

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

PAOLA BARCELOS CARNEIRO

**PREDIÇÃO *IN SILICO* DE POTENCIAIS *OFF-TARGETS* PARA VARIANTE
p.Trp402* DO GENE *IDUA* USANDO O SISTEMA DE EDIÇÃO CRISPR/CAS9**

PORTO ALEGRE

2019

PAOLA BARCELOS CARNEIRO

**PREDIÇÃO *IN SILICO* DE POTENCIAIS *OFF-TARGETS* PARA VARIANTE
p.Trp402* DO GENE *IDUA* USANDO O SISTEMA DE EDIÇÃO CRISPR/CAS9**

Trabalho de conclusão de curso apresentado como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas com ênfase em Genética e Biologia Molecular na Universidade Federal do Rio Grande do Sul.

Orientador(a): Profa. Dra. Ursula da Silveira Matte

PORTO ALEGRE

2019

CIP - Catalogação na Publicação

Carneiro, Paola
PREDIÇÃO IN SILICO DE POTENCIAIS OFF-TARGETS PARA
VARIANTE p.Trp402* DO GENE IDUA USANDO O SISTEMA DE
EDIÇÃO CRISPR/CAS9 / Paola Carneiro. -- 2019.
31 f.
Orientadora: Ursula Matte.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Biociências, Bacharelado em Ciências Biológicas,
Porto Alegre, BR-RS, 2019.

1. Mucopolissacaridose tipo I. 2. CRISPR/Cas9. 3.
Off-targets. 4. Bioinformática. I. Matte, Ursula,
orient. II. Título.

PAOLA BARCELOS CARNEIRO

**PREDIÇÃO *IN SILICO* DE POTENCIAIS *OFF-TARGETS* PARA VARIANTE
p.Trp402* DO GENE *IDUA* USANDO O SISTEMA DE EDIÇÃO CRISPR/CAS9**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de
Biociências da Universidade Federal do Rio Grande do Sul como requisito parcial
para a obtenção do título de Bacharel em Ciências Biológicas.

Aprovado em: 10 de Dezembro de 2019.

BANCA EXAMINADORA

Dr.Charley Christian Staats

Nome do professor(a) -UFRGS

Dra.Luciane Maria Pereira Passaglia

Nome do professor(a) - UFRGS

Dra.Ursula da Silveira Matte

Nome do professor(a) – UFRGS (orientadora)

AGRADECIMENTOS

Agradeço ao grupo do Laboratório Células, Tecidos e Genes (CTG) do Hospital de Clínicas de Porto Alegre pela oportunidade de fazer parte desse grupo, pelo apoio e pelos conhecimentos transmitidos.

À professora Dra. Ursula da Silveira Matte pela orientação durante este trabalho, pela oportunidade de trabalhar no grupo de pesquisa, por sempre me incentivar a aprender, por me ensinar a fazer pesquisa científica e me ajudar a construir uma grande parte da bióloga que serei.

À Ms. Martiela de Vaz Freitas pela oportunidade de poder dividir conhecimentos tanto da área computacional como biológica que, com certeza, foram essenciais para formulação deste trabalho como também pela amizade construída a partir desta parceria.

À minha família pelo apoio durante minha trajetória acadêmica, principalmente aos meus pais, Luiz e Rosibel, que me incentivam e apoiam.

Ao meu namorado Thales pelo apoio, carinho e compreensão.

Aos meus amigos Cristal, Dienifer, Hallna, José, Luane, Luisa, Marina, Monique e Pâmella pela amizade, momentos inesquecíveis e conhecimento compartilhado.

Aos meus professores da Universidade Federal do Rio Grande do Sul por terem me transmitindo conhecimentos da área biológica que foram a base para construção dessa análise.

Ao CNPq, Hospital de Clínicas de Porto Alegre e Universidade Federal do Rio Grande do Sul pelo financiamento deste trabalho, permitindo que nós, pesquisadores brasileiros, possamos desenvolver pesquisa para população mundial e fomentando a ciência em nosso país.

“Não existe nada mais fatal para o pensamento que o ensino das respostas certas. Para isso existem as escolas: não para ensinar as respostas, mas para ensinar as perguntas. As respostas nos permitem andar sobre a terra firme. Mas somente as perguntas nos permitem entrar pelo mar desconhecido”

As receitas - Rubem Alves

RESUMO

Mucopolissacaridose tipo I (MPS I) é causada pela deficiência da enzima alfa-L-iduronidase (EC. 3.2.1.76) codificada pelo gene *IDUA*. A deficiência enzimática leva ao acúmulo de heparan e dermatan sulfato dentro de lisossomo desencadeando efeitos multissistêmicos. CRISPR/Cas9 é um mecanismo do sistema imune de procariontos que promove a clivagem de uma região específica a partir de um RNA guia (sgRNA) com 20 nucleotídeos. Terapia gênica com o sistema CRISPR/Cas9 vem sendo desenvolvida para o tratamento de MPS I, contudo a análise dos efeitos *off-targets* deve ser realizada antes da sua aplicação em âmbito clínico. O objetivo desse estudo é avaliar potenciais *off-targets* para a variante mais comum encontrada em pacientes com Mucopolissacaridose I (p.Trp402*). Utilizou-se cinco preditores *in silico*: CHOPCHOP, COSMID, Cas OFFinder, CCTop and CRISPOR. Assim, 109 sequências foram obtidas como potenciais regiões de clivagem no genoma humano, além do sítio-alvo. Ainda, a avaliação da frequência alélica de variantes nessas sequências mostrou 34 sítios polimórficos. Portanto, essas análises permitem uma avaliação antecipada do possível efeito de *off-targets* na funcionalidade celular durante o uso do sistema de edição CRISPR/Cas9.

Palavras chave: MPS I, CRISPR/Cas9, *off-targets*

ABSTRACT

Mucopolysaccharidosis type I (MPS I) is caused by alpha-L-iduronidase deficiency (EC. 3.2.1.76), which is coded by IDUA gene. The enzyme deficiency leads to accumulation of heparan and dermatan sulfate (GAGs) into lysosomes triggering a multisystemic disease. CRISPR/Cas9 is a prokaryote's immune mechanism that promotes the cleavage of specific regions guided by a 20 nucleotide RNA. Therapy with CRISPR/Cas9 has been developed to treat MPS I, however a detailed investigation of off-targets effects must be performed before it reaches clinical application. The aim of this study is to evaluate possible off-target regions for the most predominant variant found in MPS I patients (p.Trp402). In silico predictors CHOPCHOP, COSMID, Cas OFFinder, CCTop and CRISPOR were used. A total of 109 off-targets was obtained for this variant. The evaluation of allele frequency in these sequences showed 34 polymorphic sites. This analysis provides an early screening of the off-target effects in cell function before CRISPR/Cas9 system can be clinically used.*

Key words: MPS I, CRISPR/Cas9, off-targets

SUMÁRIO

1- INTRODUÇÃO.....	10
2- OBJETIVOS.....	14
2.1- Objetivos gerais.....	14
2.2- Objetivos específicos.....	14
3- METODOLOGIA.....	15
3.1- Desenho experimental.....	15
4- RESULTADOS.....	17
5- DISCUSSÃO.....	23
6- CONCLUSÕES E PERSPECTIVAS.....	27
7- REFERÊNCIAS BIBLIOGRÁFICAS.....	29

INTRODUÇÃO

As Mucopolissacaridoses (MPS) estão inseridas no grupo de doenças relacionadas ao depósito lisossomal de glicosaminoglicanos (GAG) causadas pela deficiência de enzimas lisossômicas (Neufeld & Muenzer, 2001). Com exceção da Mucopolissacaridose II, que apresenta padrão de herança ligada ao X, as demais MPS são autossômicas recessivas. Dentre elas, a Mucopolissacaridose tipo I (MPS I) é causada pela deficiência da enzima alfa-L-iduronidase (EC.3.2.1.76) codificada pelo gene *IDUA* (Neufeld & Muenzer, 2001). A deficiência enzimática leva ao acúmulo de glicosaminoglicanos do tipo heparan e dermatan sulfato dentro dos lisossomos, desencadeando problemas nos processos metabólicos nos quais essa organela está envolvida. Com isso há perturbação na estabilidade e integridade celular, que acarretam disfunções multissistêmicas.

Várias mutações patogênicas já foram descritas no gene *IDUA*, e a intensidade das manifestações clínicas está associada à atividade residual da enzima conferida pelas diferentes variantes. Dentre os sintomas apresentados, três síndromes são associadas conforme o espectro clínico: Hurler, Hurler-Scheie, Scheie. Entre essas, Hurler apresenta sintomas mais graves em pacientes com MPS I, e Scheie mais tênues, sendo Hurler-Scheie a forma intermediária entre as duas anteriores (Neufeld & Muenzer, 2001). Entre as variantes identificadas em pacientes com fenótipo mais severo, p.Trp402* é um alelo nulo que apresenta ausência de detecção do produto gênico, alfa-L-iduronidase (Oussoren et al. 2013). Conjuntamente, estudos na população brasileira indicam que essa variante é a mais comumente encontrada entre os pacientes com MPS I (Matte et al., 2000; Matte et al., 2003). Além disso, esta também é a mutação mais comum na maior parte das populações estudadas, superando 50% dos alelos em alguns países (Polleto et al., 2018).

Reposição enzimática e transplante de células hematopoiéticas são terapias aprovadas e aplicadas nos casos de MPS I e visam restituição da enzima deficiente (Hobbs et al., 1981; Giuliani et al., 2010). Contudo, ambos apresentam uma eficiência limitada, especialmente sobre o sistema neurológico (Penati et al., 2017). Atualmente, novos tratamentos estão sendo desenvolvidos a partir de

outras abordagens farmacológicas e de edição gênica para amenizar os sintomas e corrigir o defeito gênico (Gonzalez, 2016; De Carvalho et al., 2017).

Dentre as alternativas terapêuticas de edição gênica, nucleases de dedo de zinco (ZFN), nucleases efetoras do tipo ativador transcricional (TALEN) e repetições palindrômicas curtas agrupadas e regularmente interespaçadas (CRISPR) estão inseridas no grupo de tecnologias de edição gênica que visam à manipulação genômica para diferentes propósitos. O uso de edição gênica utilizando o sistema CRISPR/Cas9 é promissor como alternativa de tratamento, uma vez que permite alta especificidade para correção das diferentes variantes encontradas nos pacientes com doenças genéticas, como a MPS I. Além disso é pouco oneroso e não precisa de construção e validação estrutural de proteína para cada alvo quando comparado às duas outras técnicas anteriores, apenas requerendo uma única endonuclease e pareamento de bases complementares entre RNA e DNA (Doudna & Charpentier, 2014).

CRISPR (do inglês: *Clustered Regularly Interspaced Short Palindromic Repeats*) é um mecanismo do sistema imune de procariontos contra infecções virais, que pode ser aplicado à edição gênica em outros organismos. Seu funcionamento se baseia em uma endonuclease capaz de clivar em uma região específica mediada por uma sequência de RNA (sgRNA) com 20 nucleotídeos complementar ao sítio-alvo adjacente à jusante ou à montante da sequência NGG (PAM) para *Streptococcus pyogenes*. Terapias com o sistema CRISPR/Cas9 estão sendo desenvolvidas para MPS I por intermédio de ensaios *in vitro* e *in vivo* (De Carvalho et al., 2015; De Carvalho et al., 2018; Schuh et al., 2018).

Entretanto, a edição gênica com o uso do sistema CRISPR/Cas9 traz como principal desafio a possibilidade de clivagem fora do sítio alvo (os chamados *off-targets*), associados a efeitos deletérios e prejudiciais na aplicação segura do mecanismo de edição. Logo, há um consenso de que é necessário um melhoramento na especificidade e fidelidade do sistema para sua aplicação no âmbito clínico (Liang et al., 2015; De Carvalho et al., 2015)

Com o propósito de compreender os fatores que levam à clivagem de outros sítios além da região-alvo, há experimentos que analisam alguns fatores baseados na sequência do RNA guia como a existência de uma guanina na posição 20 próxima ao motivo adjacente do protoespaçador (PAM) no sítio-alvo, o conteúdo de GC, número de *mismatches* ou *indels* à montante da sequência PAM

e presença dessa sequência, entre outros (Fu et al., 2013; Scott et al., 2017). Experimentos mostram que o número e posição de *mismatches* são parâmetros importantes na escolha do sgRNA, posto que estão associados à probabilidade de outros sítios serem clivados sendo provavelmente decorrentes da estabilidade da molécula híbrida formada. Em um estudo, a redução do número de *mismatches* pareceu favorecer a clivagem de outros sítios quando comparado com maior número de mudanças na sequência com relação a da região-alvo. No estudo, o número de até 3 *mismatches* favoreceu a clivagem de outro sítio quando comparado com maior número de incompatibilidades da sequência com relação a da região-alvo (Tsai et al., 2015). Contudo, é importante observar que as avaliações de potenciais regiões de clivagem foram aplicadas a partir da inserção de uma sequência em linhagens celulares o que pode estar limitando o número de *off-targets*. Em segundo lugar, a posição de *mismatches* também é um fator crucial, visto que *mismatches* mais distais da PAM são toleráveis e permitem a clivagem de outros sítios. Com o propósito de analisar essas regiões, estudos tentam avaliar a região proximal a PAM (*SEED*), que não pode conter *mismatches*, das sequências dos possíveis sítios *off-targets* similares à sequência do sítio-alvo (Cong et al., 2013; Wu et al., 2014). Não há consenso sobre o tamanho da região proximal a PAM, contudo *mismatches* não localizados nos 5 nucleotídeos próximos à PAM seriam passíveis de clivagem durante o processo de edição para determinado sgRNA em ensaios experimentais (Wu et al., 2014). Ainda, a variação genética entre indivíduos necessita ser considerada na avaliação da *SEED* bem como de outras regiões requeridas pelo sistema de edição. Assim, conjuntamente com a identificação de potenciais *off-targets*, há necessidade de elucidar efeitos da variabilidade genética sob essas sequências para a aplicação do sistema de edição.

As diferenças nas sequências de DNA entre os indivíduos da mesma espécie são o cerne dos estudos sobre variabilidade genética, podendo ser algumas dessas variações do tipo polimorfismos de nucleotídeo único (SNP), inserções e deleções. Estudos demonstram que a variabilidade genética pode ter efeito sobre os resultados do sistema de edição tanto no processo de clivagem do sítio-alvo como de outros sítios similares (Lessard et al., 2017; Scott and Zhang, 2017). Assim, uma investigação acurada da variabilidade genética em sítios *off-targets* se faz necessária. Com isso, bancos de dados populacionais podem ser

úteis no reconhecimento e acesso de variantes dessas sequências bem como das suas frequências alélicas ao longo das diferentes populações após a identificação dos potenciais *off-targets* pelos preditores.

Softwares públicos foram desenvolvidos para uma predição *in silico* dos potenciais sítios *off-targets* para sgRNA de interesse gerados na correção do sítio-alvo pelo sistema CRISPR/Cas9. Esses preditores utilizam o genoma de referência do organismo presente em bancos de dados para detecção prévia das regiões com similaridade que podem ser *off-targets*. Neste trabalho, 5 *softwares* foram utilizados como preditores para avaliar a região onde se localiza a variante p.Trp402* do gene *IDUA* e permitiram identificar regiões *off-targets* que podem estar associados à baixa viabilidade celular após a edição. Resultados do nosso grupo de pesquisa com edição de fibroblastos oriundos de paciente com MPS I mostraram um decréscimo da quantidade de células corrigidas com o sistema CRISPR/Cas9 após 30 dias da edição (Schuh et al., 2017). Ainda, o reconhecimento de variantes para as sequências preditas como potenciais *off-targets* foram avaliadas a partir do banco de dados populacionais 1000 genomas.

Com o intuito de identificar *in silico* potenciais sítios *off-targets* e visando avaliar a fidelidade dos preditores, um painel para Sequenciamento de Nova Geração (NGS) para essas regiões será construído para sequenciamento e permitirá compreender possíveis fatores que interferem na estabilidade celular após a edição. Ainda, nosso delineamento metodológico para potenciais *off-targets* no sistema de edição CRISPR/Cas9 poderá ser utilizado nas práticas clínicas relacionados a novos métodos alternativos de terapia gênica para pacientes com MPS I ou outras doenças genéticas.

OBJETIVOS

Objetivo geral

Analisar efeitos *off-targets* da edição gênica com o sistema CRISPR/Cas9 para a variante p.Trp402*.

Objetivos específicos

Identificar potenciais *off-targets* de sgRNA para a variante p.Trp402* do gene *IDUA* utilizando diferentes preditores *in silico* para o sistema de edição CRISPR/Cas9.

Comparar os diferentes preditores computacionais para análise *in silico* de sítios *off-targets*.

Identificar variantes alélicas das sequências preditas em sítios *off-target* e o seu impacto nos parâmetros fornecidos pelos *softwares*.

METODOLOGIA

O desenvolvimento do projeto foi realizado no Núcleo de Bioinformática do Laboratório de Células, Tecidos e Genes (CTG) do Hospital Clínicas de Porto Alegre sob a orientação da Profa. Dra. Ursula da Silveira Matte e colaboração da Ms. Martiela Vaz de Freitas.

Desenho experimental

O gene *IDUA*, localizado no cromossomo 4p.16.3 na região genômica 980785-998317 conforme a versão do genoma de referência Ghc37/hg19, foi utilizado neste trabalho. O gene contém 14 éxons e codifica duas isoformas a e b sendo a primeira com 653 aminoácidos e a segunda com 523. A variante p.Trp402*, (éxon 9 do gene) presente em pacientes com Mucopolissacaridose tipo I, origina um códon de terminação prematuro que gera uma proteína de tamanho menor com 401 aminoácidos afetando as duas isoformas. Logo, o sgRNA de interesse foi desenhado buscando a melhor sequência que engloba a região genômica onde se encontra a variante p.Trp402* (996530-996552) adjacente a sequência PAM canônica para *Streptococcus pyogenes* (NGG).

A busca da sequência do sítio complementar ao RNA guia (sgRNA) onde se localiza a variante do gene *IDUA* foi realizada no *software* CHOPCHOP (Montague, T. et al, 2014). Nesse processo, a sequência no arquivo FASTA do gene foi fornecida para a busca da região complementar ao RNA guia que contém NGG (sequência canônica PAM) a jusante do sítio-alvo.

A sequência gerada possui 23 nucleotídeos sendo 20 complementar ao sgRNA, onde se encontra a variante p.Trp402* do gene *IDUA*, e 3 referente a PAM (Figura 1). Na escolha da sequência, o conteúdo GC foi estipulado de 65% para o sgRNA. Conjuntamente, também foi avaliada a eficiência de corte, que está intimamente relacionada à presença da sequência PAM adjacente ao sítio-alvo, quantidade de sítios *off-targets* com até 3 *mismatches* e um nucleotídeo de guanina na posição 20 da sequência do sítio-alvo complementar ao sgRNA.

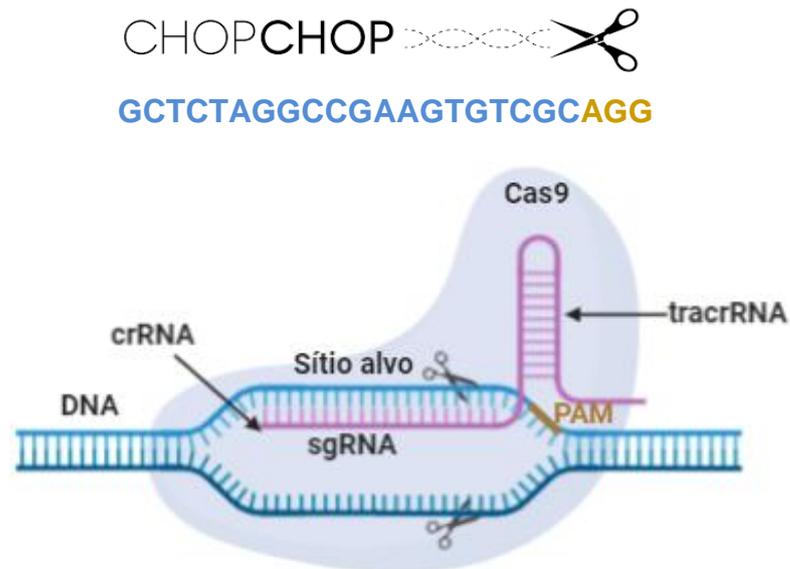


Figura 1: Esquema representativo do sistema CRISPR/Cas9, mostrando a sequência do sítio alvo onde se localiza a variante p.Trp402* (roxo claro) retornada pela busca da ferramenta CHOPCHOP para o RNA guia (sgRNA) adjacente a sequência PAM (amarelo escuro).

A predição de potenciais sítios *off-targets* para sequência *IDUA* foi obtida a partir de cinco preditores: CHOPCHOP, COSMID (Cradick et al., 2014), CRISPOR (Haeussler et al., 2016), Cas OFFinder (Park et al., 2015), CCTop (Stemmer, M. et al 2015). Para tal, utilizou-se a sequência complementar ao sgRNA para busca de sítios similares que possam anelar com o RNA guia durante o processo de edição. A avaliação de potenciais *off-targets* considerou regiões com até 6 pares de bases diferentes (*mismatches*) e sequências com até 2 *indels* similares a do sítio-alvo que os tornem potenciais sítios de clivagem. Por considerar que *mismatches* se referem a possíveis sequências *off-targets* de mesmo tamanho a do sítio-alvo que diferem apenas na composição de bases, as sequências denominadas por *indels* serão referentes a sequências que diferem em tamanho, apresentando mais ou menos pares de bases. Ainda, a escolha dos potenciais sítios *off-targets* priorizou sequências com os 5 nucleotídeos adjacentes à PAM (conhecida por *SEED*) sem a presença de *mismatches*. Posteriormente, as sequências de sítios *off-target* foram submetidas a alinhamento local para confirmar a identidade com a sequência de referência do genoma humano a partir da ferramenta BLASTn, posto que algumas ferramentas apenas fornecem a localização genômica da posição de clivagem.

A avaliação da frequência alélica de variantes nas regiões *off-targets* preditas neste trabalho foi realizada a partir da plataforma 1000 genomas fase 3

versão Gbc37/hg19. No momento da análise, a plataforma continha 2504 indivíduos provenientes de diferentes populações e 5008 alelos avaliados. Contudo, apenas as populações Europeia, América Latina e Africana foram consideradas no reconhecimento dessas variantes, visando uma análise aproximada sobre variabilidade genética brasileira. Como visão global, calculou-se a frequência alélica mundial que considera as demais populações presentes no banco de dados. Alelos com frequência igual ou maior à 1% em pelo menos uma das populações foram considerados polimorfismos nas sequências preditas como potenciais *off-targets*.

RESULTADOS

A avaliação de potenciais *off-targets* a partir dos cinco *softwares* retornou um total de 13.422 sequências sendo algumas presentes em mais de um preditor. Dessas sequências, os preditores que retornaram menor número de sequências foram CHOPCHOP e COSMID com 1 e 15 regiões, respectivamente. CCTOP e CRISPOR foram duas das três ferramentas que retornaram mais regiões, sendo 30 e 24 o número de sequências preditas, respectivamente. Com 13.352, Cas OFFinder retornou o maior número de regiões preditas. A partir do refinamento dos parâmetros na busca descritos acima (limitação do número de *mismatches* e de *indels* e região SEED sem *mismatches*), 109 sequências foram obtidas como potenciais regiões de clivagem no genoma humano além do sítio-alvo. Dessas sequências, 66 regiões *off-targets* contêm apenas *mismatches* (no máximo 6) e as restantes contêm tanto *mismatches* como *indels* (Tabela 1).

Tabela 1: Sequências encontradas, incluindo região-alvo e possíveis *off-targets* mostrando identificação da sequência (ID), sequência, número de *mismatches* (em vermelho) e/ou *indels* (M/I) (inserções em azul e deleções marcadas como travessão), além da localização genômica e região gênica (éxon/ íntron) ou intergênica que as sequências se localizam. A sequência 0 corresponde ao sítio alvo.

ID	Sequências	M/I	Localização Genômica	Região
0	GCTCTAGGCCGAAGTGTGCAGG	0/0	chr4:996530-996552	IDUA : éxon
1	GCTCTGGGCCGAAGTGTGCAGG	1/0	chr4:996530-996552	IDUA : éxon
2	GCTCTGGGCTGGGTGTGCCTGG	4/0	chr10:118919279-118919301	MIR366HG : éxon
3	TCTCTAGGCAAGTGATGCTGG	4/0	chr4:148041025-148041047	Intergênico

4	GCTCTAGGCTGAAGTGCTTCTGG	4/0	chr8:130569489-130569511	<i>CCDC26</i> : íntron
5	CCACTAGGCCAAAGTGTAGCTGG	4/0	chr19:8456438-8456460	<i>RAB11B</i> : íntron
6	GCTCCAGGAGGAAGTGTCACAGG	4/0	chr3:51437682-51437704	<i>DCAF1</i> : íntron
7	GTTCTAGGTGGAAGTGTGCTGG	4/0	chr8:13103641-13103663	<i>DLC1</i> : íntron
8	GCGCTGGCGGGATTGTCGCAGG	5/0	chr7:155601333-155601355	<i>SHH</i> : íntron
9	GATCTGGGGAGAAGGTCGCAGG	5/0	chr19:50427945-50427967	<i>IL4I1</i> : íntron
10	GCTTGGCGGCAGGTCGCAGG	5/0	chr19:2307987-2308009	<i>LINGO3</i>
11	GGTACAGCAGAAGTGTGCTGG	5/0	chr5:146345164-146345186	<i>PPPR2B</i> : íntron
12	AATCCAGTTGAAGTGTGCGCCGG	5/0	chr12:39594468-39594490	Intergênico
13	CCACCAGGCTGCAGTGTGCGCAGG	5/0	chr6:79353909-79353931	Intergênico
14	GGTCTGGGTGTAAGTGTGCGCGG	5/0	chr10:12856342-12856364	<i>CAMK1D</i> : íntron
15	GCGCCCGGCCGAGTGTGCGCGG	5/0	chr7:25991151-25991173	Intergênico
16	GCTCCACACAGTAGTGTGCGCGG	5/0	chr1:1151358-1151380	Intergênico
17	GATGGAGCCGCTGTGTCGCAGG	5/0	chr1:180223446-180223468	<i>LHX4</i> : íntron
18	AATCCAGTTCGAAGGTCGCGG	5/0	chr16:3131030-3131052	Intergênico
19	AATCCAGTTCGAAGGTCGCGG	5/0	chr10:81735284-81735306	Intergênico
20	AATCCAGTTCGAAGGTCGCTGG	5/0	chr11:59691555-59691577	Intergênico
21	AATCCAGTTCGAAGGTCGCGG	5/0	chr6:67041123-67041145	Intergênico
22	AATCCAGGCTGAAGGTCGCTGG	5/0	chrY:21455475-21455497	Intergênico
23	TTTCTAGGCTCACGTGTGCTGG	5/0	chr19:5371949-5371971	Intergênico
24	GGTCAACCCGAAGAGTCGCGG	5/0	chr8:29120598-29120620	Intergênico
25	GCCCTGGTCCGCTGTGTCGCTGG	5/0	chr17:63052331-63052353	<i>GNA13</i> : íntron
26	GCTCGAGATCAAGGTGTCGCAGG	5/0	chr11:70689763-70689785	<i>SHANK2</i> : íntron
27	GATCTGGGGAGAAGGTCGCAGG	5/0	chr19:50427945-50427967	<i>IL4I1</i> : íntron
28	CTTCCAGGCCGAGGAGTCGCGG	5/0	chr9:132640628-132640650	<i>USP20</i> : éxon
29	GGTCTAGGCCAAGCTGTGCTGG	4/0	chr17:31707604-31707626	<i>ASIC2</i> : íntron
30	GCGCTGGCGGGATTGTCGCAGG	5/0	chr7:155601333-155601355	<i>SHH</i> : íntron
31	GCTTGGCGGCAGGTCGCAGG	5/0	chr19:2307987-2308009	<i>LINGO3</i> : íntron
32	GTTCTTGGCGATGAGTCGCGG	5/0	chr6:33281749-33281771	<i>TAPBP</i> : éxon
33	GCTGAAGCCTGAGGTCGCGG	5/0	chr15:83777349-83777371	<i>TM6SF1</i> : íntron
34	GCTCTAATCATAAGCGTCGCAGG	5/0	chr16:47497815-47497837	<i>PHKB</i> : íntron
35	GCACTACACCGAAACGTCGCGG	5/0	chr20:42086382-42086404	Intergênico

36	GCTCGAGCCCAGAGGGTCGCCGG	5/0	chr5:39073917-39073939	<i>RICTOR</i> : íntron
37	GATCTAGGCTAAGGAGTCGCAGG	5/0	chr3:126655025-126655047	<i>CHCHD6</i> : íntron
38	GCTCTAAGCTGGCCTGTCGCTGG	5/0	chr2:220489550-220489572	Intergênico
39	ACTGTAGTCCTACATGTCGCTGG	6/0	chr8:13730106-13730128	Intergênico
40	GCTCCTCCAGGAAGTGTGCAGG	6/0	chr8:23288368-23288390	<i>ENTPD4</i> : éxon
41	ATTTTTGGCAGCAGTGTGCCTGG	6/0	chr12:16574919-16574941	Intergênico
42	TCTCTTTGGCAAGGGTCGCTGG	6/0	chr12:46080385-46080407	Intergênico
43	TGTCCTTGCTGAAGTGTGCAGG	6/0	chr12:119259651-119259673	<i>LINC02439</i> : íntron
44	TTTCTCCGTGCGCAGTGTGCAGG	6/0	chr12:125877331-125877353	<i>TMEM132B</i> : íntron
45	GTGCTGGGCcATTGTGTCGCTGG	6/0	chr3:87409315-87409337	Intergênico
46	GCTCTTCTGCCAGGTGTGCCTGG	6/0	chr3:108156408 -108156430	<i>MYH15</i> : éxon
47	AATCCAGGTTGAAGGGTCGCTGG	6/0	chr3:184823526-184823548	<i>C3orf70</i> : íntron
48	GCCCTAGGCATATCAGTCGCTGG	6/0	chr3:193952534-193952556	<i>LINC02036</i> : íntron
49	GCTCCTCTACcAAGTGTGCAGG	6/0	chr7: 334104-334126	Intergênico
50	GCCCCAAGTAGAAGGGTCGCAGG	6/0	chr7:534830-534852	Intergênico
51	CCTCGTCGCCGCTGTGTCGCTGG	6/0	chr7:1105256-1105278	<i>C7orf50</i> : íntron
52	GTTAAAGAACAAAGTGTGCCTGG	6/0	chr7: 3678772-3678794	<i>SDK1</i> :íntron
53	GCCCTCGTCCGTCTTGTGCAGG	6/0	chr7:27195844-27195866	<i>HOXA7</i> : éxon
54	CCTGTTATCAGAAGTGTGCAGG	6/0	chr7:44106164 -44106186	<i>DBNL</i> : éxon
55	GGACTGTGGGAAGTGTGCAGGG	6/0	chr7:55639474-55639496	<i>VOPP1</i> : íntron
56	CACCCAGGCTGGAGTGTGCAGGG	6/0	chr7: 57129464-57129486	Intergênico
57	ACTCTGGGCCAGGCTGTGCAGG	6/0	chr7:140924451-140924473	<i>TMEM178B</i> : íntron
58	GATCAAGGCTACTGTGTCGCTGG	6/0	chr4:4664672-4664694	Intergênico
59	ACTACCGGCAGGAGTGTGCCCGG	6/0	chr4: 7968537-7968559	<i>ABLIM2</i> : éxon
60	ACTCTGTGCTCATGTGTCGCAGG	6/0	chr4:38271835-38271857	Intergênico
61	ATCCTAGGCCCTGTGTCGCAGG	6/0	chr4:114864472-114864494	<i>ARJS/RP11-26P13.2</i>
62	GCTCTGCCcTAGGGGTCGCAGG	6/0	chr5:11123042-11123064	<i>CTNND2</i> : íntron
63	TGTTcAGGCCTCAGTGTGCCTGG	6/0	chr5:42461909 -42461931	<i>GHR</i> : íntron
64	AATCCAGGTTGAAGAGTCGCTGG	6/0	chr5:44986557-44986579	Intergênico
65	ACTGCAGGTGAAATGTGCAGG	6/0	chr5:79133696 – 79133718	Intergênico
66	GATGCAAGAGGAAGTGTGCCTGG	6/0	chr5:94445982-94446004	<i>MCTP1</i> : íntron

67	CTTCCGGGCCGCAGGGTCGCGGG	6/0	chr5:167956264-167956286	<i>FBLL1</i> : éxon
68	GCTC-AGGCTGAAGGGTCGCAGG	2/1	chr4:89584514-89584535	<i>HERC3</i> : íntron
69	GCTCTGGGAC—GAGTGTGCTGG	2/1	chr19:7294879-7294900	Intergênico
70	GCTCTGGCCGAAGTGA CT CAGG	2/2	chr8:47028210-47028231	Intergênico
71	CCTCTATCCCC—TGTCGCGGG	5/2	chr8:125035770-125035790	<i>FER1L6</i> : éxon
72	GCTCCCGGGCCC—TGTCGCTGG	5/2	chr8:144789064-144789084	<i>CCDC166</i> : éxon
73	ACTCTATGCTGA-GTGTGCTGG	3/1	chr12:111599505 – 111599526	<i>CUX2</i> : íntron
74	CCTCGAGGCcTG—TGTCGCTGG	4/2	chr3:50388167-50388187	<i>NPRL2/CYB561D2</i>
75	GCTGTGAGCTGA—TGTCGCAGG	4/2	chr3:64388910-64388930	Intergênico
76	CGGCTAGGCCc—GTGTGCGGG	4/2	chr3:169899326 -169899346	<i>PHC3</i> : íntron
77	GCTCTGGGTTTA—TGTCGCTGG	4/2	chr7: 75610691-75610711	<i>POR</i> : íntron
78	AAGCCAGGCCGA—TGTCGCAGG	4/2	chr7:101566496-101566516	<i>CUX1</i> : íntron
79	cCTCTGTGT CGA—TGTCGCTGG	4/2	chr7: 130487679-130487699	Intergênico
80	GCTCTAGTACAG-GTGTGCTGG	4/1	chr7:33754054-33754075	Intergênico
81	GATCTATACCCA-GTGTGCTGG	4/1	chr7:75049729-75049750	<i>POM121C</i> : íntron
82	TCTCTTGGGCcA-GTGTGCGAGG	4/1	chr4: 182426111-182426132	Intergênico
83	TCTCTTGGCCC—GTGTGCGCCGG	3/2	chr19:37763002-37763022	Intergênico
84	TCTCTTGGCCC—GTGTGCGCCGG	3/2	chr19:37770592-37770612	Intergênico
85	GCACTTGCTGA—TGTCGCAGG	3/2	chr10:71497328 -71497348	Intergênico
86	GATCTGGGCAG—GTGTGCTGG	3/2	chr14:100913107-100913127	<i>WDR25</i> : íntron
87	GAACTAGGCCGTA-TGTCGCTGG	3/1	chr9:90693036 – 90693057	Intergênico
88	GCTCAAGGCCCC—TGTCGCTGG	3/2	chrX:70369551-70369571	<i>NLGN3</i> : íntron
89	GTTTTGTTT—AGTGTGCGGG	5/2	chr8:2915404-2915424	<i>CSMD1</i> : íntron
90	GATGACAGCA—AGTGTGCGAGG	6/2	chr8:6931491- 6931511	Intergênico
91	GAAACAGCCA—AGTGTGCGGG	6/2	chr8: 9832229-9832249	Intergênico
92	ATTTTGGGCC—AGTGTGCTGG	4/2	chr7:2678363 -2678383	<i>TTYH3</i> : íntron
93	GCTGCAGGCG—AGTGTGCTGG	3/2	chr7:102180465-102180485	<i>POLR2J3</i> : éxon
94	GCTGCAGGCG—AGTGTGCTGG	3/2	chr7:102279573-102279593	Intergênico
95	ACTAAAGGCC—AGTGTGCTGG	3/2	chr21:44988824-44988844	<i>HSF2BP</i>
96	GCTGTTGCC—AGTGTGCTGG	3/2	chr10:125129485-125129505	Intergênico
97	GAGGAAGG—GAAGTGTGCTGG	4/2	chr3:39247858 -39247878	Intergênico
98	GCAATAGAG-GAAGTGTGCTGG	4/1	chr1: 229832258-229832279	Intergênico

99	GCGCTG-GCCGCAGAGTCGCCGG	3/1	chrX:153618565-153618586	Intergênico
100	GCCCT—GCCGAAGGGTCGCAGG	3/2	chr2:197183519 -197183539	<i>HECW2</i> : éxon
101	GCTCTTCGGCCCAACGTGTCGCTGG	4/2	chr4:156298074-156298098	<i>MAP9</i> : éxon
102	GCTGAGGGCCGTTGAGTGTGCAGG	4/2	chr19:38755268 -38755292	<i>SPINT2</i> : éxon
103	GGCCTGGGCCGCAAACTGTCGCCGG	4/2	chr11:66328772-66328796	<i>ACTN3</i> : éxon
104	GAAGTGGGCAACGAGGTGTCGCTGG	4/2	chr11:120708680-120708704	<i>GRIK4</i> : íntron
105	ACTCTGAGGCCAAGGTGTCGCAGG	4/1	chr8:105318003-105318026	Intergênico
106	GCACTGGGCCGAGGGTCGCGGG	4/1	chr8:48920856-48920879	Intergênico
107	GCTTCGGGCCGAGGGTCGCGGG	4/1	chr5:167956263-167956286	<i>FBL1</i> : éxon
108	GCGCGGGGCCAGAAGGTGTCGCGGG	4/1	chr1: 19283072-19283095	<i>IFFO2</i> : éxon
109	GCTATGGGCAGAGGCTGTCGCTGG	4/1	chr15:62119936 -62119959	Intergênico

FONTE: Dados de análise

Estes potenciais *off-targets* encontram-se tanto em regiões gênicas quanto intergênicas do genoma de referência (Tabela 1). A identificação das regiões, onde as sequências se localizam, demonstrou que 64 *off-targets* estão em regiões gênicas e 45 em regiões intergênicas. Dessa forma, podemos inferir que não apenas outros genes podem estar sujeitos à edição durante o processo, mas também regiões regulatórias proximais ou distais de genes.

Em seguida, a análise conduzida para estimar o impacto da variabilidade genética nas sequências preditas como potenciais *off-targets*, demonstrou a presença de 34 sítios polimórficos. Algumas dessas variantes foram encontradas na fita complementar das sequências preditas *off-targets*. Esses polimorfismos foram identificados em 29 das 109 sequências em pelos menos umas das três populações consideradas neste trabalho (Europa, América Latina e África) a partir do banco de dados 1000 Genomas. Resultados obtidos da frequência alélica em porcentagem para cada polimorfismo na Europa (F.E), América Latina (F.AL), África (F.A) bem como na população mundial (F.M) estão mostradas abaixo (Tabela 2).

Tabela 2: Frequência alélica de sequências preditas *off-targets* com polimorfismos em pelos menos umas das três populações. A tabela mostra a identificação da sequência (ID) conforme tabela 1, sequência com sítio polimórfico marcado em negrito, alelo de referência (REF), alelo alternativo (ALT) e frequências do alelo alternativo em porcentagem (%) nas três populações (F.E, F.AL, F.A) e mundial (F.M).

ID	Sequências	REF	ALT	F.E	F.AL	F.A	F.M
71	CCTCTATTCCCC—TGTCGC G GG	G	C	59,2	39,5	68	54,3
104	GAAGTGGGCA A CGAGGTGTCGCTGG	A	G	38,4	28,1	15,1	31,8

93	GCTGCAGG CG —AGTGTCTGCTGG	C	T	21	28,9	9,5	31,5
47	AATCCAGG TT GAAGGGTCTGCTGG	T	C	20,2	28,2	38,1	30,9
61	ATCCTAGGCCCTGTGT CG CAGG	C	T	31	29	17,1	25,4
107	G CTTCCGGGCCCGCAGGGTCTGCGGG	G	A	14,1	15,4	6,8	21,5
13	CCACCAGGCTGCAGTGT CG CAGG	G	A	25,1	24,8	12,2	18,3
12	AATCCAGGTTGAAGTGT CGCC GG	C	T	22,3	17,7	2,7	11,6
19	AATCCAGGTCGAAGGGT CG CCGG	C	T	7,1	15,1	1,3	8,4
5	CCACTAGGCCAAAGT GTAG CTGG	TAC	T	4,3	3,7	11,3	7
60	ACTCTGTGCTCATGTGT CG CAGG	C	T	0,3	3	12,8	5,2
48	GCCCTA GGCA TATCAGTCGCTGG	GGCA	G	0,8	1	9,1	4,1
62	GCTCTGCCCTGAGGGGT CG CAGG	C	A	0	1	9,2	2,6
51	CCTCGTCGCCGCTGTGT CG CTGG	G	C	0,1	0,6	7,7	2,1
6	AATCCAGGTCGAAGGGT CGCC GG	C	T	0,1	0,4	6,1	1,7
72	GCTC CC GGGCC—TGTCGCTGG	G	A	0	0,58	5,9	1,6
53	GCCCTCGTCCGT CT TGTCTGCAGG	T	C	0,5	0,9	3,1	1,4
37	GATCTAGGCTAAGGAGT CG CAGG	C	T	0	0,58	4,6	1,3
14	GGTCTGGGTGTAAGTGT CG CGGG	C	T	0,1	0,4	4,3	1,2
96	GCTGTTTGCC—AGTGTCTG CT GG	G	A	0	0,6	3,9	1,1
68	GCTC-AGGCTGAAGGGT CG CAGG	C	T	0	0,6	3,1	0,9
80	GCTCTAGTACAG-GTGT CG CTGG	C	T	3,3	1,2	0,1	0,9
96	GCTGTTTGCC--AGTGT CG CTGG	G	A	1,8	0,14	3,1	0,8
51	CCTC G TCGCCGCTGTGTCTGCTGG	C	G	0	0,1	2,5	0,7
14	GGTCTGGGTGTAAGTGT CG CGGG	G	A	0	0	2,8	0,7
86	GATCTGGGCAG—GTGT CG CTGG	G	A	0	0	2,6	0,7
33	GCTGAAGGCCTGAGGGT CGCC GG	C	T	0,1	0	2,2	0,6
68	GCTC-AGGCTGAAG GG TCTGCAGG	G	C	0	0	1,4	0,4
64	AATCCAGGTTGAAGAGT CG CTGG	G	A	0	0	1,4	0,4
15	GCGCCCG G CCCGAGTGTCTGCGGG	C	G	0	0	1,4	0,4
40	GCTCCT CCAGGAAGTGTCTGCAGG	GCTCCT	G	0	0,1	1,2	0,4

40	GCTCCTC CAGGA AAGTGTCGCAGG	CAGGA	C	0	0,1	1,2	0,4
89	GTTTTTGTTT--AGT G TTCGCGGG	C	T	0	0	1,6	0,4
43	TGTCCTTGCTGAAGTGTC G CAGG	G	C	0	0,1	1,4	0,4

FONTE: Dados de análise

DISCUSSÃO

Vários *softwares* foram desenvolvidos para o sistema de edição CRISPR/Cas9 e estão disponíveis *online*. Alguns desses preditores fornecem as sequências de *off-targets* classificadas a partir de alguns parâmetros já avaliados em ensaios experimentais. Alguns desses são consensos como a presença da sequência PAM canônica (NGG) para *Streptococcus pyogenes*, número e posição de *mismatches*. Contudo, outros parâmetros são diferentes e, conseqüentemente, fornecem resultados divergentes conforme o algoritmo construído.

Neste trabalho, escolhemos os cinco *softwares* mais utilizados na literatura. Alguns permitiram localizar mais regiões *off-targets* do que outros a partir do número de *mismatches* e possibilidade de reconhecer regiões que contêm *indels* como também pela falta de classificação por escore das sequências. Dessas ferramentas, Cas OFFinder retornou maior número de regiões *off-targets*, contudo apenas 94 foram selecionadas quando considerada a preservação dos cinco nucleotídeos adjacentes à PAM.

Diferente de Cas OFFinder, CCTop classifica internamente as sequências com escore de 1000 para a primeira sequência, seguido dos possíveis *off-targets* menos prováveis que são computados com escore menor até zero. A classificação leva em consideração o número e distribuição de *mismatches* no sgRNA, que diminui o escore conforme sua proximidade à PAM. Ainda, por considerar os efeitos deletérios que *off-targets* podem gerar, a distância dessas sequências de regiões de éxons anotados é considerada para o cálculo do escore (Stemmer et al., 2015). Todas as sequências preditas por CCTop foram selecionadas, uma vez que é possível definir o tamanho da *SEED* durante a submissão do sgRNA e todas as sequências deste preditor já tinham sido selecionadas previamente dos

resultados de Cas OFFinder. Para os três preditores restantes, outros parâmetros foram analisados.

Para CRISPOR, as sequências foram escolhidas a partir de dois parâmetros, sendo o valor CFD (*Cutting Frequency Determination*) que computa potenciais *off-targets* para determinado sgRNA ou conservação da *SEED*. O parâmetro CFD é proposto pelo preditor por permitir distinguir entre *off-targets* e falsos-positivos (Haeussler et al., 2016). Logo, sequências com $CFD > 0,2$ são consideradas possíveis alvo de clivagem com base no algoritmo utilizado a partir de estudos de interação entre a molécula híbrida sgRNA e DNA (Doench et al., 2016). Ainda, foi priorizada a conservação da *SEED* de 5 nucleotídeos adjacentes a PAM e, conseqüentemente, algumas sequências que não tiveram um CFD estipulado foram selecionadas. Dessa forma, essa decisão de escolha permitirá averiguar a confiabilidade dos resultados dessa ferramenta a partir da validação que será realizada pelo sequenciamento de nova geração após edição de fibroblastos oriundo de paciente com MPS I utilizando o sistema CRISPR/Cas9.

Para COSMID, a escolha foi realizada a partir da classificação das sequências que estipula penalidades conforme o número e posição de *mismatches*, procurando regiões com base em homologia de sequência. Ainda, a ferramenta também procura por regiões *off-targets* com a presença de *indels*, contudo não há computação para esse parâmetro no escore das sequências. Assim, as sequências que apresentaram escore próximo à zero ou priorização da *SEED* foram selecionadas.

Por último, CHOPCHOP apenas retornou 1 sequência, posto que apenas considera *off-targets* com até 3 *mismatches* e promove o retorno de possíveis regiões fora do alvo a partir do programa de alinhamento Bowtie (Langmead et al., 2009) que permite o rápido alinhamento de sequências curtas com genomas relativamente grandes. Dessa sequência, apenas 1 *mismatch* foi predito sendo localizada na região alvo, uma vez que o alinhamento considerou a variante não patogênica presente no genoma de referência um possível *off-targets* para a variante p.Trp402* encontrada em pacientes com MPS I. Assim, por considerar sequências com até 3 *mismatches*, sem *indels* e o tipo de alinhamento utilizado, o número de regiões preditas foi menor quando comparado com as demais ferramentas.

Dessa forma, compreende-se o porquê de diferentes preditores retornarem quantidades diferentes de potenciais regiões *off-targets* para um mesmo sgRNA de interesse como também a identificação de prováveis regiões de clivagem em detrimento de outras a partir da classificação por escore das sequências. No entanto, chama a atenção que um dos preditores mais utilizados na literatura (CHOPCHOP) retorne apenas um resultado. Esta observação reforça a necessidade de uso de um maior número de preditores.

Outro fator relevante para a escolha de possíveis candidatos *off-targets* é a proximidade de *mismatches*. Estudo demonstrou que a proximidade pode ter influência no processo de clivagem, sendo resultado da interferência na cinética e a afinidade de ligação entre a molécula híbrida quando comparado com *mismatches* dispersos na sequência do sgRNA o que é consequência da menor instabilidade do pareamento necessário (Zhang et al., 2017). Contudo, é necessário também avaliar a região de *mismatches* adjacentes para ter uma maior compreensão dos efeitos de clivagens de outros sítios sobre essa situação. Com isso, infere-se que potenciais *off-targets* com *mismatches* mais dispersos na sequência podem ter maior susceptibilidade de clivagem quando comparado com outras dependendo da posição que se localizam.

Neste trabalho, realizamos ainda uma análise para verificar se os resultados de predição de *off-targets* podem ser alterados devido à variabilidade genética existente entre os indivíduos. Essa análise demonstrou que aproximadamente 27% (n = 29) do total de sequências *off-targets* apresentam sítios polimórficos em pelo menos umas das três populações consideradas (Europa, América Latina e África). Localizou-se polimorfismos que interferem no número de *mismatches*, na *SEED* e na sequência PAM que são parâmetros importantes na clivagem dos sítios e considerados na busca de potenciais *off-targets* pelos preditores para o sistema de edição. Dessas variantes, 15 e 20 foram reconhecidas como polimorfismos em pelos menos duas das três populações e na mundial, respectivamente. Percebe-se a partir dessa análise que aproximadamente 17% (n = 18) do total potenciais *off-targets* apresentam a possibilidade de interrupção da PAM e/ou *SEED* por variantes genéticas. Com isso, observa-se que os alinhamentos realizados pelos softwares com base no genoma de referência não levam em consideração a variabilidade genética entre os indivíduos nas predições de potenciais *off-targets* para o sgRNA de interesse.

Assim, infere-se que alguns sítios preditos poderiam apresentar menor probabilidade de ser clivados em alguns indivíduos ou em determinada população onde o alelo alternativo tem uma alta frequência, uma vez que a variante gera uma menor estabilidade de anelamento do sgRNA ao sítio devido ao aumento de incompatibilidades na *SEED* ou ao longo da sequência e abolição do reconhecimento pela endonuclease quando a sequência PAM é alterada.

Por outro lado, não sabemos se existem variantes polimórficas que criem uma sequência PAM ou SEED, tornando-se um potencial *off-target*. Assim, pode-se inferir a partir do parâmetro de variabilidade genética que a estipulação de um limite no número de *mismatches* (máximo 6) e *indels* (máximo 2) pode estar subestimando a quantidade de potenciais *off-targets*. Conjuntamente com pressuposto anterior, hipotetiza-se que a procura de potenciais *off-targets* pelos *softwares* mediante a existência da sequência PAM no genoma de referência está desconsiderando regiões fora do alvo que possam apresentar sequência PAM ausente no genoma de referência, mas criada por variantes com frequências distintas entre as populações.

Todavia, é necessário também avaliar não apenas as diferentes propostas de alinhamentos e parâmetros em nível de homologia de sequências com DNA de referência, mas também o contexto biológico dessas regiões. Apesar do manual de alguns dos preditores utilizados mostrarem a importância do nível de cromatina na acessibilidade dos fatores do sistema edição no processo de clivagem dos sítios, nenhum permite avaliar os resultados com base nesse parâmetro. Avaliações experimentais detectaram clivagens de sítios pré-determinados por serem constitutivamente desprovidos de nucleossomos quando comparados com regiões mais empacotadas próximas à região promotora do gene *ho* de cepas fúngicas utilizando o sistema de edição CRISPR/Cas9 (Yarrington et al., 2018). Ainda, é interessante considerar a avaliação em nível de cromatina com o tipo celular da aplicação do sistema de edição, posto que a estrutura da cromatina, que desempenha papel maior ou menor grau de acessibilidade para clivagem das regiões, está intimamente relacionada no estabelecimento de um específico tipo celular. Dessa forma, essa informação pode ser útil para o desenvolvimento de ferramentas do sistema de edição que reforçará os resultados *in silico* em busca de potenciais *off-targets* para determinado sgRNA de interesse, permitindo ter

uma compreensão antecipada das consequências de *off-targets* na funcionalidade dos diferentes tipos celulares.

CONCLUSÕES E PERSPECTIVAS

A predição e avaliação de possíveis *off-targets* está relacionada com a aplicação segura do mecanismo de edição CRISPR/Cas9 na terapia gênica. O uso de preditores computacionais para *off-targets* é um processo que permite uma pré-avaliação dos efeitos da edição gênica pelo pesquisador. A análise de outros potenciais sítios de clivagem durante a correção gênica da variante p.Trp402* permitiu averiguar fatores que podem influenciar a baixa eficiência de edição ou efeitos deletérios para estabilidade e integridade celular *in vitro* ou *in vivo*. Dessa forma, a predição *in silico* de sítios *off-targets* desempenha um papel importante no uso do sistema CRISPR/Cas9 como método de terapia gênica em pacientes com MPS I.

Ainda, o uso de mais de uma ferramenta é necessário, visto que permitiram avaliar a fidelidade de diferentes preditores públicos disponíveis para o desenho do sgRNA e outros aspectos que devem ser considerados durante a predição *in silico*. Possíveis *indels* podem gerar potenciais sítios de clivagem durante a edição gênica. A partir disso, COSMID e Cas OFFinder permitiriam reconhecer essas regiões que não seriam visualizadas em apenas um alinhamento local da sequência do sgRNA deste estudo ou usando apenas as outras três ferramentas computacionais utilizadas.

O reconhecimento de polimorfismos nas sequências *off-targets* previstas para a variante p.Trp402* mostrou a importância da variabilidade genética na busca por esses sítios. Considera-se aqui importante a construção de ferramentas de edição que levem em consideração a frequência de variantes de nucleotídeos únicos, inserções, deleções como meio de tentar estimar potenciais sítios *off-targets* na avaliação *in silico*. A validação de algumas sequências obtidas como potências *off-targets* e do delineamento metodológico deste trabalho na busca dessas regiões será realizada a partir da montagem de um painel para sequenciamento de nova geração após cultivo celular *in vitro* utilizando fibroblastos humanos de paciente com MPS I. Dessa forma, permitirá avaliar

efeitos *off-targets* utilizando o sistema CRISPR/Cas9 bem como comparar os resultados obtidos a partir de diferentes *softwares* desenvolvidos para esse sistema de edição *in silico*. Dessa forma espera-se contribuir para o uso do mecanismo CRISPR/Cas9 em tratamento com fins terapêuticos para pacientes com MPS I.

REFERÊNCIAS BIBLIOGRÁFICAS

CONG, Le; RAN, F. Ann; COX, David et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, v.339, p.819-823, 2013, fev. 2015. doi: 10.1126/science.1231143

CRADICK, Thomas J.; QIU, Peng; LEE, Ciaran M. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Molecular Therapy. Nucleic Acids*, v.3, pp.1-26, dez. 2014. doi:10.1038/mtna.2014.64

DE CARVALHO, Talita G. et al. Genome Editing: Potential Treatment for Lysosomal Storage Diseases. *Springer*, v.1, pp.9-15, jan. 2015. doi:10.1007/s40778-014-0007-8

DE CARVALHO, Talita G. et al. CRISPR-Cas9-mediated gene editing in human MPS I fibroblasts. *Gene. Science*, v.678, pp.33-37, dez. 2018. doi: 10.1016/j.gene.2018.08.004.

DOENCH, John G.; FUSI, Nicolo; SULLENDER, Meagan et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, v.34, pp.184-191, jan. 2016. doi: 10.1038/nbt.3437

DOUDNA, Jennifer A. & CHARPENTIER, Emmanuelle. The new frontier of genome engineering with CRISPR-Cas9. *Science*, v.346, pp.1258096-1-1258096- 9, nov. 2014. doi: 10.1126/science.1258096

FU, Yanfang; FODEN, Jennifer A.; KHAYTER, Cyd et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, v.31, pp.822-826, jun.2013. doi:10.1038/nbt.2623

GIUGLIANI, Roberto et al. Terapia de reposição enzimática para as mucopolissacaridoses I, II e VI: recomendações de um grupo de especialistas brasileiros. *Revista de Associação Médica Brasileira*, v. 56, pp. 271-277. 2010. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010442302010000300009&lng=pt&nrm=is>

GONZALEZ, Esteban A. Efeitos da inibição da catepsina B sobre a doença cardiovascular na mucopolissacaridose tipo I. 2016 (Dissertação de Mestrado em Genética e Biologia Molecular) - Instituto de Biociências, Universidade do Rio Grande do Sul, Rio Grande do Sul, Porto Alegre, 2016.

HAEUSSLER, Maximilian et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, v.17, pp.1-12, jun. 2016. doi: 10.1186/s13059-016-1012-2

HOBBS, J.R et al. Reversal of clinical features of Hurler's disease and biochemical improvement after treatment by bone-marrow transplantation. *Journal of Inherited Metabolic Disease*, v. 318, pp.705-764, out . 1981 doi: 10.1016/s0140-6736(81)91046-1

LANGMEAD, Ben et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, v.10, pp.25.1-25.10, mar. 2009. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r25>

LESSARD, Samuel et al. (2017). Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *PNAS*, v.114, pp.11257–11266, dez. 2017. Disponível em: <https://www.pnas.org/content/114/52/E11257.long>

LIANG, Puping et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein cell*, v.6, pp.363–372, abr. 2015. doi: 10.1007/s13238-015-0153-5.

LIN, Yanni et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research*, v.42, pp.7473–7485, mai. 2014. doi: 10.1093/nar/gku402

MATTE, Ursula et al. Unique frequency of known mutations in Brazilian MPS I patients. *American Journal of Medical Genetics*, v.90, pp.108-109, mar. 2000. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10607946>

MATTE, Ursula et al. Identification and characterization of 13 new mutations in mucopolysaccharidosis type I patients. *Molecular Genetics and Metabolism*, v.78, pp.37-43, jan. 2003. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1096719202002007>

MONTAGUE, Tessa G. et al. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, v.42, pp. 401-407, jul. 2014. doi: 10.1093/nar/gku410

NEUFELD, Elizabeth F., & MUENZER, Joseph. The Mucopolysaccharidoses [Versão digital em The Online Metabolic and Molecular Bases of Inherited Disease]. 2001 Recuperado de : <https://ommbid.mhmedical.com/content.aspx?bookid=971§ionid=62642135&jumpsectionid=62642144>

OUSSOREN, Esmee et al. Residual α -L-iduronidase activity in fibroblasts of mild to severe Mucopolysaccharidosis type I patients. *Molecular Genetics and Metabolism*, v. 109, pp. 377-381, mai 2013. doi:10.1016/j.ymgme.

PARK, Jeongbin; BAE, Sangsu & KIM, Jin-.Soo. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*, v.31, pp.4014–4016, set. 2015. doi: 10.1093/bioinformatics/btv537.

PASQUALIM, Gabriela. Efeitos da terapia de reposição enzimática com início tardio no modelo murino de mucopolissacaridose do tipo I. 2013 (Dissertação de Mestrado em Genética e Biologia Molecular) - Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Porto Alegre, 2013.

PENATI, Rachele et al. Gene therapy for lysosomal storage disorders: recent advances for metachromatic leukodystrophy and mucopolysaccharidosis I. *Journal of Inherited Metabolic Disease*, v.40, pp.543–554, mai. 2017. doi: 10.1007/s10545-017-0052-4.

POLETTO, Édina et al.. Worldwide distribution of common IDUA pathogenic variants. *Clinical Genetics*, v.94, pp.95-102, jul. 2018. doi: 10.1111/cge.13224.

SCHUH, Roselena et al. In vivo genome editing of mucopolysaccharidosis I mice using the CRISPR/Cas9 system. *Journal of Controlled Release*, v.288, pp.23-33, out. 2018 doi:10.1016/j.jconrel.2018.08.031

SCOTT, David A. & ZHANG, Feng. Implications of human genetic variation in CRISPR-based therapeutic genome editing. *Nature Medicine*, v.23, pp.1095-1101, jul. 2017. doi:10.1038/nm.4377

STEMMER, Manuel et al. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE*, v.10, pp.1-11, abr. 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0124633>

TSAI, Shengdar Q. et al. GUIDE-Seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, v.33, pp. 187-197, dez. 2014. doi:10.1038/nbt.3117

WU, Xuebig et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, v.32, pp.670-676, abr. 2014. doi:10.1038/nbt.2889

YARRINGTON, Robert et al. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *PNAS*, v.115, pp.9351-9358, set. 2018. doi:10.1073/pnas.1810062115.

ZHANG, Liyang et al. The effect of sequence mismatches on binding affinity and endonuclease activity are decoupled throughout the Cas9 binding site. 2017. Disponível em: <https://www.biorxiv.org/content/biorxiv/early/2017/08/15/176255.full.pdf>