através de ensaios de ligação e espectrometria de massas (Nguyen, Szeto and Gras, 2021). Além disso, a disponibilidade de estruturas cristalográficas dos HLAs auxilia na compreensão de outras dinâmicas necessárias ao desenvolvimento de resposta.

Por exemplo, é possível investigar a fundo processos associados à reatividade cruzada. Eventos dessa natureza ocorrem quando um determinado TCR gera resposta contra epitopos distintos, tendo confrontado previamente apenas um deles. Assim, uma infecção anterior por um determinado patógeno poderia desencadear uma resposta contra um segundo patógeno com o qual o organismo não teria tido contato até o momento (Clute *et al.*, 2005).

Considerando que seres humanos são expostos a diferentes linhagens de *Streptococcus* ao longo da vida, é razoável considerar que memória contra patógenos de infecções anteriores a um procedimento empregando Cas9 possa desencadear uma resposta imune contra peptídeos dessa proteína. A ideia de memória associada à resposta em edição gênica é tão razoável que, em 2018, Wagner e colaboradores mediram a prevalência de células T reativas contra *Streptococcus pyogenes* em indivíduos saudáveis nos EUA. No estudo, encontraram uma alta frequência de células T efetoras reativas a SpCas9, caracterizando memória contra a proteína capaz de mitigar os efeitos de terapias mediadas por CRISPR nessa população.

Esse raciocínio não fica restrito apenas ao caso da edição gênica, mas também pode ser estendido aos vetores utilizados em terapias gênicas. Uma solução para essa questão, uma vez que a imunossupressão generalizada pode ter efeitos indesejados, como aumento da suscetibilidade a infecções ou surgimento de tumores, é a expressão transiente da proteína. Isso diminuiria o tempo de exposição da proteína, diminuindo as chances de seus subprodutos serem apresentados na superfície celular. Do mesmo modo, melhorias na eficiência dos processos de transfecção diminuiria a quantidade de proteínas necessárias para promover a edição, implicando menor quantidade de proteínas sendo degradadas no interior da célula.

Assim como o sistema imune permanece em grande parte uma caixa preta, o outro tópico abordado ao longo desse trabalho também oferece desafios. A ocorrência de *off-targets* em edição gênica permanece uma preocupação importante. Clivagens promovidas fora do DNA alvo podem levar a efeitos indesejados, especialmente aqueles relacionados à geração de mutações associadas a doenças.

Assim parece importante considerar que diversos aspectos pouco discutidos influenciam na ocorrência de *off-target*:

- o tipo celular escolhido para edição;
- o sgRNA desenhado para esse fim;
- o tipo de preditor de *off-targets* que está sendo utilizado;
- o tipo de sequenciamento empregado para detectar as regiões *off-target*.

Os tipos celulares escolhidos para o desenvolvimento de estudos de edição gênica têm um grande impacto no sucesso técnico, e isso envolve também a questão da edição em sítios indesejados. A eficiência de edição descreve essencialmente a porcentagem de células que foram editadas com sucesso em meio de cultura. Uma eficiência de edição inferior a 100% não é exatamente ruim, mas implica uma interpretação cuidadosa dos resultados. Efeitos muito sutis relacionados à edição podem ficar perdidos em meio a células que não foram editadas.

A maneira mais frequente pela qual pesquisadores buscam evitar isso, é a inclusão de marcadores de seleção, em geral, antibióticos, permitindo assim a seleção das células que foram transfectadas com sucesso. Um raciocínio oposto pode ser levado em consideração quando pensamos em *off-targets*. Ainda que possamos identificar quais células foram editadas, não há meios que permitam controlar a dinâmica de transfecção. Desse modo, a transfecção de diversas proteínas na célula pode acarretar uma maior ocorrência de edições fora do sítio alvo como mencionado anteriormente.

O principal mecanismo de transfecção em células envolve a afinidade pelos receptores da membrana celular. Se adicionarmos esse conhecimento a respeito das diferenças de transfecção em determinadas linhagens/tipos celulares, podemos inferir que enquanto a possibilidade de ocorrências de *off-targets* diminui em células imaturas (tronco primárias) devido às baixas taxas de transfecção (Papapetrou *et al.*, 2005), ela aumenta em células maduras de diversos tecidos, onde o potencial de transfecção aumenta (Yuan *et al.*, 2022).

Além disso, dois outros fatores relacionados ao tipo celular podem influenciar a ocorrência de *off-targets*. O primeiro deles é a possibilidade de os mecanismos de reparo atuarem com diferente eficiência em diferentes tipos celulares, em especial em células embrionárias (Tichy & Stambrook, 2008). Kosicki *et al.* (2018) relatam taxas significativas de mutagênese devido a efeitos *off-target*, ocasionando grandes deleções e rearranjos genômicos em células tronco e hematopoiéticas progenitoras de camundongos, além de células de linhagens diferenciadas provenientes de humanos. Adicionalmente, Wrona e colaboradores (2020) apresentaram que, durante a correção de NCF1 mutado a presença de pseudogenes no mesmo cromossomo resultou em deleções cromossômicas potencialmente prejudiciais entre o *locus* alvo e um modelo celular deficiente em p47phox. Além disso, alterações na cromatina podem facilitar (ou dificultar) o acesso a regiões *off-target* de maneira desigual entre diferentes tipos celulares (Daer *et al.*, 2017).

O desenho do sgRNA é outro fator que pode interferir na ocorrência de edições gênicas indesejadas. Ainda que muito das limitações nesse sentido estejam associadas à necessidade do uso de guias muito específicas relacionadas a determinadas doenças, é possível desenhar guias com as mais diversas PAM (essenciais para o correto direcionamento da clivagem) aumentando ou diminuindo sua especificidade na localização do alvo. Isso implica o conhecimento de outras estruturas proteicas capazes de clivar o DNA e, claro, mais conhecimento sobre sua imunogenicidade.

A compreensão de porque essas edições fora do alvo acontecem é tão importante que métodos experimentais e preditivos têm sido desenvolvidos na tentativa de entender, quantificar

e diminuir a ocorrência de *off-targets*. O trabalho mais recente nesse sentido (Bravo *et al.*, 2022) desenvolveu uma técnica chamada "determinação de estrutura guiada por cinética" (*kinetics-guided structure determination*, em inglês). A técnica utiliza microscopia microeletrônica para visualizar a proteína Cas9 em ação enquanto interage com uma região de *off-target*.

Neste trabalho os autores descrevem que ao encontrar *mismatch* na posição distal a PAM, mais especificamente entre as posições 18 e 20, ao invés de a proteína desistir do processo e buscar um novo sítio, uma alteração estrutural ocorre e resulta no aumento da estabilidade da proteína. Isso faz com que a proteína permaneça por mais tempo em contato com o DNA, promovendo a clivagem. Sem essa estabilidade adquirida, a Cas9 não executa todas as outras etapas necessárias para clivar o DNA.

Foi através dessa descoberta que a SuperFi-Cas9 foi desenhada. Previamente, outras Cas9 já haviam sido construídas com a intenção de aumentar atividade e diminuir a ocorrência de *off-targets*: eSpCas9 (Slaymaker *et al.*, 2016), Cas9-HF1 (Kleinstiver *et al.*, 2016), evoCas9 (Casini *et al.*, 2018), Sniper-Cas9 (Lee *et al.*, 2018), e HypaCas9 (Ikeda *et al.*, 2019). Essas variantes da proteína contêm mutações que visam interromper a interação proteína-DNA ou RNA-DNA, desestabilizando os domínios proteicos de interação com o DNA. Isso resulta no impedimento da principal modificação estrutural no domínio HNH da nuclease, essencial para promoção da clivagem.

A SuperFi-Cas9 foi desenhada de modo a afastar do DNA a porção proteica associada a essa estabilização que ocorre a favor dos *mismatch*, diminuindo a ocorrência de *off-targets*. O resultado, de acordo com os autores, é uma proteína capaz de clivar o sítio alvo com a mesma eficiência da Cas9 natural, mas com muito menos *off-targets*. Isso é certamente algo que não poderíamos ver nas análises descritas no trabalho apresentado no capítulo 4 desta tese, e que nunca foram notadas antes em trabalhos que envolvem simulações. O fato de usarmos dados estáticos provenientes dos bancos de dados e com ferramentas implementadas para diversas finalidades diminui muito a possibilidade de descobertas desta natureza.

Mesmo assim, é muito importante a descrição do pipeline sugerido. Assim pesquisadores com pouco ou nenhum conhecimento em programação também podem investigar seus alvos de pesquisa. É importante salientar que apenas a predição dos *off-targets* relacionados ao sgRNA da estrutura escolhida foram feitas utilizando linha de comando. Entretanto, isso também poderia ter sido feito através de qualquer ferramenta online disponível para essa finalidade, não inviabilizando o prosseguimento da análise.

Neste trabalho, um dos objetivos era compreender como as alterações energéticas decorrentes do mau pareamento entre o sgRNA e o DNA fora do alvo influenciam a ocorrência

de *off-targets*. A escolha por ferramentas online surge como um desafio devido ao tamanho da proteína e pela necessidade de que se descrevam métodos em bioinformática que considerem o potencial das ferramentas já existentes para análises estruturais.

As ferramentas foram escolhidas levando em consideração sua disponibilidade e periodicidade de atualização, além do suporte às estruturas envolvidas. O Web 3DNA (Li *et al.*, 2019), por exemplo, foi atualizado recentemente para que fosse mais responsivo, rápido e intuitivo. A maioria dos programas disponíveis para modelagem de DNA não permitem o uso de moldes cristalográficos ou não lidam bem com a resolução de pareamentos não canônicos, nos casos em que são permitidos. Em nenhum dos outros programas testados de forma adicional obtivemos os outputs desejado, fosse para modelagem: make-na (REF) e 3D-Dart (REF); ou análise dos resultados: Curves+ (Blanchet *et al.*, 2011).

Mesmo a ferramenta utilizada relata que essa é uma questão que ainda precisa ser melhorada. Entretanto, a facilidade de usar um molde proveniente de uma estrutura cristalográfica, foi suficiente para solucionar o problema de gerar um complexo DNA/RNA com pareamentos não usuais. Ao invés de ter que resolver as duas estruturas e ainda manter rigorosamente os cálculos necessários para fornecer um output confiável, o software precisou resolver parte do problema, modelando apenas as estruturas de DNA que nos interessavam.

A escolha da ferramenta para a docagem do DNA na estrutura sgRNA:Cas9 envolveu testes de diversas ferramentas que dessem suporte a estruturas de ácidos nucleicos. Dentre todas as ferramentas avaliadas, HDOCK (Yan *et al.*, 2020) foi a única a finalizar a docagem teste com sucesso. A docagem teste consistia de dois inputs: um arquivo contendo apenas a proteína, e outro arquivo contendo o ligante DNA/RNA. Para visualização das superfícies e cálculo do potencial eletrostático, duas ferramentas diferentes foram avaliadas, Pymol (Schrödinger and DeLano, 2020) e UCSF Chimera (Pettersen *et al.*, 2004). Ambas forneceram os mesmo resultados, com poucas variações entre as visualizações.

Os resultados descritos neste trabalho são parciais. Até aqui é possível perceber que *mismatch* entre o DNA e o sgRNA afetam a energia da estrutura. No entanto, sem análises adicionais, ainda não é possível inferir qual o impacto dessas alterações na dinâmica da proteína. Sabemos, entretanto, que não há proposta que leve em consideração ferramentas web já publicadas direcionadas a análise das interações entre a Cas9 e seus substratos descritos na literatura.

Tão importante quanto tentar entender a ocorrência de *mismatch*, é tentar prevêê-los. O processo de localização de possíveis locais de edição que vão além daquele desejado é inato ao

processo de design de guias eficientes. Não surpreende, assim, que ao longo dos anos diversas ferramentas tenham sido desenvolvidas para prever sítios indesejados de edição.

As ferramentas de predição se baseiam em regras que foram sendo descobertas conforme o uso da ferramenta expandia. Majoritariamente essas ferramentas se baseiam na similaridade dos sítios *off-target* com o sítio alvo, empregando ferramentas bem estabelecidas para alinhamento de sequências como método de busca desses locais. Entretanto, nem todas as ferramentas empregadas têm um bom desempenho, diferindo enormemente na quantidade de resultados fornecidos para um mesmo input.

Doench e colaboradores (2016), por exemplo, descrevem em seu trabalho que Bowtie não possui capacidade de identificar sítios fora do alvo com mais de duas incompatibilidades. De forma que, uma boa solução para essa questão seria o emprego da ferramenta CAS-OFFinder (Bae, Park and Kim, 2014) nas predições. Outras ferramentas disponíveis, não abordadas no trabalho apresentado, como CRISPR-DO (Ma et al., 2016) e CRISPOR (Concordet and Haeussler, 2018) utilizam o algoritmo de Burrow-Wheelers (BWA) (Li and Durbin, 2009) para buscar sítios no genoma passíveis de serem editados, tornando sua capacidade de predição superior.

Muito precisa ser feito com relação a avaliação de *off-targets in silico*, uma vez que não há informações acerca das características que devem ser levadas em consideração no momento da predição de *off-targets*. Se similaridade, acessibilidade a cromatina, conservação de PAM, número de *mismatch* distais ou regiões exônicas, não há consenso. Não há consenso nem mesmo sobre quais ferramentas devemos utilizar para prever *off-targets*. Pode-se optar por ferramentas que foram desenhadas baseadas nas regras que foram sendo descobertas a respeito do posicionamento dos *mismatch* e seu impacto, ou então por ferramentas baseadas em *machine learning* que levam em consideração dados experimentais.

Por sua vez, os métodos experimentais desenvolvidos para detecção de *off-targets* parecem ter desempenho semelhante, ainda que cada um tenha seus pontos fracos e fortes. Um dos primeiros trabalhos feitos para descobrir DBS gerados *in vitro*, Digenome-seq (Kim et al, 2015), analisou padrões de fragmentação de bibliotecas geradas a partir da purificação de DNA genômico. Mais recentemente, duas técnicas foram desenvolvidas baseadas na ligação de adaptadores nas regiões de DBS resultantes da clivagem por Cas9 para sua detecção: SITE-Seq (Cameron et al., 2017) e CIRCLE-Seq (Tsai et al., 2017).

GUIDE-seq (Tsai et al., 2015) é a técnica envolvendo detecção de *off-targets* que melhor aumenta a sensibilidade do sistema, inundando-o de oligodeoxinucleotídeos de fita dupla. Assim, quando a clivagem do DNA ocorre, esses elementos podem ser inseridos, facilitando

sua detecção. Esta, é uma técnica da preferência de muitos grupos de pesquisa para identificar sítios fora do alvo experimentalmente.

Um fator que dificulta o entendimento dos reais efeitos dos *off-targets* no genoma humano está relacionado ao fato de que os métodos publicados para detecção de *off-targets* não utilizam um conjunto único de sgRNAs ou células em suas análises. Análises mais homogêneas teriam a possibilidade de uma compreensão mais ampla e sistemática da comparabilidade entre esses métodos. Os dados experimentais disponíveis e os métodos para detecção de *off-targets* ainda são poucos e, segundo Bao e colaboradores (2020), dispersos. Isso dificulta a comparação dos métodos relacionados à verificação da ocorrência de *off-targets*.

Idealmente, todas as técnicas experimentais desenvolvidas, deveriam ser testadas usando um mesmo conjunto de sgRNAs e tipos celulares diferenciados, para que tantos seus próprios mecanismos fossem testados, como preditores existentes pudessem ser analisados com base em um conjunto de dados homogêneos. Isso permitiria que os preditores fossem classificados quanto ao seu desempenho e direcionados para tipos de guias, células e contextos específicos.

Além disso, uma preocupação extra surge quando pensamos em edição ocorrendo fora do sítio alvo: quanto os estudos que visam a compreensão de mecanismos celulares são afetados pela ocorrência de modificações não relatadas nas células estudadas? Não há na literatura relatos de preocupação nesse sentido.

Fica evidente, no entanto, que a produção de mais dados melhorariam não apenas os métodos experimentais, mas todos aqueles que se baseiam nesses dados para a predição de *off-targets* e, até o momento, para que a predição de *off-targets* seja realmente eficiente, parece ser preciso que ela seja, antes de sua aplicação propriamente dita, sempre validada por algum método experimental.

Referências adicionais

- Actor, J. K. (2012). Basic Virology. Elsevier's Integrated Review Immunology and Microbiology, 121–128. doi:10.1016/b978-0-323-07447-6.00013-2
- Arbuckle, J. H., & Medveczky, P. G. (2011). The molecular biology of human herpesvirus-6 latency and telomere integration. *Microbes and infection*, 13(8-9), 731–741. <https://doi.org/10.1016/j.micinf.2011.03.006>
- Asokan, A., Schaffer, D. V., & Samulski, R. J. (2012). The AAV vector toolkit: poised at the clinical crossroads. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(4), 699–708. <https://doi.org/10.1038/mt.2011.287>
- Atchison R. W., Casto B. C. and Hammon W. M. (1965) Adenovirus-associated defective virus particles. *Science* 149:754–755.
- Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics (Oxford, England)*, 30(10), 1473–1475. <https://doi.org/10.1093/bioinformatics/btu048>
- Bao XR, Pan Y, Lee CM, Davis TH, Bao G. Tools for experimental and computational analyses of off-target editing by programmable nucleases. *Nat Protoc.* 2021;16(1):10-26. doi:10.1038/s41596-020-00431-y
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, 315(5819), 1709–1712. doi:10.1126/science.1138140
- Becker, S., & Boch, J. (2021). TALE and TALEN genome editing technologies. *Gene and Genome Editing*, 2, 100007. doi:10.1016/j.ggedit.2021.100007
- Berns, K. I., & Muzyczka, N. (2017). AAV: An Overview of Unanswered Questions. *Human Gene Therapy*, 28(4), 308–313. doi:10.1089/hum.2017.048
- Blaese, R. M., Culver, K. W., Miller, A. D., Carter, C. S., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P., Greenblatt, J. J., Rosenberg, S. A., Klein, H., Berger, M., Mullen, C. A., Ramsey, W. J., Muul, L., Morgan, R. A., & Anderson, W. F. (1995). T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science (New York, N.Y.)*, 270(5235), 475–480. <https://doi.org/10.1126/science.270.5235.475>
- Blanchet, C., Pasi, M., Zakrzewska, K., & Lavery, R. (2011). CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic acids research*, 39(Web Server issue), W68–W73. <https://doi.org/10.1093/nar/gkr316>
- Blicharz, L., Rudnicka, L., & Samochocki, Z. (2019). Staphylococcus aureus: an underestimated factor in the pathogenesis of atopic dermatitis?. *Postepy dermatologii i alergologii*, 36(1), 11–17. <https://doi.org/10.5114/ada.2019.82821>

- Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, 151(Pt 8), 2551–2561. <https://doi.org/10.1099/mic.0.28048-0>
- Boylston, A. (2012). The origins of inoculation. *Journal of the Royal Society of Medicine*, 105(7), 309–313. doi:10.1258/jrsm.2012.12k044
- Brandt, D. Y., Aguiar, V. R., Bitarello, B. D., Nunes, K., Goudet, J., & Meyer, D. (2015). Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. *G3 (Bethesda, Md.)*, 5(5), 931–941. <https://doi.org/10.1534/g3.114.015784>
- Bravo-Egana, V., Sanders, H., & Chitnis, N. (2021). New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing. *Human immunology*, 82(7), 478–487. <https://doi.org/10.1016/j.humimm.2021.01.010>
- Bravo, J., Liu, M. S., Hibshman, G. N., Dangerfield, T. L., Jung, K., McCool, R. S., Johnson, K. A., & Taylor, D. W. (2022). Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature*, 603(7900), 343–347. <https://doi.org/10.1038/s41586-022-04470-1>
- Bravo, J., Liu, M. S., Hibshman, G. N., Dangerfield, T. L., Jung, K., McCool, R. S., Johnson, K. A., & Taylor, D. W. (2022). Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature*, 603(7900), 343–347. <https://doi.org/10.1038/s41586-022-04470-1>
- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., ... van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, 321(5891), 960–964. doi:10.1126/science.1159689
- Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L., & Gao, G. (2021). Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy*, 6(1). doi:10.1038/s41392-021-00487-6
- Cameron, P., Fuller, C. K., Donohoue, P. D., Jones, B. N., Thompson, M. S., Carter, M. M., Gradia, S., Vidal, B., Garner, E., Slorach, E. M., Lau, E., Banh, L. M., Lied, A. M., Edwards, L. S., Settle, A. H., Capurso, D., Llaca, V., Deschamps, S., Cigan, M., Young, J. K., ... May, A. P. (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nature methods*, 14(6), 600–606. <https://doi.org/10.1038/nmeth.4284>
- Casanovas MorenoTorres, I., Jiménez Guerra, G., Foronda GarcíaHidalgo, C., & Serrano García, M. L. (2020). Detection of Streptococcus pyogenes in throat swab samples using antigen detection technique [Detección de Streptococcus pyogenes en muestras faringoamigdalares mediante técnica de detección de antígeno]. *Revista española de quimioterapia : publicacion oficial de la Sociedad Espanola de Quimioterapia*, 33(2), 137–138. <https://doi.org/10.37201/req/078.2019>
- Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., Inga, A., & Cereseto, A. (2018). A highly specific

- SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology*, 36(3), 265–271. <https://doi.org/10.1038/nbt.4066>
- Chen, H. H., Mack, L. M., Kelly, R., Ontell, M., Kochanek, S., & Clemens, P. R. (1997). Persistence in muscle of an adenoviral vector that lacks all viral genes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5), 1645–1650. <https://doi.org/10.1073/pnas.94.5.1645>
- Chirmule, N., Propert, K., Magosin, S., Qian, Y., Qian, R., & Wilson, J. (1999). Immune responses to adenovirus and adeno-associated virus in humans. *Gene therapy*, 6(9), 1574–1583. <https://doi.org/10.1038/sj.gt.3300994>
- Cho, S. W., Kim, S., Kim, J. M., & Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31(3), 230–232. <https://doi.org/10.1038/nbt.2507>
- Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), 132–141. <https://doi.org/10.1101/gr.162339.113>
- Ciuffi A. (2016). The benefits of integration. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 22(4), 324–332. <https://doi.org/10.1016/j.cmi.2016.02.013>
- Clute, S. C., Watkin, L. B., Cornberg, M., Naumov, Y. N., Sullivan, J. L., Luzuriaga, K., Welsh, R. M., & Selin, L. K. (2005). Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. *The Journal of clinical investigation*, 115(12), 3602–3612. <https://doi.org/10.1172/JCI25078>
- Coffin JM, Hughes SH, Varmus HE. (1997) *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press;
- Collias, D., & Beisel, C. L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications*, 12(1). doi:10.1038/s41467-020-20633-y
- Concordet, J. P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic acids research*, 46(W1), W242–W245. <https://doi.org/10.1093/nar/gky354>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), 819–823. doi:10.1126/science.1231143
- Cruvinel, W., Mesquita, D., Jr, Araújo, J. A., Catelan, T. T., de Souza, A. W., da Silva, N. P., & Andrade, L. E. (2010). Immune system - part I. Fundamentals of innate immunity with emphasis on molecular and cellular mechanisms of inflammatory response. *Revista brasileira de reumatologia*, 50(4), 434–461.
- Crystal R. G. (2014). Adenovirus: the first effective in vivo gene delivery vector. *Human gene therapy*, 25(1), 3–11. <https://doi.org/10.1089/hum.2013.2527>

- Daer, R. M., Cutts, J. P., Brafman, D. A., & Haynes, K. A. (2017). The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS synthetic biology*, 6(3), 428–438. <https://doi.org/10.1021/acssynbio.5b00299>
- DeFilippis, V. R., & Villareal, L. P. (2000). An Introduction to the Evolutionary Ecology of Viruses. *Viral Ecology*, 125–208. doi:10.1016/b978-012362675-2/50005-7
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., ... Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602–607. doi:10.1038/nature09886
- Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology*, 190(4), 1390–1400. <https://doi.org/10.1128/JB.01412-07>
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, 34(2), 184–191. <https://doi.org/10.1038/nbt.3437>
- Eid, A., & Mahfouz, M. M. (2016) Genome editing: the road of CRISPR/Cas9 from bench to clinic. *Experimental & Molecular Medicine*, 48(10), e265–e265. doi:10.1038/emm.2016.111
- Friedland, A. E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., ... Bumcrot, D. (2015). Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biology*, 16(1). doi:10.1186/s13059-015-0817-8
- Friedmann, T., Jonlin, E. C., King, N. M. P., Torbett, B. E., Wivel, N. A., Kaneda, Y., & Sadelain, M. (2015). ASGCT and JSGT Joint Position Statement on Human Genomic Editing. *Molecular Therapy*, 23(8), 1282. doi:10.1038/mt.2015.118
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, 31(9), 822–826. <https://doi.org/10.1038/nbt.2623>
- Gallo, R. C. (2005). History of the discoveries of the first human retroviruses: HTLV-1 and HTLV-2. *Oncogene*, 24(39), 5926–5930. doi:10.1038/sj.onc.1208980
- Gardner, M. R., Mendes, D. E., Muniz, C. P., Martinez-Navio, J. M., Fuchs, S. P., Gao, G., & Desrosiers, R. C. (2022). High concordance of ELISA and neutralization assays allows for the detection of antibodies to individual AAV serotypes. *Molecular therapy. Methods & clinical development*, 24, 199–206. <https://doi.org/10.1016/j.omtm.2022.01.003>
- Greely, H. T. (2019). CRISPR'd babies: human germline genome editing in the “He Jiankui affair”*. *Journal of Law and the Biosciences*. doi:10.1093/jlb/lisz010

Gupta, R. M., Musunuru K. (2014) Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest.* Oct;124(10):4154-61. doi: 10.1172/JCI72992

Hale C. R., Zhao P., Olson S., Duff M. O., Graveley B. R., Wells L., et al. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139 945–956. 10.1016/j.cell.2009.07.040

Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., Cofsky, J. C., Kyrpides, N. C., Banfield, J. F., & Doudna, J. A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science (New York, N.Y.)*, 362(6416), 839–842. <https://doi.org/10.1126/science.aav4294>

Hirano, H., Gootenberg, J. S., Horii, T., Abudayyeh, O. O., Kimura, M., Hsu, P. D., ... Nureki, O. (2016). Structure and Engineering of *Francisella novicida* Cas9. *Cell*, 164(5), 950–961. doi:10.1016/j.cell.2016.01.039

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), 827–832. <https://doi.org/10.1038/nbt.2647>

Hu, Z., Liu, Y., Zhai, X., Dai, J., Jin, G., Wang, L., Zhu, L., Yang, Y., Liu, J., Chu, M., Wen, J., Xie, K., Du, G., Wang, Q., Zhou, Y., Cao, M., Liu, L., He, Y., Wang, Y., Zhou, G., ... Shen, H. (2013). New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nature genetics*, 45(12), 1499–1503. <https://doi.org/10.1038/ng.2809>

Ikeda, A., Fujii, W., Sugiura, K., & Naito, K. (2019). High-fidelity endonuclease variant HypaCas9 facilitates accurate allele-specific gene modification in mouse zygotes. *Communications biology*, 2, 371. <https://doi.org/10.1038/s42003-019-0627-8>

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12), 5429–5433. doi:10.1128/jb.169.12.5429-5433.1987

Jackson, D. A., Symons, R. H. & Berg P. (1972) Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 69(10), 2904–2909. doi:10.1073/pnas.69.10.2904

Jansen, R., Embden, J. D. A. van, Gastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565–1575. doi:10.1046/j.1365-2958.2002.02839.x

Jeziorski, E., Foulongne, V., Ludwig, C., Louhaem, D., Rodiere, M., Sitbon, M., & Courgnaud, V. (2016). Searching for Common Mammalian Retroviruses in Pediatric Idiopathic Diseases. *Viruses*, 8(3), 86. doi:10.3390/v8030086

Jiang, H., & Chess, L. (2009). How the immune system achieves self-nonsel self discrimination during adaptive immunity. *Advances in immunology*, 102, 95–133. [https://doi.org/10.1016/S0065-2776\(09\)01202-4](https://doi.org/10.1016/S0065-2776(09)01202-4)

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816–821. doi:10.1126/science.1225829

Kajon, A., & Lynch, J. (2016). Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Seminars in Respiratory and Critical Care Medicine*, 37(04), 586–602. doi:10.1055/s-0036-1584923

Kaufmann, S. H. E. (2019). Immunology's Coming of Age. *Frontiers in Immunology*, 10. doi:10.3389/fimmu.2019.00684

Kernbach, M., Ramsay, C., Rohr, J. R., & Martin, L. B. (2018). Eco-Immunology: Past, Present, and Future. Reference Module in Earth Systems and Environmental Sciences. doi:10.1016/b978-0-12-409548-9.10890-5

Kim N. K. (2017). Gene-Editing: Interpretation of Current Law and Legal Policy. *Development & reproduction*, 21(3), 343–349. <https://doi.org/10.12717/DR.2017.21.3.343>

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., Hwang, J., Kim, J. I., & Kim, J. S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods*, 12(3), 237–243. <https://doi.org/10.1038/nmeth.3284>

Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H.-Y., ... Kim, J.-S. (2017). In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nature Communications*, 8, 14500. doi:10.1038/ncomms14500

Kim, H. J., Lee, H. J., Kim, H., Cho, S. W., & Kim, J. S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research*, 19(7), 1279–1288. <https://doi.org/10.1101/gr.089417.108>

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), 490–495. <https://doi.org/10.1038/nature16526>

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P., Li, Z., Peterson, R. T., Yeh, J. R., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523(7561), 481–485. <https://doi.org/10.1038/nature14592>

Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*. doi:10.1038/nbt.4192

Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., Lee, K., Jung, I., Kim, D., Kim, S., & Kim, J. S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications*, 9(1), 3048. <https://doi.org/10.1038/s41467-018-05477-x>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li, S., Olson, W. K., & Lu, X. J. (2019). Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic acids research*, 47(W1), W26–W34. <https://doi.org/10.1093/nar/gkz394>

Li, Y., Ge, X., Zhang, H., Zhou, P., Zhu, Y., Zhang, Y., Yuan, J., Wang, L. F., & Shi, Z. (2010). Host range, prevalence, and genetic diversity of adenoviruses in bats. *Journal of virology*, 84(8), 3889–3897. <https://doi.org/10.1128/JVI.02497-09>

Liang P (2015) CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell*. DOI: 10.1007/s13238-015-0153-5

Louten, J. (2016). Poxviruses. *Essential Human Virology*, 273–290. doi:10.1016/b978-0-12-800947-5.00015-6

Lynch, J. P., 3rd, & Kajon, A. E. (2016). Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Seminars in respiratory and critical care medicine*, 37(4), 586–602. <https://doi.org/10.1055/s-0036-1584923>

Ma, J., Köster, J., Qin, Q., Hu, S., Li, W., Chen, C., Cao, Q., Wang, J., Mei, S., Liu, Q., Xu, H., & Liu, X. S. (2016). CRISPR-DO for genome-wide CRISPR design and optimization. *Bioinformatics (Oxford, England)*, 32(21), 3336–3338. <https://doi.org/10.1093/bioinformatics/btw476>

Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *CRISPR*, 47–75. doi:10.1007/978-1-4939-2687-9_4

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J., & Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, 9(6), 467–477. <https://doi.org/10.1038/nrmicro2577>

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121), 823–826. doi:10.1126/science.1232033

Markakiou, S., Gaspar, P., Johansen, E., Zeidan, A. A., & Neves, A. R. (2020). Harnessing the metabolic potential of *Streptococcus thermophilus* for new biotechnological applications. *Current opinion in biotechnology*, 61, 142–152. <https://doi.org/10.1016/j.copbio.2019.12>.

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, 322(5909), 1843–1845. doi:10.1126/science.1165771

Meier, A. F., Fraefel, C., & Seyffert, M. (2020). The Interplay between Adeno-Associated Virus and its Helper Viruses. *Viruses*, 12(6), 662. <https://doi.org/10.3390/v12060662>

Meyer, D., C Aguiar, V. R., Bitarello, B. D., C Brandt, D. Y., & Nunes, K. (2018). A genomic perspective on HLA evolution. *Immunogenetics*, 70(1), 5–27. <https://doi.org/10.1007/s00251-017-1017-3>

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., Almendros C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. 2009 Mar;155(Pt 3):733-740. doi: 10.1099/mic.0.023960-0

Mojica, F. J. M., Diez-Villasenor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, 36(1), 244–246. doi:10.1046/j.1365-2958.2000.01838.x

Naso, M. F., Tomkowicz, B., Perry, W. L., 3rd, & Strohl, W. R. (2017). Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy*, 31(4), 317–334. <https://doi.org/10.1007/s40259-017-0234-5>

Nguyen, A. T., Szeto, C., & Gras, S. (2021). The pockets guide to HLA class I molecules. *Biochemical Society transactions*, 49(5), 2319–2331. <https://doi.org/10.1042/BST20210410>

Núñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature Structural & Molecular Biology*, 21(6), 528–534. doi:10.1038/nsmb.2820

Ormond, K. E., Mortlock, D. P., Scholes, D. T., Bombard, Y., Brody, L. C., Faucett, W. A., Garrison, N. A., Hercher, L., Isasi, R., Middleton, A., Musunuru, K., Shriner, D., Virani, A., & Young, C. E. (2017). Human Germline Genome Editing. *American journal of human genetics*, 101(2), 167–176. <https://doi.org/10.1016/j.ajhg.2017.06.012>

Orrù, V., Steri, M., Sole, G., Sidore, C., Viridis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., Floris, M., Mentzen, W. I., Urru, S. A., Olla, S., Marongiu, M., Piras, M. G., Lobina, M., Maschio, A., Pitzalis, M., Urru, M. F., ... Cucca, F. (2013). Genetic variants regulating immune cell levels in health and disease. *Cell*, 155(1), 242–256. <https://doi.org/10.1016/j.cell.2013.08.041>

Pacouret, S., Bouzelha, M., Shelke, R., Andres-Mateos, E., Xiao, R., Maurer, A., Mevel, M., Turunen, H., Barungi, T., Penaud-Budloo, M., Broucque, F., Blouin, V., Moullier, P., Ayuso, E., & Vandenberghe, L. H. (2017). AAV-ID: A Rapid and Robust Assay for Batch-to-Batch Consistency Evaluation of AAV Preparations. *Molecular therapy : the journal of the American Society of Gene Therapy*, 25(6), 1375–1386. <https://doi.org/10.1016/j.ymthe.2017.04.001>

Papapetrou, E. P., Zoumbos, N. C., & Athanassiadou, A. (2005). Genetic modification of hematopoietic stem cells with nonviral systems: past progress and future prospects. *Gene therapy*, 12 Suppl 1, S118–S130. <https://doi.org/10.1038/sj.gt.3302626>

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12.

Poletti, V., & Mavilio, F. (2017). Interactions between Retroviruses and the Host Cell Genome. *Molecular therapy. Methods & clinical development*, 8, 31–41.
<https://doi.org/10.1016/j.omtm.2017.10.001>

Prugnotte, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., & Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Current biology : CB*, 15(11), 1022–1027. <https://doi.org/10.1016/j.cub.2005.04.050>

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), 2281–2308.
<https://doi.org/10.1038/nprot.2013.143>

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D423-31. doi: 10.1093/nar/gku1161

Rosenberg, S. A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R. A., Moen, R., Karson, E. M., Lotze, M. T., Yang, J. C., & Topalian, S. L. (1990). Gene transfer into humans--immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *The New England journal of medicine*, 323(9), 570–578. <https://doi.org/10.1056/NEJM199008303230904>

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for genome editing, regulation and targeting. *Nat. Biotechnol.* 32, 347–355.

Sanders R. (2015) CRISPR inventor calls for pause in editing heritable genes. Acceso: <<https://news.berkeley.edu/2015/12/01/crispr-inventor-calls-for-pause-in-editing-heritable-genes/>>

Schröder A.R., Shinn P., Chen H., Berry C., Ecker J.R., Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* 2002;110:521–529

Schrödinger, L., & DeLano, W. (2020). PyMOL. Retrieved from:
<http://www.pymol.org/pymol>

Sentmanat, M.F., Peters, S.T., Florian, C.P. et al. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep* 8, 888 (2018). <https://doi.org/10.1038/s41598-018-19441-8>

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E. Koonin, E. V. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell*, 60(3), 385–397. doi:10.1016/j.molcel.2015.10.008

Sibbald B. (2001). Death but one unintended consequence of gene-therapy trial. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 164(11), 1612.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., & Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research*, 43(18), e118. <https://doi.org/10.1093/nar/gkv575>

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science (New York, N.Y.)*, 351(6268), 84–88. <https://doi.org/10.1126/science.aad5227>

Song L., Joly Y. (2021) After He Jianku: China's biotechnology regulation reforms. *Medical Law International* 21(2):174-192. doi:10.1177/0968533221993504

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), 62–67. doi:10.1038/nature13011

Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., ... Zhang, F. (2019). Engineering of CRISPR-Cas12b for human genome editing. *Nature Communications*, 10(1). doi:10.1038/s41467-018-08224-4

Summers, W. C. (2009). Virus Infection. *Encyclopedia of Microbiology*, 546–552. doi:10.1016/b978-012373944-5.00323-0

Sun, H., Fu, S., Cui, S., Yin, X., Sun, X., Qi, X., ... Wan, Y. (2020). Development of a CRISPR-SaCas9 system for projection- and function-specific gene editing in the rat brain. *Science Advances*, 6(12), eaay6687. doi:10.1126/sciadv.aay6687

Sung, Y. K., & Kim, S. W. (2019). Recent advances in the development of gene delivery systems. *Biomaterials research*, 23, 8. <https://doi.org/10.1186/s40824-019-0156-z>

The International HIV Controllers Study, 2010 The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. *Science* 330: 1551–1557. doi:10.1126/science.1195271

Tichy, E. D., & Stambrook, P. J. (2008). DNA repair in murine embryonic stem cells and differentiated cells. *Experimental cell research*, 314(9), 1929–1936. <https://doi.org/10.1016/j.yexcr.2008.02.007>

Tran, N. T., Heiner, C., Weber, K., Weiland, M., Wilmot, D., Xie, J., ... Tai, P. W. L. (2020). AAV-Genome Population Sequencing of Vectors Packaging CRISPR Components Reveals Design-Influenced Heterogeneity. *Molecular Therapy - Methods & Clinical Development*, 18, 639–651. doi:10.1016/j.omtm.2020.07.007

Truong, V. A., Hsu, M.-N., Kieu Nguyen, N. T., Lin, M.-W., Shen, C.-C., Lin, C.-Y., & Hu, Y.-C. (2019). CRISPRai for simultaneous gene activation and inhibition to promote stem cell chondrogenesis and calvarial bone regeneration. *Nucleic Acids Research*. doi:10.1093/nar/gkz267

Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., & Joung, J. K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature methods*, 14(6), 607–614. <https://doi.org/10.1038/nmeth.4278>

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., ... Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, 33(2), 187–197. doi:10.1038/nbt.3117

Wagner DL, Amini L, Wendering DJ, et al. High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat Med*. 2019 Feb;25(2):242-248. doi:10.1038/s41591-018-0204-6

Wang, D., Tai, P., & Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. *Nature reviews. Drug discovery*, 18(5), 358–378. <https://doi.org/10.1038/s41573-019-0012-9>

Wang, Q., & Finer, M. H. (1996). Second-generation adenovirus vectors. *Nature Medicine*, 2(6), 714–716. doi:10.1038/nm0696-714

Watson, J. D. & Crick, F. H. C (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356), 737–738. doi:10.1038/171737a0

Weiss, R. A. (1987). Retroviruses and human disease. *Journal of Clinical Pathology*, 40(9), 1064–1069. doi:10.1136/jcp.40.9.1064

WHO. (2019) WHO expert advisory committee on developing global standards for governance and oversight of human genome editing: report of the first meeting. <https://www.who.int/publications/i/item/WHO-SCI-RFH-2019-01>

Wilson, R. F. (2010). The Death of Jesse Gelsinger: New Evidence of the Influence of Money and Prestige in Human Research. *American Journal of Law & Medicine*, 36(2-3), 295–325. doi:10.1177/009885881003600202

Wrona, D., Pastukhov, O., Pritchard, R. S., Raimondi, F., Tchinda, J., Jinek, M., Siler, U., & Reichenbach, J. (2020). CRISPR-Directed Therapeutic Correction at the NCF1 Locus Is Challenged by Frequent Incidence of Chromosomal Deletions. *Molecular therapy. Methods & clinical development*, 17, 936–943. <https://doi.org/10.1016/j.omtm.2020.04.015>

Yan, K., Feng, J., Liu, X., Wang, H., Li, Q., Li, J., Xu, T., Sajid, M., Ullah, H., Zhou, L., Zhou, L., & Chen, Y. (2021). Inhibition of Hepatitis B Virus by AAV8-Derived CRISPR/SaCas9 Expressed From Liver-Specific Promoters. *Frontiers in microbiology*, 12, 665184. <https://doi.org/10.3389/fmicb.2021.665184>

Yan, Y., Tao, H., He, J., & Huang, S. Y. (2020). The HDOCK server for integrated protein-protein docking. *Nature protocols*, 15(5), 1829–1852. <https://doi.org/10.1038/s41596-020-0312-x>

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research*, 40(12), 5569–5576. doi:10.1093/nar/gks216

Yuan, Q., Maas, R., Brouwer, E., Pei, J., Blok, C. S., Popovic, M. A., Paauw, N. J., Bovenschen, N., Hjortnaes, J., Harakalova, M., Doevendans, P. A., Sluijter, J., van der

- Velden, J., & Buikema, J. W. (2022). Sarcomere Disassembly and Transfection Efficiency in Proliferating Human iPSC-Derived Cardiomyocytes. *Journal of cardiovascular development and disease*, 9(2), 43. <https://doi.org/10.3390/jcdd9020043>
- Zabner J et al. A chimeric type 2 adenovirus vector with a type 17 fiber enhances gene transfer to human airway epithelia. *J Virol* 1999; 73: 8689–8695
- Zaboikin, M., Zaboikina, T., Freter, C., & Srinivasakumar, N. (2017). Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents. *PLOS ONE*, 12(1), e0169931. doi:10.1371/journal.pone.0169931
- Zarei, A., Razban, V., Hosseini, S. E., & Tabei, S. M. B. (2019). Creating cell and animal models of human disease by genome editing using CRISPR/Cas9. *The Journal of Gene Medicine*, 21(4), e3082. doi:10.1002/jgm.3082
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163(3), 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>
- Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., ... Zhang, F. (2016). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nature Biotechnology*, 35(1), 31–34. doi:10.1038/nbt.3737
- Zhang, W.-W. (1999). Development and application of adenoviral vectors for gene therapy of cancer. *Cancer Gene Therapy*, 6(2), 113–138. doi:10.1038/sj.cgt.7700024
- Zhang, Y., Heidrich, N., Ampattu, B. J., Gunderson, C. W., Seifert, H. S., Schoen, C., Vogel, J., & Sontheimer, E. J. (2013). Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Molecular cell*, 50(4), 488–503. <https://doi.org/10.1016/j.molcel.2013.05.001>
- Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res*. 2020 Jun;30(6):492-506. doi: 10.1038/s41422-020-0332-7. Epub 2020 May 20. PMID: 32433595; PMCID: PMC7264227.