Broad-Spectrum and Personalized Guide RNAs for CRISPR/Cas9 HIV-1 Therapeutics

Will Dampier,[1–3,*] Neil T. Sullivan,[1,2,*] Joshua Chang Mell,[1,4–6] Vanessa Pirrone,[1,2]
Garth D. Ehrlich,[1,4–7] Cheng-Han Chung,[1,2] Alexander G. Allen,[1,2] Mathew DeSimone,[3]
Wen Zhong,[1,2] Katherine Kercher,[1,2] Shendra Passic,[1,2] Jean W. Williams,[1,2] Zsofia Szep,[8,9]
Kamel Khalili,[10,11] Jeffrey M. Jacobson,[10–12] Michael R. Nonnemacher,[1,2] and Brian Wigdahl[1,2,6,9]

Abstract

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas9 system has been used to excise the HIV-1 proviral genome from latently infected cells, potentially offering a cure for HIV-infected patients. Recent studies have shown that most published HIV-1 guide RNAs (gRNAs) do not account for the diverse viral quasispecies within or among patients, which continue to diversify with time even in long-term antiretroviral therapy (ART)-suppressed patients. Given this observation, proviral genomes were deep sequenced from 23 HIV-1-infected patients in the Drexel Medicine CNS AIDS Research and Eradication Study cohort at two different visits. Based on the spectrum of integrated proviral DNA polymorphisms observed, three gRNA design strategies were explored: based on the patient's own HIV-1 sequences (personalized), based on consensus sequences from a large sample of patients [broad-spectrum (BS)], or a combination of both approaches. Using a bioinformatic algorithm, the personalized gRNA design was predicted to cut 46 of 48 patient samples at 90% efficiency, whereas the top 4 BS gRNAs (BS4) were predicted to excise provirus from 44 of 48 patient samples with 90% efficiency. Using a mixed design with the top three BS gRNAs plus one personalized gRNA (BS3 + PS1) resulted in predicted excision of provirus from 45 of 48 patient samples with 90% efficiency. In summary, these studies used an algorithmic design strategy to identify potential BS gRNAs to target a spectrum of HIV-1 long teriminal repeat (LTR) quasispecies for use with a small HIV-1-infected population. This approach should advance CRISPR/Cas9 excision technology taking into account the extensive molecular heterogeneity of HIV-1 that persists *in situ* after prolonged ART.

Keywords: CRISPR/Cas9, HIV-1, excision, gRNA, gRNA pipeline, patient-derived sequence

[1]Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[2]Center for Molecular Virology and Translational Neuroscience, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[3]School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, Pennsylvania.
[4]Center for Genomic Sciences, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[5]Center for Advanced Microbial Processing, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[6]Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania.
[7]Department of Otolaryngology—Head and Neck Surgery, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[8]Division of Infectious Diseases and HIV Medicine, Department of Medicine, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[9]Center for Clinical and Translational Medicine, Institute for Molecular Medicine and Infectious Disease, Drexel University College of Medicine, Philadelphia, Pennsylvania.
[10]Department of Neuroscience, Center for Neurovirology, and Comprehensive NeuroAIDS Center, Lewis Katz School of Medicine, Temple University, Philadelphia, Pennsylvania.
[11]Center for Translational AIDS Research, Lewis Katz School of Medicine, Temple University, Philadelphia, Pennsylvania.
[12]Section of Infectious Disease, Department of Medicine, Lewis Katz School of Medicine, Temple University, Philadelphia, Pennsylvania.
*These authors contributed equally to this work.

Introduction

PATIENTS WHO ADHERE TO a highly active antiretroviral therapy (HAART) regimen typically achieve viral loads below the level of detection of clinical diagnostic assays, but unfortunately almost all patients rebound within 2–4 weeks after interruption of therapy. Early expectations were that a consistent HAART regimen would clear the viral infection. However, this has not been the case, despite improved treatment, probably because of the formation of reservoirs that contain latently infected cells within a number of tissues. The most prominent latently infected cell pool has been shown to reside within the resting CD4+ memory T cell compartment. Clonal proliferation of HIV-1-infected memory T cells maintains and expands the latent reservoir in the absence of efficient detection by the immune system.[1,2] In addition to the CD4+ memory T cell compartment, cells of the central nervous system (CNS; microglial, perivascular macrophages, and astrocytes),[3] cells from the gastrointestinal tract,[4] and many other sites comprise additional reservoirs containing latently infected cells.[5–7]

Current efforts to remove HIV-1 from the CD4+ memory T cell compartment primarily focus on a "shock-and-kill" strategy, in which drugs are used to induce reactivation of virus from latently infected cells. The intended outcome of this therapeutic approach centers on the elimination of latently infected cells after reactivation of HIV-1 transcription, replication, and infectious virus production because of cytopathic effects induced by the virus or by the host immune response. In principle, the latency reversal agents (LRAs) currently under evaluation include histone deacetylase inhibitors, protein kinase C agonists, farnesyl transferase inhibitors, and others as previously reviewed.[6,8] All of these experimental approaches involve varying degrees of activation of viral antigen production that is then targeted by the host immune system. However, in practice, some form of immune boosting will likely be required for this type of strategy to work optimally.[9–11] A recent study has shown that the cytotoxic T lymphocyte (CTL) response depends on which LRA was used, indicating that not all LRAs will be able to induce a robust CTL response.[12] Additional studies involving reversal agent induced CTL suppression has been reported with different classes of LRAs,[13] having shown that a large proportion of the integrated proviruses in the latent reservoir are nonfunctional or unable to produce replication-competent virus under defined experimental circumstances at a given time point; the fraction of latently infected cells was shown to increase the longer the patient was suppressed with ART.[14,15] However, recent studies have indicated that some defective proviruses still have the ability to produce viral proteins, and this may be involved in shaping the overall CTL response, shifting its focus away from cells that have replication-competent proviral insertions.[16] Additional studies have suggested that only some types of CD4+ T cells infected with HIV-1 are being reactivated with the shock-and-kill approach. In this regard, viral outgrowth assays have shown that the proportion of CD4+ T cells reactivated from the resting CD4+ T cell HIV-1 reservoir was typically >100-fold lower than the number of infected cells detectable by quantitative polymerase chain reaction (PCR).[17] In addition, the efficacy of the "shock-and-kill" strategy with respect to eliminating or reducing the number of infected

cells in the monocyte–macrophage lineage or within cells of the CNS remains unknown.

Given these observations, an alternative objective has been to find novel approaches that reduce the size of the latently infected cell population that do not require activation of HIV gene expression or reactivation of virus production.[18] This has largely entailed precise editing of the genome using one of the three main gene-editing systems. The zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are protein-based systems that can efficiently induce double-stranded DNA breaks (DSBs) but require in-depth design and modeling to ensure proper configuration of numerous tandem protein modules to recognize specific sequences in the genome. Unlike ZFNs and TALENSs, the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated Cas9 system is an RNA–protein-based gene-editing system, which has been widely implemented because of the ease of guide RNA (gRNA) design; implementation of specific gRNAs into a variety of CRISPR vector systems; and adaptability with respect to a large number of downstream applications.[19]

The CRISPR/Cas9 system is being used throughout agricultural, biomedical, and basic research domains with ambition of this technology to be used in clinical trials. The two main components in the CRISPR/Cas9 system are the Cas9 endonuclease and the gRNA. The DNA sequence targeted by the CRISPR/Cas9 gRNA, also known as the protospacer, is complementary to the gRNA. This targeting event is facilitated by the recognition of a three-nucleotide protospacer adjacent motif (PAM) followed by a complementary sequence that stretches for ∼20 nucleotides immediately adjacent.[20] Once this binding reaction has been completed, Cas9 induces a DSB and the cellular DNA damage response repairs the break by nonhomologous end joining, thus inducing insertions and deletions (Indels).

The potential of CRISPR/Cas9 as a tool to eliminate the integrated HIV-1 provirus from latently infected cells has been demonstrated *in vitro*, *in vivo*, and more recently *ex vivo*.[21–24] It has also been demonstrated to be capable of disrupting the CCR5 and/or CXCR4 coreceptors for viral entry.[25–35] In addition to inducing Indel mutations around the target site, CRISPR/Cas9 has also been used to excise nearly the entire proviral genome and to activate latent genomes.[21–59] The most promising results have demonstrated that HIV-1-specific gRNAs can target provirus excision from *ex vivo* patient samples.[23] In addition, the use of experimental animal models has demonstrated that CRISPR/Cas9 can be delivered to infected cells in the peripheral blood as well as the CNS and gut associated lymphoid tissue to induce Indels and excise provirus.[22,24] These experiments have produced ground-breaking results with respect to efforts to eliminate or "cure" an HIV-1-infected cell. However, in all these studies, there have always been varying amounts of residual provirus remaining in the infected cell population under investigation. We propose that this is, in part, because of the fact that the gRNAs used do not adequately account for the natural genetic diversity within the integrated HIV-1 proviral genomes within an infected patient, because viral quasispecies (vQS) continuously arise by mutation throughout the course of the disease even after prolonged ART.[60–63]

One major question in designing gRNAs for proviral excision is whether the vQS present within patients are

homogeneous enough to use a small number of gRNAs directed against all proviral reservoirs remaining after successful HAART-mediated suppression? A recent study has sought to measure how genetic changes accumulate in the HIV-1 genomes of patients before and after HAART.[60] It was demonstrated that mutations continue to accumulate long after HAART-mediated suppression and can be clinically important.[60,63] A study of 36 HIV-1-infected patients enrolled in the Drexel Medicine CNS AIDS Research and Eradication Study (CARES) cohort found that—within the peripheral blood mononuclear cell (PBMC) compartment—the predominant long terminal repeat (LTR) sequence from integrated provirus exhibited decreased mutation rates after introduction of HAART, but genetic changes continued to accumulate, even after >6 years of therapy with a median of 10–20 unique mutations per year over the entire LTR region.[60] A similar study by Josefsson et al. examining 12 patients on long-term HAART reported that within CD4+ memory T cells and gut reservoirs, the gag-pol region of the genome mutated at a rate of 50–70 nucleotide changes per 10 kb per year before initiation of HAART, which was reduced to ~ 1 nucleotide change per 10 kb per year during long-term HAART therapy.[64] The 10-fold difference in observed mutation rates after HAART-mediated suppression likely reflects the examination of distinct cell types and HIV-1 genomic regions within the two studies. As a single point mutation can disrupt CRISPR activity, it is important to tailor therapeutic molecules to individual patient vQS or otherwise ensure that gRNA design acts broadly across target sites despite genetic variation at any location across the HIV-1 proviral genomes. Understanding the importance of genetic variation for CRISPR/Cas9 design is further delineated by recent studies that show how variation in the human genome also needs to be considered for gRNAs with potential for therapeutic applications to minimize or avoid unwanted off-target introduction of DSBs. We have previously shown with eight peripheral blood samples from six HIV-1-infected patients that gRNAs tailored to individual patients could be used to target the detectable range of vQS present in each of the eight patient samples, as determined by next-generation sequencing.[61] In this study, longitudinally deep-sequencing data from 23 patients enrolled in the Drexel CARES cohort were used to determine whether broad-spectrum (BS) and personalized gRNA packages can be designed to target the HIV-1 vQS in patient-derived sequences.

Materials and Methods

Patient selection

Patients enrolled in the Drexel Medicine CARES cohort were recruited from the Partnership Comprehensive Care Practice of the Division of Infectious Disease and HIV Medicine in the Department of Medicine at Drexel University College of Medicine (Philadelphia, PA) and the Center for Clinical and Translational Medicine in the Institute for Molecular Medicine and Infectious Disease at the Drexel University College of Medicine. Patients in the Drexel Medicine CARES cohort were recruited under protocol 1201000748 (B.W., Principal Investigator), which adheres to the ethical standards of the Declaration of Helsinki (1964, amended most recently in 2008), which was developed by the World Medical Association as previously described.[65] All patients provided

written consent upon enrollment. For this study, 23 patients were selected that used the following selection criteria: (i) patient samples were available for analysis from three or more patient visits, (ii) patients were currently on HAART, (iii) patients whose last recorded CD4+ T cell count was ≥350 cells/μL, (iv) patients whose last recorded viral load was ≤50 copies/mL, (v) patients who had a visit during the 2016 calendar year, (vi) patients had a normal neurocognitive assessment,[66] (vii) patients were hepatitis C virus-negative at all visits, and (viii) patients who were found to test negative for drugs of abuse at their most recent visit; however, some individuals do have a history of drug abuse (Table 1). Each patient then had samples analyzed from two time points, with two patients having three longitudinal samples.

Deep-sequencing methodology

Amplification of patient HIV-1 LTRs and deep sequencing were performed as previously described.[60,63,65,67] In brief, patient PBMCs were isolated by Ficoll gradient centrifugation. Genomic DNA was isolated and a HIV-1 LTR-specific two-round nested PCR was performed. The amplified LTRs were then purified and tagmented using the Nextera XT Library Prep procedure with the Nextera XT Index procedure v2 to produce the sequencing libraries. Sequencing was performed using the Illumina NextSeq 500 instrument in single-ended mode using a 150 Mid Output v2 sequencing procedure, as previously described by the manufacturer (Illumina). This produced ~ 10 million reads per sample, providing the ability to query diversity of the HIV-1 LTR within PBMCs. All samples were sequenced to an average depth of at least $1,000\times$ across the LTR (positions 36–586) (Supplementary Fig. S1A; Supplementary Data are available online at www.liebertpub.com/aid). The sequences obtained from each patient sample were aligned to the HXB2 reference genome, GenBank accession number K03455, for HIV-1 using the BWA mem, version 0.7.12 algorithm with default settings.[68] The frequency of each base at each position was calculated using the samtools mpileup algorithm, version 1.3.[69] Supplementary Figure S1B

TABLE 1. DEMOGRAPHICS OF THE SUBSET OF PATIENTS SELECTED FOR LTR SEQUENCING AND GUIDE RNA DESIGN

|	Longitudinal sampling n = 23 patients, 48 samples
Age, years	49 ± 8.5
History of drug use	
Yes	18.7%
No	81.2%
Gender	
Male	56%
Female	40%
CD4 (cells/mL)	
Latest	621 ± 355
Nadir	221 ± 203
Viral load (copies/μL)	
Latest	Undetectable (<48)
Peak	93,280 ± 505,977
Years seropositive	21.9 ± 8.09
Years on ART	12.8 ± 6.6

ART, antiretroviral therapy; LTR, long terminal repeat.

shows the average entropy at each position within the LTR across the 48 sequenced samples. Entropy was calculated to demonstrate the extensive intrapatient variation that exists throughout the LTR; the higher the entropy the higher the variation. The pileup files were parsed using the pysam package dropping all Indels and bases with a quality score >10. It was assumed that sequencing errors would be rare when compared with the consensus nucleotide. Indels were considered a fifth base at the point of insertion that, for calculation purposes, did not match any nucleotide. The final aligned sequences have been made available at the NCBI short-read archive (SRA), and all samples were linked under Bioproject ID PRJNA309974.

gRNA design

From the positional frequency data of the vQS, gRNAs were designed in a two-stage process. First, the consensus sequence of the 21 nt 5′ of all GG dinucleotides in the reference sequence (on both strands) was extracted from each patient sample to produce a candidate list of gRNAs. The consensus sequence was selected as the gRNA, as opposed to the most common 21-mer, because the consensus sequence will have the lowest number of mismatches when compared across all reads; this is not always true of the most common 21-mer. Second, the methodology, previously introduced in Dampier *et al.*, was expanded to work with next generation sequencing (NGS) data.[62] This was accomplished by utilizing the Massachusetts Institute of Technology (MIT) penalty function[70] (Fig. 1) to find gRNAs with a BS activity. An MIT penalty score of 1 indicates a perfect match and assumes 100% predicted cleavage (Fig. 1C, D). To predict what fraction of a patient's vQS would be cleaved by a particular gRNA candidate, a two-step process was used. First, all aligned reads overlapping the gRNA target site were extracted from the alignment. Second, a sliding window was used on each read to determine the likelihood that a query gRNA would be predicted to cleave that patient-derived target sequence even in the presence of small Indels within
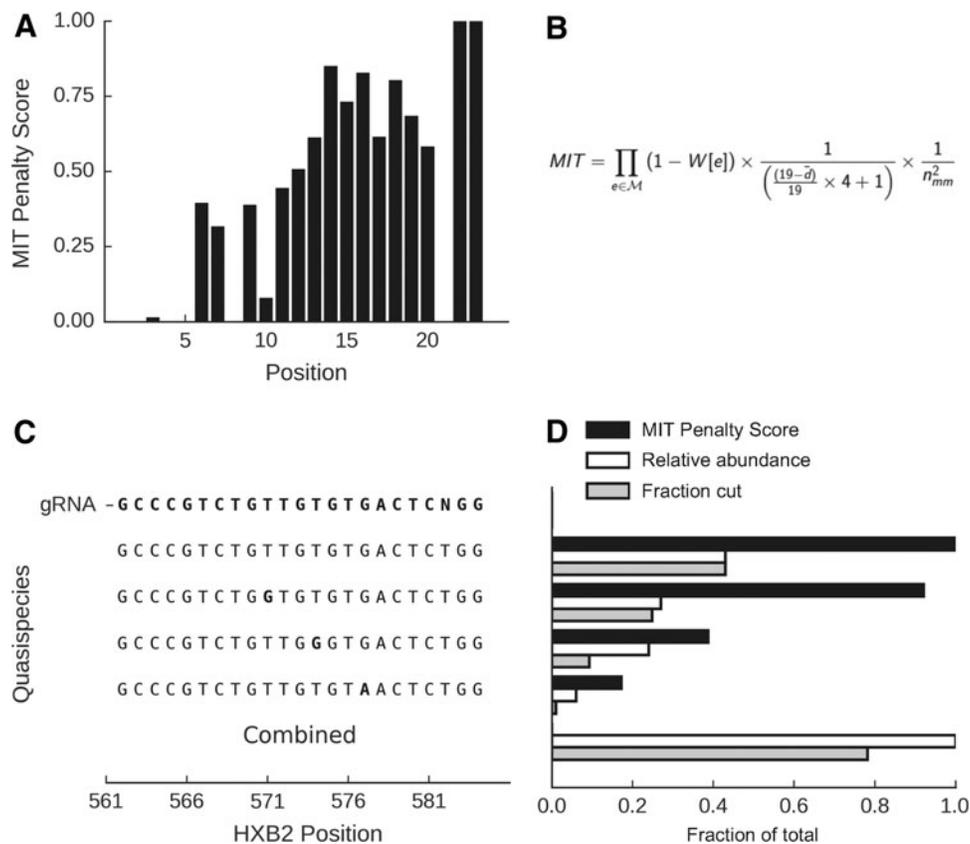
FIG. 1. gRNAs were designed against patient-derived HIV-1 quasispecies. The position-specific penalties associated with mismatches within the protospacer used the MIT penalty score (**A**) as described by the equation (**B**) derived by Hsu *et al.*[70] W[e] is a penalty function that evaluates to 0 if position e is a match between the gRNA and the target and the position-specific penalty M in the event of a mismatch at position e. d is the arithmetic mean of pairwise distance between gRNA and target across the entire 20-mer and n represents the number of mismatches between the target and the gRNA. The MIT penalty score ranges from 1, in the case of a perfect match, to 0, in the case of a complete mismatch, and is interpreted as the percentage of strands that would be predicted to be cleaved. (**C**) All patient quasispecies sequences with a frequency >1% from sample A0010-R06 spanning the target of the gRNA (HXB2 561–581); mutations from the ideal target are marked in *bold*. (**D**) For each of the quasispecies of sample A0010-R06, the MIT penalty score is shown in *black*, the relative abundance of the sequence is shown in *white*, and the fraction of the quasispecies expected to be cut by the gRNA is shown in *gray*. For example, the most abundant quasispecies accounts for 40% of the total and the gRNA exactly matches the sequence (*row 1*). As such, the entire quasispecies has been assumed to be cut. The second most abundant has a penalty score of 0.9, as such only 36% of that fraction was predicted to be removed. gRNA, guide RNA; MIT, Massachusetts Institute of Technology.

the read. This result was then averaged across all reads that cover the target region. Base quality scores were not considered in this step. It is mathematically more likely that a sequencing error would decrease the binding score and as such these values underestimate the likely predicted cleavage efficiency. The predicted cleavage of a gRNA for a particular patient-derived sequence was determined up to the limit of detection, which is determined by the sequencing depth across the target position. Multiplexes of gRNAs were considered as independent observations and aggregated by calculating the likelihood that any gRNA would be predicted to cut the target sequence. The final list of gRNAs was then screened against the human genome using a BLAST query against HG19. Considering a gRNA has 20 nucleotide positions, the gRNAs that had 17 or more nucleotide matches at any site in the human genome were excluded, as they have a high chance of binding to the human genome. The gRNAs that had 16 or less matches to the human genome were kept, as these have minimal if any chance of binding to the human genome and causing off-target effect.

gRNA selection

The final candidate list of gRNAs was sorted by the total percentage of observed vQS that would be predicted to be cleaved within that particular sample (personalized) or across all samples (BS) in the study. With the assumption that the final delivery vector could accommodate a multiplexed delivery of four gRNAs (maximum currently able to be placed into a lentiviral vector system), three types of gRNA packages were designed: the top four gRNAs with the greatest total cleavage potential within each individual (personalized; PS1–PS4) or across all samples (BS; BS4), as well as the top three BS gRNAs plus one personalized gRNA for each individual (3 + 1 package; BS3 + PS1). The personalized gRNA was selected from the earlier of the two visits from each patient and the most effective patient-specific gRNA was added to the collection.

Results

A personalized gRNA package was effective at vQS recognition over an extended period of time

Many of the gRNAs specific to HIV-1 have been designed from a single input sequence or from a small number of laboratory strains. In this study and unique when compared with others, anti-HIV-1 gRNAs specific for the HIV-1 LTR were designed based on thousands of patient-derived HIV-1 sequences. To design HIV-1-specific gRNAs, 23 patients with at least two visits were selected for deep sequencing of the HIV-1 LTR from the Drexel Medicine CARES cohort (described in the section ''Materials and Methods'' and Table 1). PBMC isolation from peripheral blood samples and HIV-1-specific LTR PCR for NGS was performed as previously described.[63,65,67] Using the algorithm already described, gRNAs were designed for each of the 48 samples to optimize cleavage efficiency in the face of LTR genetic diversity. The algorithm identified regions of high conservation within the LTR and ensured they matched the seed region (positions 15–20) within the gRNA (positions within the protospacer associated with high penalty), whereas LTR regions with high diversity were accommodated with proto-

spacer positions with little or no associated penalty (Fig. 1). Of the 48 samples examined, 45 were cleavable to the limit of detection with 3–8 gRNAs (Fig. 2A). This implies that the vector design needs to accommodate at least four configurable gRNAs, which has been shown for a lentiviral vector previously, and more recently using an adeno-associated virus (AAV) vector.[24,71] When the top 4 gRNAs were individually used for each sample to calculate the predicted percentage cutting against its own sample (personalized), 43 of the 48 were predicted to be cleaved to 99% efficiency and all 48 samples would be cleaved to 90% efficiency (Fig. 2B).

If it is necessary to devise a personalized set of gRNAs to excise the integrated provirus, it is obvious that there will be a delay between sample isolation, sequencing of the vQS, design of the gRNAs, and construction of the therapeutic for delivery. At present, the length of this time frame is not known. However, as indicated in our previous study,[61] designing gRNAs for one visit may not be as effective on subsequent visits because of continued evolution of proviral sequences. However, that statement was made with only a few longitudinal samples evaluated. Given that these 48 samples were from 23 patients with two or three visits each, this larger data set was used to more accurately address this question. Utilizing the current data set, there was not a consistent pattern with respect to the effectiveness of personalized therapies as time progresses (Fig. 2C). In a total of 16 of the visit pairs, a decreased effectiveness was observed while 12 either maintained or improved in their effectiveness regarding predicted cleavage efficiency. This implies that there may be a time limit between the sampling and delivery of a gRNA package for some patients. This result also reconfirms that some patients on HAART still have ongoing, although low level, genetic variation that could have a functional impact on the efficacy of therapy. However, the level of genetic variation was also shown to be patient specific, with some personalized gRNAs remaining effective at cleaving the vQS even up to 70 months later (Fig. 2C). Therefore, continued NGS analysis of longitudinal samples is important to designing gRNAs to tackle the entire vQS within and among infected individuals.

A BS gRNA package can be designed to eliminate the vast majority of vQS in a given patient

When the LTRs from all 48 samples were pooled together, there were five gRNAs that had a cutting fraction >10%, with the optimal one being able to cut ∼90% of the quasispecies observed in all samples (Fig. 3A). The top four gRNAs were predicted to cut the majority of patient-derived sequence in all samples and were used in further analysis as the BS package (BS4). The gRNA sequences are included in Supplementary Table S1. One of the major questions in gRNA design is how well gRNAs designed from one set of data can potentially cut proviral DNA from nonevaluated specimens. We utilized the methodology previously described in Dampier et al. to evaluate how well this package of gRNAs would function across the patient-derived sequences from the Los Alamos National Laboratory Database (https://hiv.lanl.gov/content/sequence/HIV/mainpage.html).[62] For the four gRNAs contained in the BS4 package, these were predicted to cleave 89%, 83%, 74%, and 76% of the samples in the database. These four gRNAs were then further explored to determine how effective they would be across all the HIV-1 NGS data
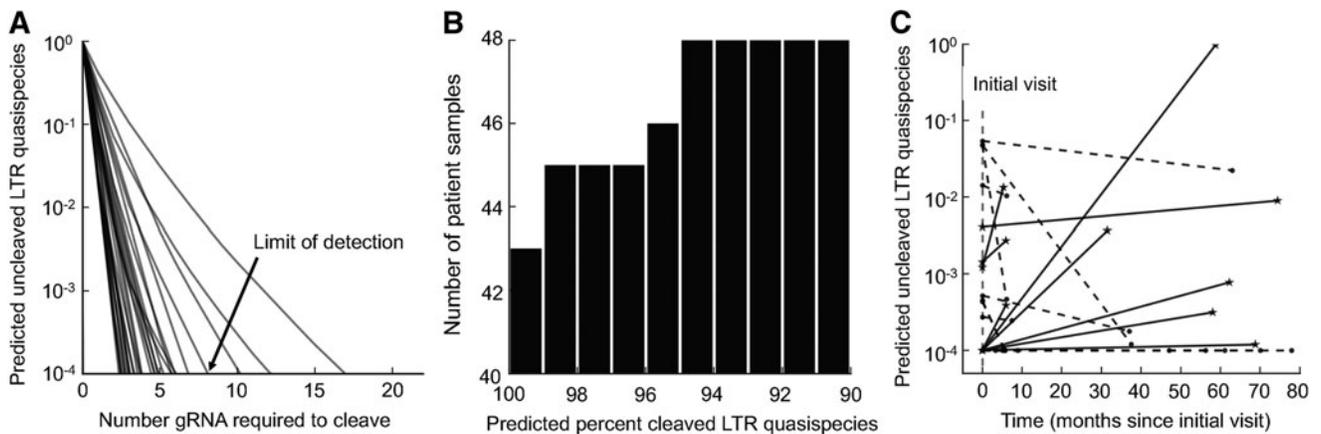
FIG. 2. Personalized gRNAs can target the quasispecies with longitudinal efficacy. **(A)** The number of gRNAs needed to cleave the fraction of the viral quasispecies for each patient. Most patients require three to eight gRNAs at the current level of NGS detection (1:10,000). The three *lines* to the *right* of the *arrow* indicate the level of detection of the three that are not captured by the three to eight gRNAs. **(B)** A histogram indicating the number of samples that pass a particular cleavage threshold when tested with a package of four gRNAs personalized to the sample. **(C)** The samples from 23 patients were examined longitudinally by designing a package of four personalized gRNAs against one visit and evaluating the effectiveness on a subsequent visit from the same patient. *Solid lines* indicate instances in which the treatment has become less effective and *dotted lines* indicate the treatment has either been maintained or improved in effectiveness. The effectiveness was limited by the level of detection (10^{-4}). NGS, next generation sequencing.

derived within this study. Given that LTR-A and LTR-B were designed from previously published HIV-1 sequences[21–23] versus the gRNAs designed specifically from the CARES cohort here, LTR-A and LTR-B combined (LTR-A/B) were examined with respect to how well they would cut the CARES patient samples versus the BS gRNAs from those designed against the pool of 48 samples (Fig. 3B). To keep

the comparison equal, only the top two BS gRNAs were used when comparisons were made with LTR-A and LTR-B. When using a cutoff of 90% cleavage: LTR-A/B were only predicted to be effective against five patient samples, whereas the top 2 gRNAs from the BS package (BS-2) would be effective against 44 patient samples. A more conservative cutoff of targeting 99% of the vQS reveals that
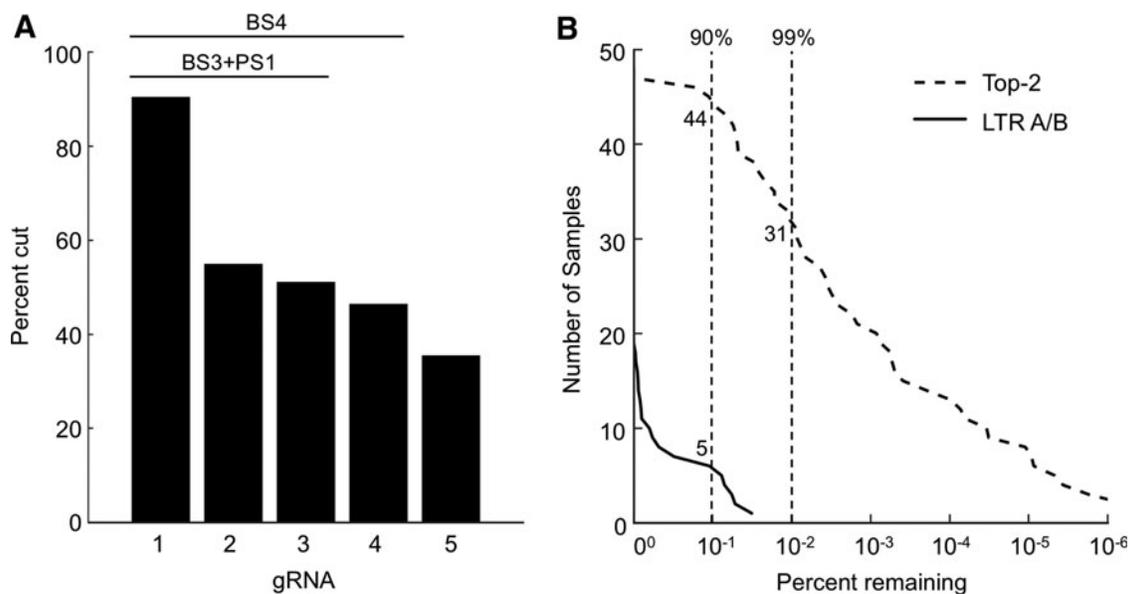

FIG. 3. A package of four gRNAs has potential BS cutting efficiency. **(A)** All five gRNAs have >10% predicted cutting fraction of patient-derived sequences from the CNS AIDS Research and Eradication Study cohort. **(B)** The top two BS gRNAs were compared with gRNAs LTR-A/B[21–23] to determine how many samples would be predicted to be cleaved to a particular cutoff. The results obtained with the top two gRNAs (Top-2) are shown as *dotted line* and the results obtained with LTR-A/B gRNAs are shown in *black*. The graph is extended to include predicted cuts <10%, so all 48 samples are presented. Results obtained with cutting efficiencies of 90% and 99% are indicated. BS, broad-spectrum.

the BS-2 package would be effective against 31 samples. It should be noted that BS-2 gRNAs were specifically designed from this data set and it is currently unclear how they would perform on an unsampled individual. However, these predictive analyses suggest that the efficacy of gRNA design can be enhanced to improve the BS activity of a given gRNA package designed to target the complex HIV-1 quasispecies.

Owing to practical limitations in delivering a purely personalized treatment and the high success rate of the BS package, a hybrid method (utilizing some consensus gRNAs and some personalized gRNAs) was used to compare with a BS (BS4)-only approach. In theory, this process could involve the production, validation, and experimental utilization of a package of three BS gRNAs along with a more extensive library of personalized gRNAs. The design process was examined by combining the top three BS gRNAs (BS3) with one personalized gRNA (PS1) to see whether this approach would increase the efficacy of the package of gRNAs (Fig. 4). The targets of BS3 and PS1 anti-HIV-1 gRNAs in each patient are given in Supplementary Table S1. These analyses demonstrated that the BS3 + PS1 package produced comparable results to those obtained utilizing the personalized gRNA or BS4 at 90% efficiency (45 patient samples predicted to cut with BS3 + PS1 vs. 46 with personalized, vs. 44 with BS4). However, at higher efficiency cutoff of 99%, the BS4 package underperforms personalized therapy (Table 2). The personalized gRNAs are not clustered in any particular
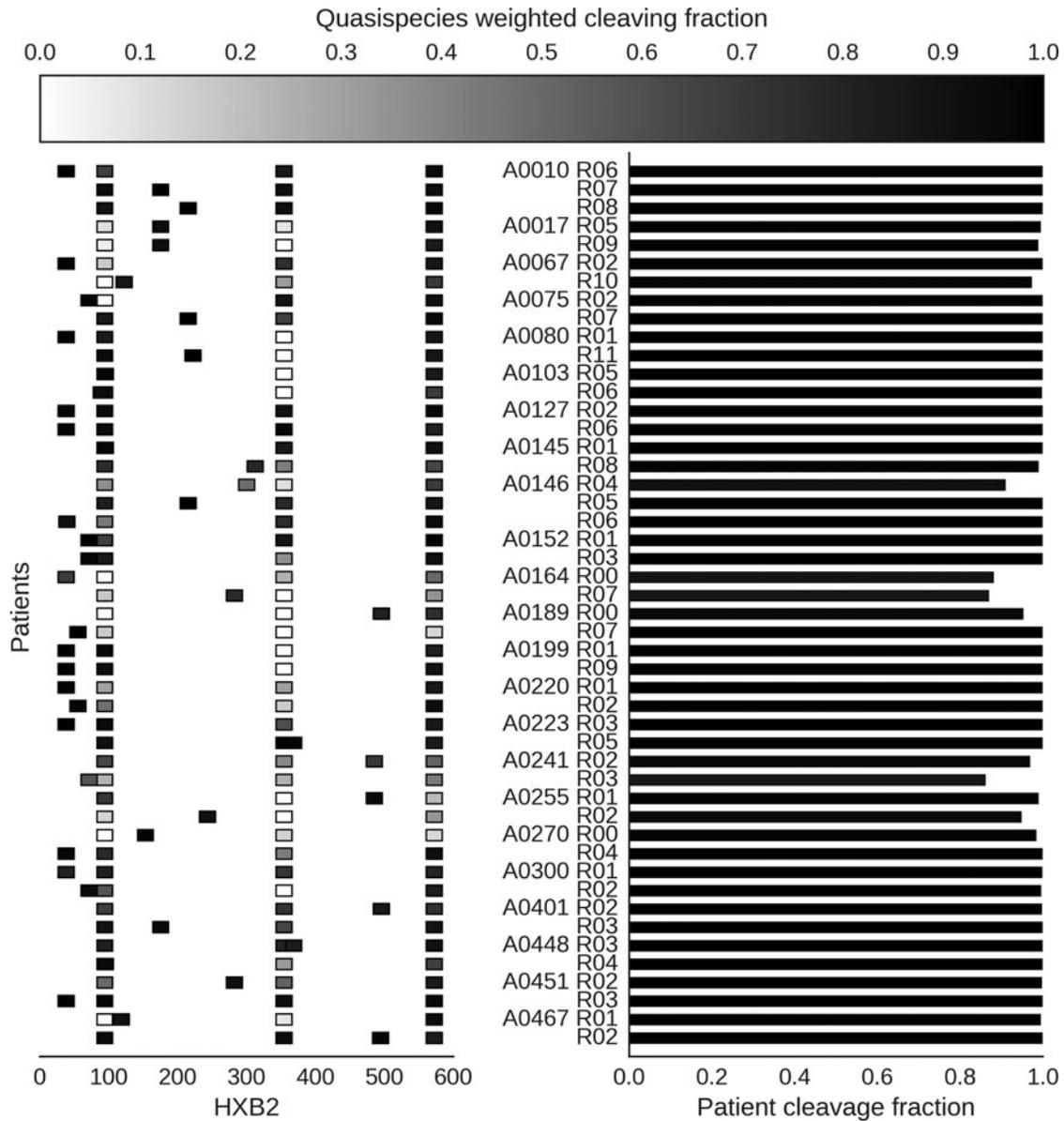
FIG. 4. Combining three BS gRNAs with a single personalized gRNA (BS3 + PS1) is predicted to cleave the majority of the patient quasispecies. On the *left*, the positions of each gRNA for each patient are shown with the *shade* indicating the weighted average of the quasispecies that was predicted to be cleaved. On the *right*, the combined weighted cleavage fraction was calculated for each patient assuming each position was independent.

Table 2. Overall Cutting Efficiencies
of the Methods Tested

	Predicted cutting efficiency	
	90%	99%
Personalized	46	43
BS (top four, BS4)	44	31
BS3 + PS1 (top three plus one personalized)	45	36

BS, broad-spectrum.

part of the LTR as indicated by the large variation in their optimal targeting positions (Fig. 4).

Discussion

When considering gRNA design for HIV-1 excision therapy, one of the most important concerns is the extent that the therapy will be limited by the variation in the HIV-1 genome over time because of both the high mutation rate of the HIV-1 reverse transcriptase[72] and recombination events.[73] This rapid rate of mutation and a host of selective pressures ultimately lead to a swarm of similar, yet distinct, HIV-1 genomes within a given patient. This study has shown that it is possible to select a set of anti-HIV-1 LTR gRNAs that are expected to cleave all or nearly all vQS within an individual by taking advantage of the inherent promiscuity of the CRISPR/Cas9 system. We have demonstrated the use of patient-derived sequence information to inform gRNA design that can enhance CRISPR/Cas9 gRNA specificity rather than using a single input sequence to try and capture a diverse population. Furthermore, this study shows that a small package of gRNAs can be selected such that they will likely cleave a majority of the vQS within and across the patients tested.

To our knowledge, this is the first study to have designed HIV-1-specific gRNAs that are predicted to bind to a spectrum of vQS (within the limit of detection) for a vast majority of patients tested. Although current HIV-1-specific gRNAs are effective in a number of assays performed,[2,21–30,32–45,47–59] they were not screened to test patient-derived sequences and indeed were sometimes specifically designed using the genomes of laboratory strains in an attempt to inhibit the expression of a particular HIV-1 gene, the activity of the viral promoter (LTR), or to abrogate viral replication. The gRNAs that are reported herein (Supplementary Table S1) are ideal for testing in a number of different therapeutic avenues for three reasons: (i) they can recognize all vQS within an infected patient to a given limit of detection, (ii) they are likely to prevent or dramatically delay virus escape due to being multiplexed and targeting conserved regions in a patient-specific manner, and (iii) could be easily incorporated into CRISPR/Cas9-specific lentiviral vectors or AAV that can accommodate up to four gRNAs.

There has been extensive research into *in silico* methods to predict the likelihood of CRISPR/Cas9 cleavage of particular gRNA:target pairs. A position-specific scoring matrix of CRISPR/Cas9 has been developed to estimate the likelihood of off-target effects from a given gRNA using systematically mismatching gRNA variants on the same target.[70] The promiscuous nature of gRNA targeting has been one of the primary concerns with regard to therapeutic gene-editing technology. A more recent study further expands the gRNA matrix to 12×20 possible substitutions as well as 64 possible PAMs.[74] In addition, these gRNA search algorithms were developed to improve off-target prediction by incorporating scoring matrices and evidence-based rules as previously reviewed.[75,76] These algorithms were incorporated into the pipeline presented here.

Along with the challenge of designing the gRNAs necessary for the excision of all integrated proviruses, the other major challenge will be delivery of the gRNAs. One way that has been proposed is the use of lentiviral and AAV vectors. The lentiviral delivery system would allow for the infection of similar cell types as HIV naturally infects, potentially limiting the delivery of therapy to uninfected target cells. However, the efficiency of vectors with respect to delivery to latently or persistently infected cells may be complicated by both the decreased infection rates of these types of cells, which have been previously infected by HIV-1 and show a more resistant phenotype to secondary HIV-1 infection. The lentiviral delivery system is also limited in the number of gRNAs that can be packaged into a single vector construct. Spreading the gRNAs across many vectors may result in a cell not receiving the gRNAs required for proper targeting. The nature of this limit has yet to be fully defined. Finally, even if the previous hurdles are overcome, an ideal therapy needs to deliver the gRNAs to tissue resident cells that are present at low numbers that may require lentiviral migration across vascular barriers.

The CRISPR/Cas9 system has been shown *in vitro*, *in vivo*, and *ex vivo* as a novel and promising method for treating and possibly curing HIV infection.[21–24] In this analysis, we have demonstrated that the proposed gRNA design pipeline can lead to the production of anti-HIV-1 gRNAs that can span multiple viral genotypes as well as being patient specific. The BS4 or BS3 + PS1 strategy *in silico* clearly surpasses current gRNAs with respect to cleaving patient-derived HIV sequences that are detectable by NGS across relatively large number of HIV-1-infected patients. With an increased number of longitudinal patient-derived PBMC samples that have been subjected to NGS, this gRNA design pipeline has been shown to be capable of designing optimal gRNAs to target a broader spectrum of patients, from multiple cohorts. Future studies will be needed to examine the potential of gRNAs to target multiple HIV subtypes, beyond subtype B that is common in North America and Europe, to hopefully achieve a functional and/or sterilizing HIV cure.

Sequence Data

Final sequences have been submitted to the SRA and all samples were linked under Bioproject ID PRJNA309974.

Acknowledgments

Author Contributions

W.D., M.R.N., and B.W. conceptualized the article. W.D., M.D., and M.R.N. analyzed the data and wrote the first draft of the article. W.D., N.T.S., M.R.N., V.P., C.-H.C., A.G.A., G.D.E., J.C.M., and B.W. contributed to writing and made critical revisions of the article. Z.S. and J.M.J. aided in recruiting patients. M.D., W.Z., K.K., S.P., and J.W.W. contributed to patient sample preparation and experimentation. All authors reviewed and approved the final version of the article.

Author Disclosure Statement

No competing financial interests exist.

References

1. Hosmane NN, Kwon KJ, Bruner KM, *et al.*: Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J Exp Med 2017;214:959–972.

2. Maldarelli F, Wu X, Su L, *et al.*: HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science 2014;345:179–183.

3. Avalos CR, Abreu CM, Queen SE, *et al.*: Brain macrophages in simian immunodeficiency virus-infected, antiretroviral-suppressed macaques: A functional latent reservoir. mBio 2017;8 [Epub ahead of print]; DOI: 10.1128/mBio.01186-17.

4. Estes JD, Kityo C, Ssali F, *et al.*: Defining total-body AIDS-virus burden with implications for curative strategies. Nat Med 2017;23:1271–1276.

5. Barton K, Winckelmann A, Palmer S: HIV-1 reservoirs during suppressive therapy. Trends Microbiol 2016;24:345–355.

6. Datta PK, Kaminski R, Hu W, *et al.*: HIV-1 latency and eradication: Past, present and future. Curr HIV Res 2016; 14:431–441.

7. Mzingwane ML, Tiemessen CT: Mechanisms of HIV persistence in HIV reservoirs. Rev Med Virol 2017;27 [Epub ahead of print]; DOI: 10.1002/rmv.1924.

8. Schwartz C, Bouchat S, Marban C, *et al.*: On the way to find a cure: Purging latent HIV-1 reservoirs. Biochem Pharmacol 2017;146:10–22.

9. Ferrari G, Pollara J, Tomaras GD, Haynes BF: Humoral and innate antiviral immunity as tools to clear persistent HIV infection. J Infect Dis 2017;215(Suppl 3):S152–S159.

10. Riley JL, Montaner LJ: Cell-mediated immunity to target the persistent human immunodeficiency virus reservoir. J Infect Dis 2017;215(Suppl 3):S160–S171.

11. Shan L, Deng K, Shroff NS, *et al.*: Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. Immunity 2012; 36:491–501.

12. Jones RB, Mueller S, O'Connor R, *et al.*: A subset of latency-reversing agents expose HIV-infected resting CD4+ T-cells to recognition by cytotoxic T-lymphocytes. PLOS Pathog 2016;12:e1005545.

13. Kwaa AK, Goldsborough K, Walker-Sperling VE, Pianowski LF, Gama L, Blankson JN: The effect of Ingenol-B on the suppressive capacity of elite suppressor HIV-specific CD8+ T cells. PLoS One 2017;12:e0174516.

14. Cohn LB, Silva IT, Oliveira TY, *et al.*: HIV-1 integration landscape during latent and active infection. Cell 2015;160: 420–432.

15. Ho YC, Shan L, Hosmane NN, *et al.*: Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell 2013;155:540–551.

16. Pollack RA, Jones RB, Pertea M, *et al.*: Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. Cell Host Microbe 2017;21:494–506.e4.

17. Eriksson S, Graf EH, Dahl V, *et al.*: Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. PLoS Pathog 2013;9:e1003174.

18. Bruner KM, Hosmane NN, Siliciano RF: Towards an HIV-1 cure: Measuring the latent reservoir. Trends Microbiol 2015; 23:192–203.

19. Manjunath N, Yi G, Dang Y, Shankar P: Newer gene editing technologies toward HIV gene therapy. Viruses 2013;5: 2748–2766.

20. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337:816–821.

21. Hu W, Kaminski R, Yang F, *et al.*: RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc Natl Acad Sci U S A 2014;111:11461–11466.

22. Kaminski R, Bella R, Yin C, *et al.*: Excision of HIV-1 DNA by gene editing: A proof-of-concept in vivo study. Gene Ther 2016;23:690–695.

23. Kaminski R, Chen Y, Fischer T, *et al.*: Elimination of HIV-1 genomes from human T-lymphoid cells by CRISPR/Cas9 gene editing. Sci Rep 2016;6:22555.

24. Yin C, Zhang T, Qu X, *et al.*: In vivo excision of HIV-1 provirus by saCas9 and multiplex single-guide RNAs in animal models. Mol Ther 2017;25:1168–1186.

25. Choi JG, Dang Y, Abraham S, *et al.*: Lentivirus pre-packed with Cas9 protein for safer gene editing. Gene Ther 2016; 23:627–633.

26. Hofer U, Henley JE, Exline CM, Mulhern O, Lopez E, Cannon PM: Pre-clinical modeling of CCR5 knockout in human hematopoietic stem cells by zinc finger nucleases using humanized mice. J Infect Dis 2013;208(Suppl 2): S160–S164.

27. Hou P, Chen S, Wang S, *et al.*: Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. Sci Rep 2015;5:15577.

28. Hutter G, Bodor J, Ledger S, *et al.*: CCR5 targeted cell therapy for HIV and prevention of viral escape. Viruses 2015;7:4186–4203.

29. Kang H, Minder P, Park MA, Mesquitta WT, Torbett BE, Slukvin II: CCR5 disruption in induced pluripotent stem cells using CRISPR/Cas9 provides selective resistance of immune cells to CCR5-tropic HIV-1 virus. Mol Ther Nucleic Acids 2015;4:e268.

30. Li C, Guan X, Du T, *et al.*: Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using

adenovirus-delivered CRISPR/Cas9. J Gen Virol 2015;96: 2381–2393.

31. Liu Z, Chen S, Jin X, *et al.*: Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4(+) T cells from HIV-1 infection. Cell Biosci 2017; 7:47.

32. Wang Q, Chen S, Xiao Q, *et al.*: Genome modification of CXCR4 by *Staphylococcus aureus* Cas9 renders cells resistance to HIV-1 infection. Retrovirology 2017;14:51.

33. Wang W, Ye C, Liu J, Zhang D, Kimata JT, Zhou P: CCR5 gene disruption via lentiviral vectors expressing Cas9 and single guided RNA renders cells resistant to HIV-1 infection. PLoS One 2014;9:e115987.

34. Xu L, Yang H, Gao Y, *et al.*: CRISPR/Cas9-mediated CCR5 ablation in human hematopoietic stem/progenitor cells confers HIV-1 resistance in vivo. Mol Ther 2017;25: 1782–1789.

35. Yu S, Yao Y, Xiao H, *et al.*: Simultaneous knockout of CXCR4 and CCR5 Genes in CD4+ T cells via CRISPR/Cas9 confers resistance to both X4- and R5-tropic human immunodeficiency virus type 1 infection. Hum Gene Ther 2018;29:51–67.

36. Bialek JK, Dunay GA, Voges M, *et al.*: Targeted HIV-1 latency reversal using CRISPR/Cas9-derived transcriptional activator systems. PLoS One 2016;11:e0158294.

37. Ebina H, Kanemura Y, Misawa N, *et al.*: A high excision potential of TALENs for integrated DNA of HIV-based lentiviral vector. PLoS One 2015;10:e0120047.

38. Ebina H, Misawa N, Kanemura Y, Koyanagi Y: Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. Sci Rep 2013;3:2510.

39. Honda A, Hirose M, Sankai T, *et al.*: Single-step generation of rabbits carrying a targeted allele of the tyrosinase gene using CRISPR/Cas9. Exp Anim 2015;64:31–37.

40. Huang Z, Nair M: Nair M. A CRISPR/Cas9 guidance RNA screen platform for HIV provirus disruption and HIV/AIDS gene therapy in astrocytes. Sci Rep 2017;7:5955.

41. Ji H, Jiang Z, Lu P, *et al.*: Specific reactivation of latent HIV-1 by dCas9-SunTag-VP64-mediated guide RNA targeting the HIV-1 promoter. Mol Ther 2016;24:508–521.

42. Kaminski R, Chen Y, Salkind J, *et al.*: Negative feedback regulation of HIV-1 by gene editing strategy. Sci Rep 2016; 6:31527.

43. Kim V, Mears BM, Powell BH, Witwer KW: Mutant Cas9-transcriptional activator activates HIV-1 in U1 cells in the presence and absence of LTR-specific guide RNAs. Matters (Zur) 2017;2017.

44. Lebbink RJ, de Jong DC, Wolters F, *et al.*: A combinational CRISPR/Cas9 gene-editing approach can halt HIV replication and prevent viral escape. Sci Rep 2017;7:41968.

45. Liao HK, Gu Y, Diaz A, *et al.*: Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells. Nat Commun 2015;6:6413.

46. Mandal PK, Ferreira LM, Collins R, *et al.*: Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell 2014;15:643–652.

47. Qu D, Li C, Sang F, *et al.*: The variances of Sp1 and NF-kappaB elements correlate with the greater capacity of Chinese HIV-1 B′-LTR for driving gene expression. Sci Rep 2016;6:34532.

48. Saayman SM, Lazar DC, Scott TA, *et al.*: Potent and targeted activation of latent HIV-1 using the CRISPR/dCas9 activator complex. Mol Ther 2016;24:488–498.

49. Schumann K, Lin S, Boyer E, *et al.*: Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc Natl Acad Sci U S A 2015;112:10437–10442.

50. Ueda S, Ebina H, Kanemura Y, Misawa N, Koyanagi Y: Anti-HIV-1 potency of the CRISPR/Cas9 system insufficient to fully inhibit viral replication. Microbiol Immunol 2016;60:483–496.

51. Wang G, Zhao N, Berkhout B, Das AT: A combinatorial CRISPR-Cas9 attack on HIV-1 DNA Extinguishes all infectious provirus in infected T cell cultures. Cell Rep 2016; 17:2819–2826.

52. Wang G, Zhao N, Berkhout B, Das AT: CRISPR-Cas9 can inhibit HIV-1 replication but NHEJ repair facilitates virus escape. Mol Ther 2016;24:522–526.

53. Wang Z, Pan Q, Gendron P, *et al.*: CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. Cell Rep 2016;15:481–489.

54. Yan M, Wen J, Liang M, Lu Y, Kamata M, Chen IS: Modulation of gene expression by polymer nanocapsule delivery of DNA cassettes encoding small RNAs. PLoS One 2015;10:e0127986.

55. Ye L, Wang J, Beyer AI, *et al.*: Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. Proc Natl Acad Sci U S A 2014;111:9591–9596.

56. Yin C, Zhang T, Li F, *et al.*: Functional screening of guide RNAs targeting the regulatory and structural HIV-1 viral genome for a cure of AIDS. AIDS 2016;30:1163–1174.

57. Yoder KE, Bundschuh R: Host double strand break repair generates HIV-1 strains resistant to CRISPR/Cas9. Sci Rep 2016;6:29530.

58. Zhang Y, Yin C, Zhang T, *et al.*: CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. Sci Rep 2015;5:16277.

59. Zhu W, Lei R, Le Duff Y, *et al.*: The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. Retrovirology 2015;12:22.

60. Dampier W, Nonnemacher MR, Mell J, *et al.*: HIV-1 genetic variation resulting in the development of new quasispecies continues to be encountered in the peripheral blood of well-suppressed patients. PLoS One 2016;11:e0155382.

61. Dampier W, Nonnemacher MR, Sullivan NT, Jacobson JM, Wigdahl B: HIV Excision utilizing CRISPR/Cas9 technology: Attacking the proviral quasispecies in reservoirs to achieve a cure. MOJ Immunol 2014;1:pii:00022.

62. Dampier W, Sullivan NT, Chung CH, Mell JC, Nonnemacher MR, Wigdahl B: Designing broad-spectrum anti-HIV-1 gRNAs to target patient-derived variants. Sci Rep 2017;7:14413.

63. Nonnemacher MR, Pirrone V, Feng R, *et al.*: HIV-1 promoter single nucleotide polymorphisms are associated with clinical disease severity. PLoS One 2016;11:e0150835.

64. Josefsson L, von Stockenstrom S, Faria NR, *et al.*: The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. Proc Natl Acad Sci U S A 2013;110:E4987–E4996.

65. Li L, Aiamkitsumrit B, Pirrone V, *et al.*: Development of co-selected single nucleotide polymorphisms in the viral promoter precedes the onset of human immunodeficiency virus type 1-associated neurocognitive impairment. J Neurovirol 2011;17:92–109.

66. Antell GC, Dampier W, Aiamkitsumrit B, *et al.*: Evidence of divergent amino acid usage in comparative analyses of R5- and X4-associated HIV-1 Vpr sequences. Int J Genomics 2017;2017:4081585.

67. Aiamkitsumrit B, Dampier W, Martin-Garcia J, *et al.*: Defining differential genetic signatures in CXCR4- and the CCR5-utilizing HIV-1 co-linear sequences. PLoS One 2014; 9:e107389.

68. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26:589–595.

69. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987–2993.

70. Hsu PD, Scott DA, Weinstein JA, *et al.*: DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013;31:827–832.

71. Kabadi AM, Ousterout DG, Hilton IB, Gersbach CA: Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res 2014;42:e147.

72. Mansky LM, Temin HM: Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 1995;69:5087–5094.

73. Carvajal-Rodriguez A, Crandall KA, Posada D: Recombination favors the evolution of drug resistance in HIV-1 during antiretroviral therapy. Infect Genet Evol 2007;7:476–483.

74. Doench JG, Fusi N, Sullender M, *et al.*: Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol 2016;34:184–191.

75. Bolukbasi MF, Gupta A, Wolfe SA: Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. Nat Methods 2016;13:41–50.

76. Tycko J, Myer VE, Hsu PD: Methods for Optimizing CRISPR-Cas9 genome editing specificity. Mol Cell 2016; 63:355–370.

Address correspondence to:
Brian Wigdahl
Department of Microbiology and Immunology
Drexel University College of Medicine
245 North 15th Street
Philadelphia, PA 19102

E-mail: bw45@drexel.edu