Computational analysis of Cas proteins unlocks new potential in HIV-1-targeted gene therapy

Will Dampier[1,2], Rachel Berman[1,2], Michael Nonnemacher[1,2,3], and Brian Wigdahl[1,2,3,4]

[1]Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA, USA, [2]Center for Molecular Virology and Gene Therapy, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, PA, USA, [3]Center for Clinical and Translational Medicine, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, PA, USA, [4]Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA

ABSTRACT

The human immunodeficiency virus type 1 (HIV-1) pandemic has been slowed with the advent of anti-retroviral therapy (ART). However, ART is not a cure and as such has pushed the disease into a chronic infection. One potential cure strategy that has shown promise is the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas gene editing system. It has recently been shown to successfully edit and/or excise the integrated provirus from infected cells and inhibit HIV-1 in vitro, ex vivo, and in vivo. These studies have primarily been conducted with SpCas9 or SaCas9. However, additional Cas proteins are discovered regularly and modifications to these known proteins are being engineered. These alternative Cas molecules have different requirements for protospacer adjacent motifs (PAMs) which impacts the possible targetable regions of HIV-1. Other modifications to the Cas protein or gRNA impact the tolerance for mismatches between gRNA and the target. While reducing off-target risk, this impacts the ability to account for HIV-1 genetic variability. We strive to examine these parameter choices using a computational approach for surveying the suitability of a Cas editor for HIV-1 gene editing. The Nominate, Diversify, Narrow, Filter (NDNF) pipeline measures the safety, broadness, and effectiveness of a pool of potential gRNAs for any PAM. These studies revealed that broader PAMs improve the targeting potential of editors like SaCas9 and LbCas12a have larger pools of useful gRNAs, while broader PAMs reduced the pool of useful SpCas9 gRNAs yet increased the number of targetable locations. Investigation of the mismatch tolerance of Cas editors indicates a 2-missmatch tolerance is an ideal balance between on-target sensitivity and off-target specificity. It is our recommendation that researchers in the HIV-1 gene editing field explore the wider world of Cas editors.

INTRODUCTION

HIV continues to be an ongoing public health problem in the United States and around the world. CRISPR-Cas9 therapy has emerged as a potent tool for editing both viral and human targets to effect a cure. While early editors like SpCas9 are large and have promiscuous recognition, newer editors like SaCas9 are smaller and have more restrictive recognition. Beyond editor choice, biomedical engineers have begun to mutate Cas editors to introduce a number of effects including modifying the PAM. This study developed an in silico measurement system to evaluate Cas:gRNA pairs based on PAM and mismatch recognition. Concurrently, a literature search was conducted to survey the currently available Cas editors and their applications in HIV.

Type	Name	Biological origin	Size (aa)	Spacer (nt)	PAM/PFS	HIV-1 Target	Host Target
II-A	SpCas9	*Streptococcus pyogenes*	1368	20	5′-(PS)-NGG	**Yes**	CCR5, APOBEC3G, Tetherin
II-A	SpCas9-NG	*Streptococcus pyogenes*	1368	20	5′-(PS)-NG		CCR5
II-A	SaCas9	*Staphylococcus aureus*	1053	21-23	5′-(PS)-NNGRRT	**Yes**	CCR5
II-A	SaCas9-KKH	*Staphylococcus aureus*	1053	20-23	5′-(PS)-NNNRRT		CCR5
V-A	LbCas12a (Cpf1)	*Lachnospiraceae bacterium*	1228	23-25	5′-TTTV-(PS)	**Yes**	
V-A	AsCas12a (Cpf1)	*Acidaminococcus sp.*	1307	24	5′-TTTN-(PS)	**Yes**	CCR5
V-B	AaCas12b (C2c1)	*Alicyclobacillus acidiphilus*	1129	20	5′-TTN-(PS)		CCR5
V-E	Cas12e (PlmCasX2)	*Planctomycetes*	978	16-24	5′-TTCN-(PS)		CCR5
VI-A	LbuCas13a (C2c2)	*Leptotrichia buccalis*	1159	24	No PFS requirement	**Yes**	
VI-D	RspCas13d (CasRx)	*Ruminococcus flavefaciens*	922	20	No PFS requirement	**Yes**	

Table 1. Cas editors used in HIV. A selection of the results of a literature review exploring the different Cas editors available for gene editing. This selection is limited to those that have been currently used in HIV.

Library based selection reveals safer and broader gRNAs

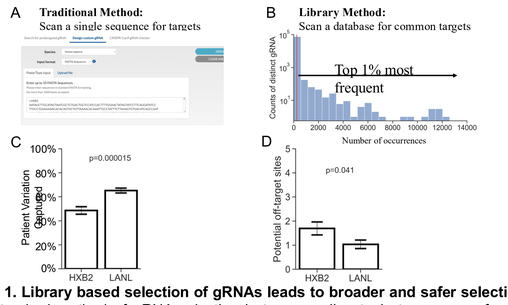
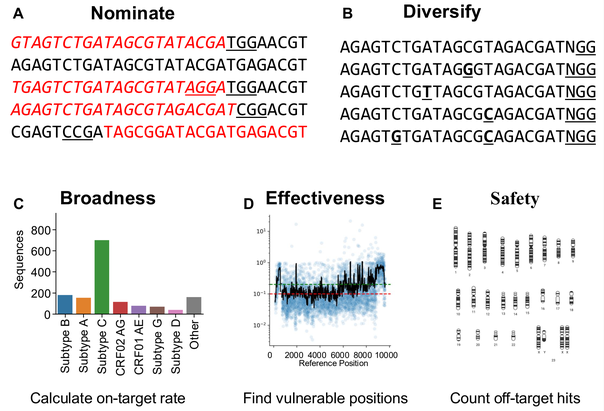


Figure 1. Library based selection of gRNAs leads to broader and safer selections **A)** A standard method of gRNA selection is to use online tools to scan a reference genome. **B)** A library approach scans a large collection of sequences and counts the number of times each occurs. An earlier publication (Chung & Allen 2019, Sci Reports) examined the difference between the two for SpCas9 in which gRNAs were using a library approach selected from the Los Alamos HIV Sequence Database (LANL) or using the single reference strain HXB2. **C)** A bar plot comparing the percentage of a patient cohort targetable by each gRNA. **D)** A bar plot of the number of off-target hits for gRNAs derived from the two approaches.

Evaluation methodology for novel Cas systems



Calculate on-target rate | Find vulnerable positions | Count off-target hits

Figure 2. Schematic steps of the evaluation pipeline. In order to evaluate Cas:gRNA pairs, a pipeline was developed to measure the number of broad, effective, and safe pairs. **A)** First, sites are nominated by searching the nomination set for all potential protospacers in each sequence. **B)** Next, random edits are added to each nominated protospacer to reach a constant 50,000 protospacers. The resulting library is evaluated across three dimensions: **C)** Broadness is measured as the percentage of targetable sequences in a validation set. **D)** Effectiveness is measured by referring to a dataset of known impacts of mutations. **E)** Safety is measured by conducting an in silico off-target evaluation against the human genome.

The HIV-1 genome is not uniformly susceptible to mutation

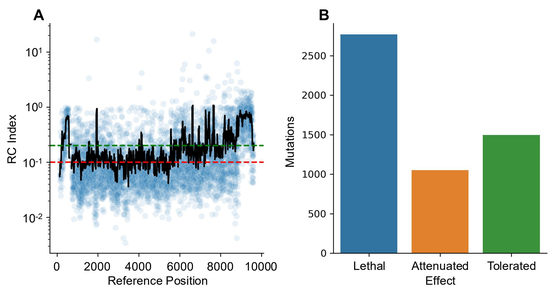


Figure 3. The effect of mutations on HIV replication capacity. A) The RC index of each induced mutation was plotted in blue. The black line represents a 40bp rolling average centered on each point. The red line indicates the maximum RC index of a lethal mutation (0.1). The green line indicates the minimum value of a tolerated mutation (0.2). Between the green and red lines indicates attenuated. B) The number of lethal, attenuated, and tolerated mutations is plotted.

Globally represented validation cohort of full length genomes

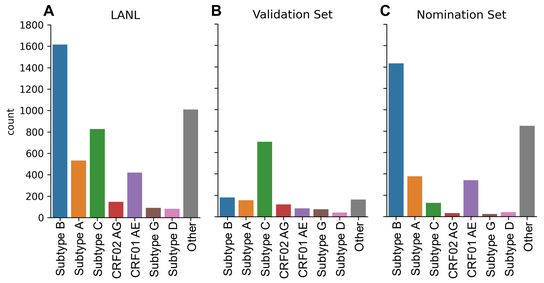


Figure 4. HIV Subtype distribution of full genomic sequences in the LANL dataset. A) The entire 2021 LANL full genome dataset. B) A subset of the LANL database drawn to be a globally representative validation dataset. C) The remaining sequences used as a nomination set.

Different Cas molecules have different targeting profiles

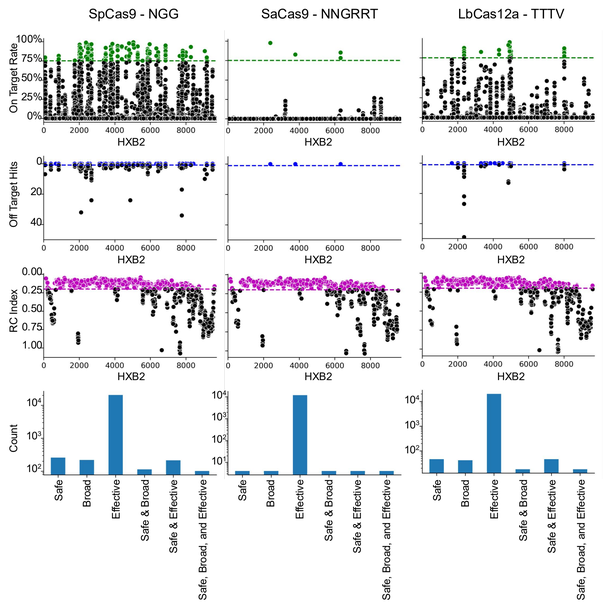


Figure 5. Cas9 editor influences the number of safe, broad, and effective targets. Each column indicates the results of the NDNF pipeline for SpCas9 (left), SaCas9 (middle), and Cpf1 (right). The top row indicates the broad gRNAs by plotting the on-target rate for each gRNA; those in green are above a 75% cutoff and considered broad. The second row indicates the safety of each gRNA by plotting the off-target count for each gRNA. Those with no hits are considered safe. The third row indicates the effectiveness of each gRNA by plotting the average RC index of the 40bp window centered on the cut-site. The fourth row plots the counts of safe, broad, and effective gRNAs and their overlaps.

Two mismatch tolerance provides the optimal balance of on-target rate and off-target safety

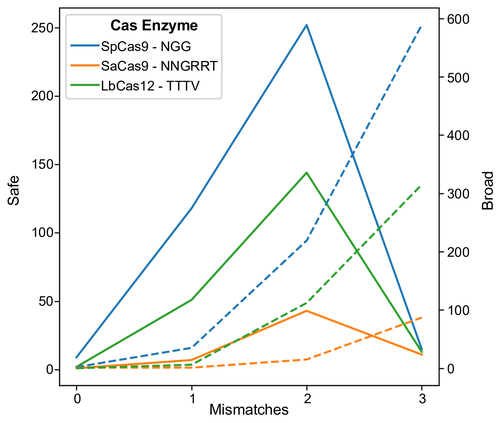


Figure 6. Higher promiscuity leads to more broad targets but fewer safe ones. The number of safe gRNAs was calculated and plotted for the wild-type PAMs each of SpCas9 (blue), SaCas9 (orange), and Cpf1 (green) in solid lines using the left axis. The number of broad gRNAs was calculated and plotted for the wild-type PAMs each of SpCas9 (blue), SaCas9 (orange), and Cpf1 (green) in dotted lines using the right axis.

Modified PAMs increase the number of safe, broad, and effective gRNAs

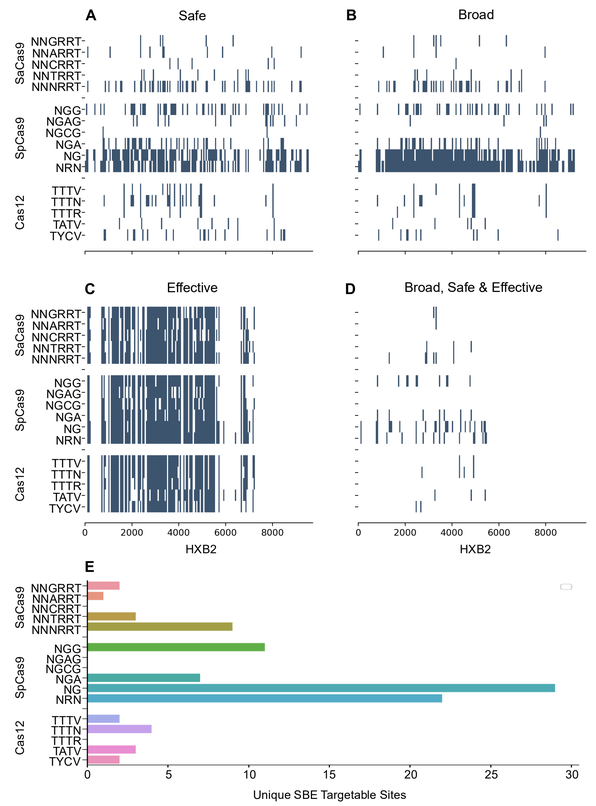


Figure 7. The pattern of targetable regions is altered by PAM mutations. For each of the SpCas9 and SaCas9 variants identified, the pattern of broad (A), safe (B), and effective (C) is plotted against target position in the HXB2 reference genome. D) Indicates the overlapping sites of safe, broad, and effective gRNAs. E) A bar graph showing the number of unique genomic positions targetable by SBE gRNAs for each PAM choice.

Increasing PAM specificity decreases the number of targets

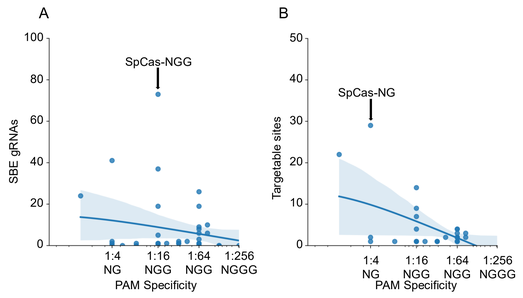


Figure 7. Increase PAM specificity leads to a decrease in SBE gRNAs and loss of targetable sites. The number of A) SBE gRNAs and targetable sites B) was plotted against the PAM specificity. The X-ticks are representative PAMs for each level of specificity. The blue line represents a y~log(x) regression and the shadow indicates a 95% CI

CONCLUSIONS

- New editors open up new targeting possibilities in HIV.
- The optimal balance between broad and safe happens at 2 mismatches.
- Continued research into new editors may further expand the HIV targeting toolbox.

ACKNOWLEDGEMENTS

This work is supported by:
NIMH R01 MH110360 (Contact Multi-PI, BW),
NIMH P30 MH092177 (CNHC/CTRSC, Drexel Component PI, BW),
NIMH T32 MH079785 (Drexel Component PI, BW), and R01 NS089435 (Contact Multi-PI, MRN)