Roychoudhury *et al. BMC Biology* (2018) 16:75
https://doi.org/10.1186/s12915-018-0544-1

BMC Biology

RESEARCH ARTICLE Open Access

Viral diversity is an obligate consideration in CRISPR/Cas9 designs for targeting the HIV reservoir

Pavitra Roychoudhury[1], Harshana De Silva Feelixge[2], Daniel Reeves[2], Bryan T. Mayer[2], Daniel Stone[2], Joshua T. Schiffer[2,3,4] and Keith R. Jerome[1,2*]

Abstract

Background: RNA-guided CRISPR/Cas9 systems can be designed to mutate or excise the integrated HIV genome from latently infected cells and have therefore been proposed as a curative approach for HIV. However, most studies to date have focused on molecular clones with ideal target site recognition and do not account for target site variability observed within and between patients. For clinical success and broad applicability, guide RNA (gRNA) selection must account for circulating strain diversity and incorporate the within-host diversity of HIV.

Results: We identified a set of gRNAs targeting HIV LTR, *gag*, and *pol* using publicly available sequences for these genes and ranked gRNAs according to global conservation across HIV-1 group M and within subtypes A–C. By considering paired and triplet combinations of gRNAs, we found triplet sets of target sites such that at least one of the gRNAs in the set was present in over 98% of all globally available sequences. We then selected 59 gRNAs from our list of highly conserved LTR target sites and evaluated in vitro activity using a loss-of-function LTR-GFP fusion reporter. We achieved efficient GFP knockdown with multiple gRNAs and found clustering of highly active gRNA target sites near the middle of the LTR. Using published deep-sequence data from HIV-infected patients, we found that globally conserved sites also had greater within-host target conservation. Lastly, we developed a mathematical model based on varying distributions of within-host HIV sequence diversity and enzyme efficacy. We used the model to estimate the number of doses required to deplete the latent reservoir and achieve functional cure thresholds. Our modeling results highlight the importance of within-host target site conservation. While increased doses may overcome low target cleavage efficiency, inadequate targeting of rare strains is predicted to lead to rebound upon cART cessation even with many doses.

Conclusions: Target site selection must account for global and within host viral genetic diversity. Globally conserved target sites are good starting points for design, but multiplexing is essential for depleting quasispecies and preventing viral load rebound upon therapy cessation.

Keywords: CRISPR/Cas9, Gene therapy, Endonucleases, Gene editing, HIV, Latent reservoir, Viral genetic diversity, Computational biology, Mathematical modeling, Genomics

* Correspondence: kjerome@fredhutch.org
[1]Department of Laboratory Medicine, University of Washington, Seattle, USA
[2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, USA
Full list of author information is available at the end of the article

Background

Despite the success of combination antiretroviral therapy (cART) in suppressing HIV viremia, reservoirs of latently infected cells remain the major barrier for HIV cure [1]. The HIV latent reservoir is composed of long-lived infected cells harboring replication-competent proviruses with limited transcription that can reactivate and reseed the reservoir upon cART interruption [2, 3]. A promising therapeutic strategy for achieving cure involves depleting the reservoir by direct disruption of proviral genomes using engineered DNA-editing enzymes such as CRISPR/Cas9 nucleases. A growing body of research shows that endonuclease-induced mutation of essential viral genes or excision of provirus can render the virus unable to replicate [4–12]. If performed on a large scale, this approach could yield pharmacologically significant reservoir reduction. However, viral reservoirs are highly diverse, even in well-suppressed individuals [13, 14], and this diversity remains a major challenge for the application of genome editing strategies towards an HIV cure. Effective targeting of all viral genetic variants within an infected individual will be crucial for achieving sufficient reservoir reduction to prevent viral rebound upon cART cessation [15, 16] and preventing the emergence of resistance to this therapy [11].

Thus far, studies used to demonstrate the viability of gene-editing strategies against HIV have primarily targeted single molecular clones that provide ideal endonuclease target site recognition [7, 8]. Multiple classes of gene-editing enzymes have been studied, but the CRISPR/Cas9 system has gained popularity in recent years due to its effectiveness, relative simplicity, and ease of use. Several computational tools now exist to identify CRISPR target sites, to predict the activity of guide RNAs (gRNAs) targeting those sites, and to identify and score gRNAs based on multiple factors including predicted off-target activity [17–19]. However, no available tools allow guide selection based on predicted target site conservation or predicted clinical efficacy based on viral diversity. The identification and characterization of the most conserved target sites on a group- or subtype-specific basis will allow rapid selection of gRNAs when deep sequencing of a patient's reservoir is not practical or feasible. Furthermore, because the virus can evolve resistance to endonuclease targeting [11], multiple sites may need to be targeted concurrently in order to prevent the emergence of resistance. Therefore, the selection of multiplexed sets of gRNAs must account for the diversity of circulating strains across a wide range of infected people, and dosing strategies must consider within-host diversity of HIV to maximize the probability of a functional cure.

Here, we present a CRISPR gRNA design strategy that selects target sites not only by predicted efficacy and specificity but also by prevalence in the population. We first created a database of highly conserved target sites in HIV LTR, *gag*, and *pol* focusing on group- and subtype-level conservation using information about the global sequence diversity of HIV. We used this database to identify highly conserved target site pairs and triplets to create multiplex gRNA designs predicted to maximize targeting and reduce the probability of treatment resistance. From this analysis, we identified and tested 59 LTR guides using a fluorescent reporter to quantify activity in vitro. We then used deep-sequence data from HIV-infected individuals to determine within-host target site conservation and probability of cleavage by individual gRNAs in our list. Finally, we used a mathematical model to predict the number of doses that would be required to achieve functional cure thresholds, while accounting for varying levels of target site diversity and enzyme efficacy.

Results

Broadly targeting spCas9 gRNAs against HIV gag, pol, and LTR

We performed a screen to identify globally conserved target sites for *Streptococcus pyogenes* (spCas9) in LTR, *gag*, and *pol* using alignments for these regions obtained from the HIV LANL database. LTR was chosen for its utility in excision of the provirus [8, 20, 21], while *gag* and *pol* were chosen based on their conservation between HIV strains [22]. The publicly available LANL alignments contain HIV sequences from thousands of infected persons (from about 1200 for LTR to more than 8000 for *pol*) and include strain and geographic information. From these alignments, we computed majority consensus sequences for LTR, *gag*, and *pol* of HIV-1 group M and subtypes A–C. We identified a total of 246 unique gRNA target sites in LTR, 573 in *gag*, and 897 in *pol*. For each target site identified, we determined the number of exact hits in the overall alignment of all group M sequences and for each subtype and ranked target sites by overall prevalence (Fig. 1). Target sites were found to be most conserved in *pol* (Table 1), where a single target site was present in up to 86.5% ($n = 4416$) of all group M sequences. The most conserved target sites in LTR and *gag* occurred in up to 70.6% ($n = 1216$) and 71.1% ($n = 8435$), respectively, of group M sequences.

We determined predicted on-target cleavage efficiency and off-target activity for each guide sequence (Fig. 2) using the sgRNA designer tool [17]. Predicted on-target activity scores were in the range [0,1] where a score of 1 was associated with successful knockout in the experiments of Doench et al. [17, 23] and gRNAs with scores < 0.2 were generally excluded because they were shown to be predictive of poor activity. Mean predicted activity scores across all identified guides were 0.50 (SD 0.12, $n = 246$) for LTR, 0.49 (SD 0.13, $n = 573$) for *gag*, and 0.47 (SD 0.13, $n = 897$) for *pol*. From the list of gRNAs

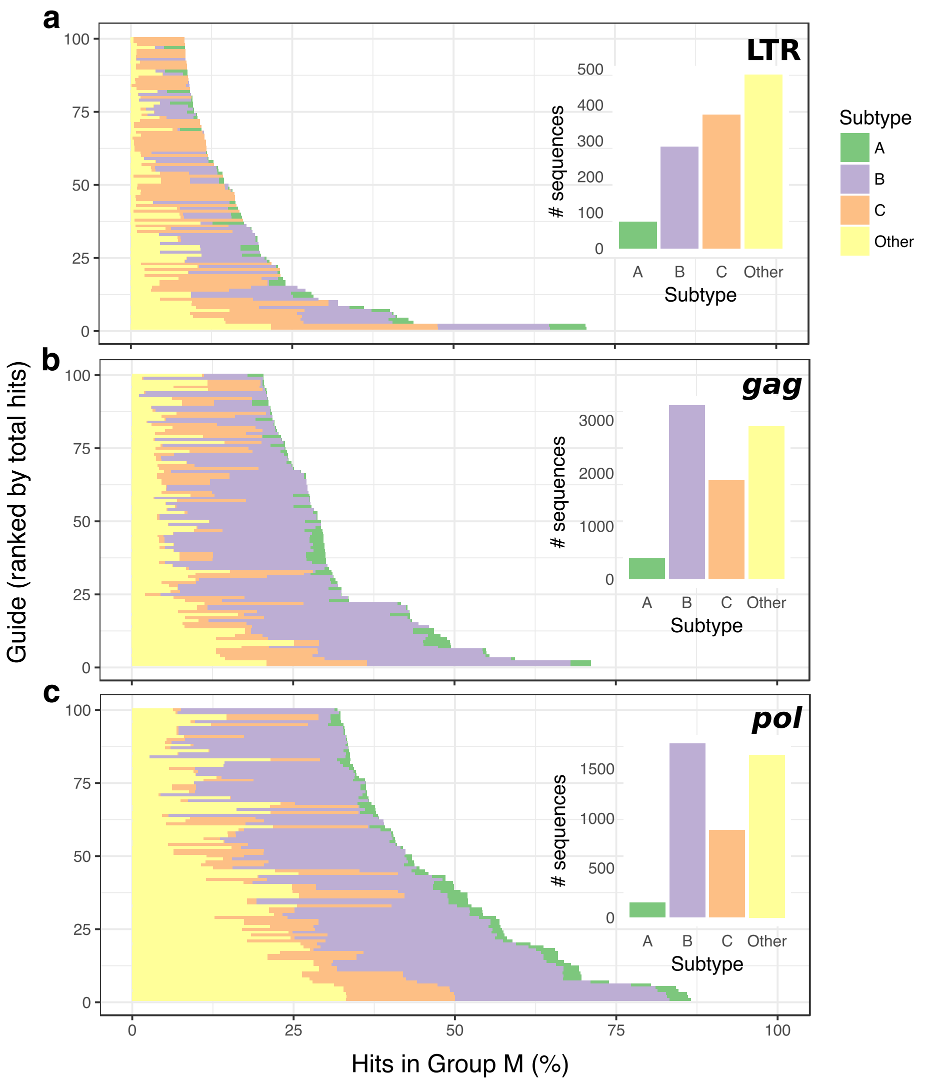

Fig. 1 Top 100 gRNA target sites in HIV LTR (**a**), *gag* (**b**), and *pol* (**c**) ranked by prevalence (bottom to top) within an alignment of available sequences within group M for each genomic region. The *x*-axis shows the percentage of all sequences in group M that contain an exact match to the target site. Within each horizontal bar, shading indicates what percentage of sequences with target sites hits belong to each subtype. Inset bar plots show the total number of sequences of each subtype in the alignment

identified, we excluded 10 from *gag* and 26 from *pol* from further analyses due to high predicted off-target activity scores. No significant correlation was observed between predicted activity and target site conservation (Additional file 1: Table S1A).

Multiplexed gRNA designs

For each gene, we determined the number of sequences that could be targeted by pairs and triplets of gRNAs in group M overall, and in each subtypes A–C (Table 1). We determined that just two strategically selected

Table 1 Maximum targeting possible with 1, 2, or 3 gRNAs

	Subtype A				Subtype B				Subtype C				Group M			
	n	Single	Pair	Triplet	*n*	Single	Pair	Triplet	*n*	Single	Pair	Triplet	*n*	Single	Pair	Triplet
LTR	75	90.7	100.0	100.0	284	74.3	92.6	98.6	373	84.5	96.0	98.9	1216	70.6	83.0	88.8
gag	404	86.4	96.3	99.5	3280	80.9	95.2	98.5	1865	75.7	94.0	98.4	8453	71.1	88.2	95.5
pol	150	96.0	100.0	100.0	1750	88.4	98.6	99.8	878	84.6	97.6	99.9	4416	86.5	96.5	99.2

n = number of sequences in the alignment; the remaining columns show the percentage (out of total sequences) that can be targeted with single, paired, or triplet gRNA combinations

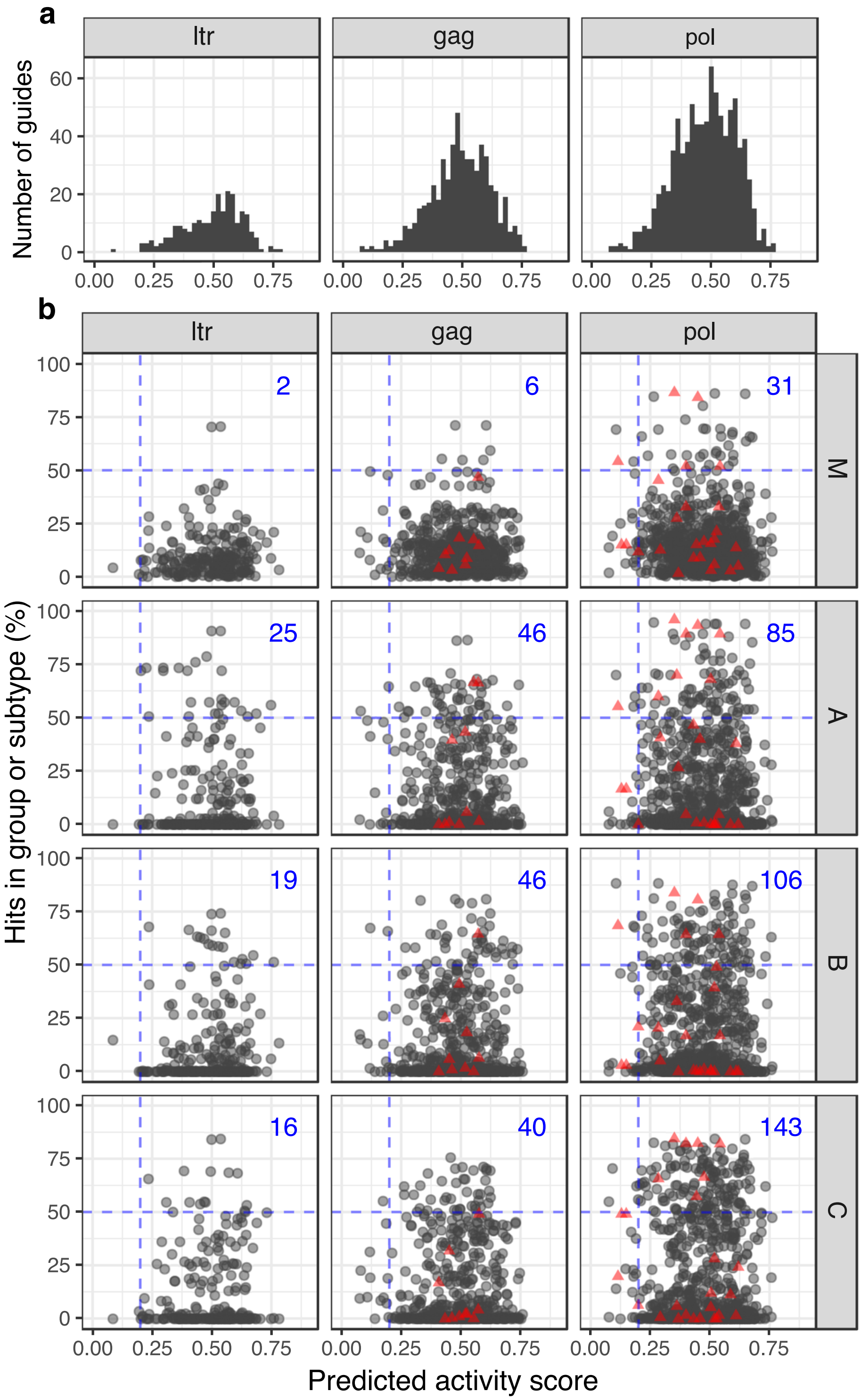

Fig. 2 a Histogram of predicted activity of all gRNAs identified in LTR, *gag*, and *pol* across all four consensus sequences (group M, subtypes A–C) for each gene. **b** Predicted activity score vs. target site conservation for individual gRNAs grouped by subtype and gene. Red triangles indicate gRNAs excluded due to predicted off-target activity. Numbers in blue represent the total number of guides with predicted activity score > 0.2 and where target sites occur in more than 50% of sequences in the group or subtype alignment gRNAs are sufficient for targeting 100% of LTR and *pol* sequences in the current global alignment for subtype A, and three gRNAs are able to target over 98% of all sequences in subtypes A–C. However, when considering all group M sequences, the maximum percentage of sequences targeted by triplet sets of gRNAs drops to 88.8% for LTR, 95.5% for *gag*, and 99.2% for *pol* (Table 1 and Additional file 1: Table S2). The two most conserved LTR sites in the whole of group M (ranks 1 and 2) were also the most prevalent target sites in the individual subtypes, but this was not the case for *gag* and *pol* (Additional file 1: Table S2).

Overall, better coverage of group M or subtypes A–C sequences was achieved when pair or triplet gRNAs targeted *pol*, suggesting that *pol* is an ideal therapeutic target for targeted mutagenesis with multiplexed guide RNAs. We determined that a minimum set of eight gRNA target sites would be required to guarantee that every *pol* sequence in the group M global was targeted at least once.

Functional testing of selected gRNAs

From our list of 246 gRNAs targeting LTR, we identified 59 gRNAs for functional testing by first considering the most conserved target sites in group M and each subtype. We then included any gRNAs that increase the number of sequences targeted when combined in pairs or triplets with the previous list (Additional file 2: Figure S1A). In order to test the activity of these guides in vitro, we designed LTR-GFP fusion reporter constructs using consensus sequences for group M and subtypes A–C (Fig. 3a, Additional file 2: Figure S1B). We tested the ability of each gRNA to knock down reporter GFP expression in HEK293 cells following co-transfection with a plasmid expressing spCas9 mCherry containing each HIV-specific gRNA and the LTR-GFP fusion reporter. The activity of each gRNA was measured in terms of percent knockdown of median GFP fluorescence intensity relative to negative controls at 24 h post-transfection in Cas9 expressing (mCherry positive, Additional file 2: Figure S1C) cells.

We compared measured gRNA activity to predicted activity scores from the sgRNA designer (Fig. 3b); there was a trend towards weak positive correlation between predicted and measured activity (Pearson's $r = 0.25$, $n = 59$, 95% CI = 0.00–0.48, Additional file 1: Table S1B). We observed a reduction of GFP fluorescence intensity with 52 out of 59 gRNAs (Fig. 3c, Additional file 1: Table S4), with a maximum knockdown of 76.3% (mean = 15.3%, SD = 16.0%, $n = 59$). Maximum knockdown was achieved at target site CAAAGACTGCTGACACAGAAGGG, which was identified in the consensus sequence of subtype C and found to occur in 23.1% of group M sequences and 68.4% of subtype C sequences in the 2016 LANL alignment. We observed clustering of the most active guides within the LTR; target sites for gRNAs with GFP knockdown > 30% were found at positions 74–75, 319–344, and 446 relative to the start of the 5′ LTR. Although some active guides appear to coincide with regions of high-residue conservation within the LTR (Fig. 3c), we found no significant correlation between GFP knockdown and target site prevalence within all available sequences in Group M (Pearson's $r = -0.03$, $n = 59$, 95% CI = −0.28–0.23, Additional file 1: Table S1C).

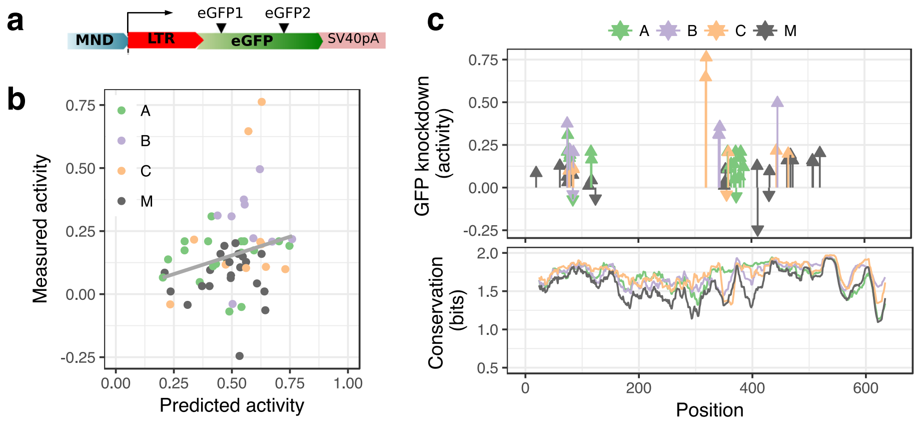

Fig. 3 a LTR-GFP fusion reporter to test gRNAs for activity in vitro. **b** Activity was measured in terms of percent knockdown of median GFP fluorescence intensity relative to negative controls. We found positive but statistically non-significant correlation between computationally predicted activity scores and measured activity. **c** We achieved reduction of GFP fluorescence intensity (positive activity) with a majority of gRNA designs and observed clustering of tested target sites in two areas of the LTR with the most active guides being clustered around the center of the LTR. With a small number of gRNAs, we observed negative activity (increase in GFP fluorescence). Lower panel shows residue conservation (in 0–2 bits) across the LTR for alignments of subtype sequences or all sequences in group M

In silico testing of candidate gRNAs on within-host patient sequences

In order to simulate the application of this gene-editing approach on a diverse within-host virus population, we used a published dataset of HIV sequences obtained from HIV-infected blood donors in Brazil [24], focusing on the *pol* gene (because it is the most highly conserved) for 10 patients. We started with our list of all *pol* target sites that we identified above from group and subtype consensus sequences from 2016 LANL alignments, labeling each target site according to the consensus sequence it was identified from (300, 317, 304, and 328 target sites from group M and subtype A–C consensus sequences, respectively, 1249 sites total, 897 unique sites). From this combined list of globally conserved target sites, we determined whether each site was present in each patient's HIV consensus sequence (Additional file 1: Tables S5 and S6) [24]. Across infected persons, an average of 89.4 group M target sites (i.e., 29.80% of all group M sites identified) and 119.9 subtype B sites (39.44% of all subtype B target sites identified) were found to be also present within patient consensus sequences (SD 11.14 sites/3.24% and 9.84 sites/3.71%, respectively, $n = 10$ patients), while subtype A and C sites were identified less frequently (Fig. 4a). Since subtype B is highly prevalent in Brazil, this was not surprising. Five target sites were found to be present in all 10 patient consensus sequences (Additional file 1: Table S6), and one of these (GATGGCAGGTGATGATTGTGTGG) was also highly conserved in the global alignment for subtype B (present in 87.09% of LANL sequences). These five target sites were found to occur between positions 2294 and 2981 in *pol*. In addition, we identified gRNA target sites directly from the patient's consensus sequence. The number of directly identified sites for each patient ranged between 276 and 313 (mean = 299.30, SD = 10.83, $n = 10$). Out of 1712 unique sites generated from the 10 patients' consensus sequences, 351 were present in our list of globally conserved sites. Of the remaining sites, 1135 were only present in a single individual and 87 sites were found in more than 5 individuals. With one exception (GTTTCTTGCCCTGTCTCTGCTGG), every site that was present in more than 5 individuals was also present in our global list.

Next, we used deep-sequence data from each of these individuals [24] to determine the degree of conservation of each target site within the patient's virus quasispecies population. In order to accurately quantify rare target site variants, we identified 4 out of 10 patient datasets where mean coverage across all identified target sites was above 5000× (Additional file 1: Table S2, Additional file 3: Figure S2B). For each of these patients, we determined within-host target site conservation by computing the percentage of reads in the alignment containing an exact match to the site. Within-host target site conservation was found to vary dramatically for individual gRNAs and between individual patients, ranging between 5.5 and 95.6% with a mean of 83.5% (SD 14.3%, $n = 2298$) (Fig. 4b).

Within-host target site conservation was an average of 3.4% higher for sites identified from our global list (range of means = 84.7–86.5%, $n = 4$ patients) compared to sites that were only present in the patient's sequence (mean = 81.6%, $n = 4$, $p = 0.026$), but the difference between groups was not statistically significant (*F* test, $p = 0.15$). Target sites identified from group M or subtype B consensus sequences tended to be more conserved than sites identified from the patient sequence, but the differences were not statistically significant (both 3.7% higher, with $p = 0.087$ and $p = 0.054$, respectively). Within-host

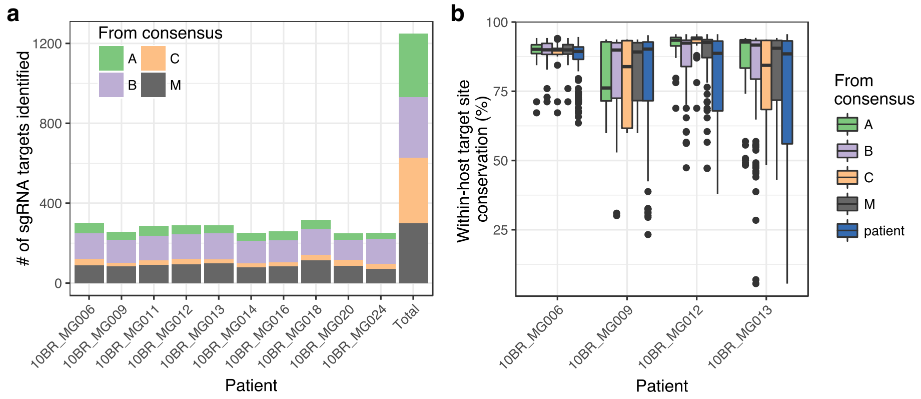

Fig. 4 a Number of previously identified target sites from global consensus sequences of group M and subtypes A–C that were present in each patient's HIV consensus sequence. **b** Within-host target site conservation for each identified target site using deep-sequence data for 4 patients, summarized using box plots. Black dots indicate outlier target sites (outside 1.5 × IQR), and target sites are grouped and colored according to which consensus sequence they were identified from (the group- or subtype-level consensus from LANL alignments, or from the patient's HIV consensus sequence) target site conservation was nearly identical using group M or subtype B sites ($p = 0.98$). All p values were > 0.1 after multiple test corrections.

Modeling reservoir depletion with CRISPR-based therapy

We developed a mathematical model to understand the effect of experimentally controllable parameters on reservoir depletion with hypothetical weekly dosing of various candidate CRISPR/Cas9 therapies targeting HIV. The model simulates the clearance of the latent reservoir by including many (up to 10^4) quasispecies carrying replication-competent DNA. These species are unevenly abundant and are assumed to follow a log-normal distribution so that each quasispecies contains 1–1000 members. Further, each quasispecies is cleared from the reservoir so that the total reservoir clearance rate recapitulates the experimentally measured reservoir half-life of 3–4 years [25, 26]. In the absence of CRISPR therapy, the model simulates a fluctuating but, on average, slowly decaying HIV reservoir with varying compositions [27]. We then simulated reservoir clearance with varying enzyme efficacy (ϵ, the probability of successful mutagenic DNA cleavage at the target site) and varying coverage proportion (ρ, the proportion of sequences that would respond to enzyme). The measure of target site conservation was based on our analysis of patient samples. Parameter ranges for ϵ were based on ranges of predicted cleavage efficiency from the sgRNA designer tool (Fig. 2) and measured activity (Fig. 3) described above.

Including CRISPR, our simulations suggest that treatments with gRNAs targeting a single site will be insufficient to achieve functional cure even at high levels of target site conservation and enzyme efficacy (Fig. 5a, Additional file 4: Figure S3). Enzyme efficacy is relatively unimportant in this case, only affecting the number of treatments needed to remove the sensitive quasispecies. Once removed, additional treatments provide no additional benefit and insensitive quasispecies dominate the reservoir (Fig. 5a). However, if 100% coverage of all quasispecies can be achieved through the selection of a multiplexed set of gRNAs that can be delivered simultaneously, the number of treatments to deplete the reservoir to the first cure threshold (100-fold decrease [16]) can be achieved in 1–5 treatments depending on efficacy (Fig. 5c), whereas the second threshold (2000-fold decrease [15]) may require 5–10 treatments depending on efficacy. For all modeled assumptions, coverage is vital to reservoir depletion. Whereas suboptimal efficiency can be surmounted by repeated doses, the diversity of the reservoir constitutes the largest barrier to depletion.

Discussion

Gene editing using CRISPR/Cas9 has the potential to effect a functional cure for HIV through targeted mutagenesis or proviral genome excision [28]. This approach has now been demonstrated in multiple proof-of-concept in vitro and in vivo studies [7, 9–12, 20, 29, 30]. While laboratory demonstration of gRNA activity has largely relied on clonal populations of lab-adapted HIV strains, clinical applications of this method will need to consider the wide intra- and inter-host diversity of HIV. The global diversity of HIV-1 is reflected in the classification of viruses into four broad groups (M, N, O, and P) that are 25–40%

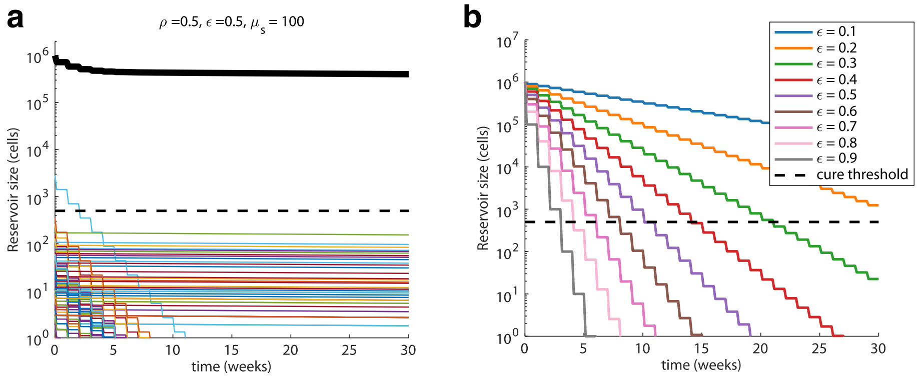

Fig. 5 Simulated reservoir depletion with anti-HIV CRISPR therapy. **a** Example simulation based on predicted target site conservation ("potency," $\rho = 0.5$) and enzyme efficacy to each target site ($\epsilon = 0.5$). CRISPR therapy is dosed weekly, and the average strain contains 100 infected cells ($\mu_s = 100$). Thin colored lines represent single strains, $L_s(t)$, and the thick black line represents the total reservoir, $L(t) = \sum_s L_s(t)$. Strains targeted by CRISPR are cleared rapidly, but untargeted strains remain unaffected and the total reservoir size does not decrease below estimated depletion thresholds for functional cure. The dashed line represents a stringent threshold for latent reservoir reduction where patients are expected to remain suppressed for years without cART [15, 16]. See Additional file 4: Figure S3 for simulations varying all parameters. **b** If 100% coverage ($\rho = 1$) of target sites can be achieved (either through multiplexing of targets or due to a target site that is highly conserved), enzyme efficacy becomes relevant, dictating the number of doses to cure. At or better than predicted efficacy $\epsilon > 0.5$, doses range between 1 and 5 doses for a median 1 year remission and 5–10 doses for a potentially lifelong absence of viral rebound based on previously estimated thresholds. However, even for 100% coverage, efficacy at 10% or less per dose requires substantial dosing (> 30) to achieve thresholds

divergent, and within-group subtypes that are up to 15% divergent [22]. This remarkable global diversity of HIV is the result of within-host evolution and adaption to immune pressure, and transmission of genetic variants from the host quasispecies over multiple rounds of viral replication. Target sites chosen for gene editing will therefore also need to reflect this genetic variability within and between individuals.

Globally conserved target sites are good starting points for gRNA design; if their high frequencies in the population are the result of selection, endonuclease-induced mutations are more likely to be highly deleterious to the virus. Indeed, it has been shown that highly conserved target sites are associated with improved antiviral activity and, importantly, delayed viral escape [10, 29]. Identification of sites that are conserved at a global or subtype level may also allow for future deployment of these therapies in situations where obtaining individual patient HIV sequence data may not be feasible or practical. To this end, we identified gRNA target sites in HIV LTR that were highly conserved in global consensus sequences and tested the activity of these guides in vitro. Using a separate set of deep-sequence data [24], we showed that sites identified from our list of globally conserved targets that were present in the patient's sequence also showed greater within-host conservation. For computational efficiency, our approach looks for exact matches, but future enhancements could incorporate position-dependent penalties to account for the ability of Cas9 to bind in the presence of mismatches to the target site.

The experimental setup used to test candidate gRNAs was designed to allow us to compare gRNAs against each other while minimizing the confounding factors such as cell line-derived variation. We performed the assays under low transfection efficiency conditions and gated on mCherry-positive cells in order to limit plasmid copy numbers that could affect the ability to observe changes in GFP fluorescence intensity by flow cytometry. Since we have previously seen variations in transfection efficiency between different target site reporter plasmids when transfected under the same conditions, we incorporated two internal GFP-specific gRNAs as controls to be analyzed with each reporter. This allowed us to compare the relative activity across all of the LTR-specific gRNAs since they could not all be tested against each of the LTR reporters. We found that within the described transfection efficiency range, we saw comparable levels of relative GFP knockdown when using the two GFP control gRNAs.

Gene therapy approaches designed to cure an infected individual will need to ensure that all relevant within-host variants are targeted. Although early initiation of long-term cART has been shown to reduce the rate of HIV evolution, the virus is still thought to accumulate about 0.97 mutations/kb/year [13, 14]. Using a mathematical model, we showed that variants that are not recognized and cleaved will be the major barrier to achieving functional cure thresholds. These variants, if replication-competent, have the potential to reactivate upon cART interruption and reseed the reservoir. Our model makes assumptions about the underlying distribution of quasispecies abundance, which is not fully understood. Yet, because CRISPR works on a fraction of quasispecies, our conclusions appear robust to simulated reservoirs with different absolute number of species (see Additional file 4: Figure S3). Estimating time to rebound based on reservoir reduction is challenging and various estimates of thresholds for depletion exist [15, 16, 31–33]. In our simulations, we have included estimates for median 1 year and median lifetime remission from HIV rebound [15, 16]. These thresholds were developed from natural reservoirs and might not correspond exactly to the perturbed CRISPR-treated reservoirs. Most importantly, the depletion itself depends on targeting viral quasispecies diversity. While we endeavor to estimate targeting proportions in the present work, further experiments are needed to fully understand the in vivo process.

Besides cleavage efficiency, target site conservation, and reservoir size, a number of other factors will also contribute to the clinical success of this type of gene therapy for HIV cure [28, 34–36]. For example, we have also not explicitly incorporated gene delivery in the current model but instead assumed that it is captured within the cleavage efficiency parameter ϵ. However, we have shown previously [37] that gene delivery of endonucleases using viral vectors is prone to large bottlenecks at the points of vector packaging, viral entry, and gene expression. Optimization of gene delivery is therefore another important step needed for the clinical success of gene therapies against HIV. We and others have shown that multiple doses will be needed to deplete the reservoir to achieve functional cure thresholds [15, 16, 37]. Dosing regimens will need to optimize efficacy while minimizing potential toxicity and off-target effects.

HIV has also been shown to rapidly escape endonuclease targeting in vitro [10, 11, 29]. Although this risk is reduced by keeping the patient on cART, it is still important for endonuclease-based therapies to target multiple sites concurrently in order to achieve sustained reservoir depletion and prevent the emergence of treatment resistance. Our simulations support these findings and show that even enzymes with high on-target efficiency will fail to produce a functional cure if there are target site variants present at frequencies as low as 1%. Two recent proof-of-principle studies showed that an approach with dual gRNAs targeting multiple genes can delay or completely prevent viral escape [12, 38]. We identified paired and triplet sets of gRNA target sites that occur in over 98% of the population. Since these sites are likely to also be highly conserved within-host (as our results suggest), they would be good candidates for testing in vitro for activity. Although our mathematical model can incorporate multiplexed gRNAs by changing the coverage (ρ), it does not explicitly include dynamic emergence of treatment-resistant variants. Our model framework is amenable to emergent resistance but was not included for lack of information on these dynamics. Nor does the model include potential anatomic sanctuary sites where HIV diversity changes in time. The modeled CRISPR therapy assumes constant suppressive cART, and we rely on previous observations that potent cART prevents most ongoing evolution [13, 39–43].

A number of recent studies have designed LTR-based CRISPR strategies and shown broad antiviral activity against HIV in a number of different model systems [7, 8, 12, 20, 21, 38, 44, 45]. LTR is an attractive target because there are two copies per provirus genome, and this allows a single gRNA to potentially cleave two independent regions, leading to a deletion of a majority of the provirus or mutations in one or both LTRs. Each of these potential outcomes is beneficial as they can all impact HIV replication and reactivation. However, we have shown here that *pol* may be a better genomic target for directed mutagenesis due to target site conservation, which allows targeting of a majority of variants with reasonable numbers of gRNAs in multiplexed designs. As a result, we believe that targeting multiple sites within *pol* may be a better approach than targeting LTR alone, which generally contains less conserved sites.

The weak correlation between predicted and measured activity scores is likely due to differences in the methods, cell lines, and experimental conditions used to generate the two sets of scores. The predicted activity score generated by the sgRNA designer tool is based on a broad genome-wide CRISPR-based screen that was used to train a machine learning model [17]. In spite of the differences in approaches, the fact that the scores are correlated is encouraging because it helps to further validate this broadly used metric.

One of the limitations of our within-host analysis is that we do not have detailed information about the patient cohort [24] such as treatment status, age at HIV diagnosis, and time of cART initiation and interruption, if any. These factors could potentially impact reservoir diversity. However, the current analysis is primarily aimed at demonstrating the importance and feasibility of designing gRNAs targeting a diverse viral population. Future work needs to address this in greater detail, possibly incorporating treatment-related variables to select gRNA designs.

Conclusions

In summary, we have performed a detailed computational analysis to identify optimal CRISPR target sites, taking into consideration both within-host and global viral diversity. We determined the in vitro activity of a set of gRNAs targeting highly conserved sites and showed a weak but positive correlation between measured and predicted activity. We used a mathematical model to simulate clinical application of this therapy and showed that although increased dose may overcome low target cleavage efficiency, inadequate targeting of rare strains is predicted to lead to rebound upon cART cessation even with many doses. Our results have applications beyond HIV and CRISPR since genetic diversity is an important consideration for any gene therapy platform targeting a heterogeneous population, whether it is a persistent viral disease such as hepatitis B virus, or even cancer.

Methods

HIV sequence datasets and pre-processing

For our analysis of global target site conservation, we obtained sequences from the Los Alamos National Laboratory (LANL) database. For each region of interest (*gag*, *pol*, LTR), we downloaded pre-made LANL alignments of all available group M sequences (2016 version). We extracted a majority consensus sequence using Geneious v10 [46] for all sequences in group M and for each subtype. We did not consider groups N, O, or P in our analyses because they represent a small fraction of HIV infections globally compared to group M and there are limited sequences available for these groups. However, our algorithms are easily adapted to run on any alignment provided.

For within-host analyses of target site conservation, we used deep-sequencing data (Additional file 1: Table S5) from a study of HIV-infected blood donors in Brazil [24]. Raw paired-end reads for each patient were trimmed to remove adapters and low-quality regions using Trimmomatic v0.32.2 [47] and mapped using Bowtie2 v0.2 [48] to the consensus sequence deposited by the authors to GenBank. These pre-processing steps (Additional file 3: Figure S2) were performed within the Galaxy software framework (https://galaxyproject.org/).

gRNA target site analysis

We developed a custom script to identify gRNA target sites for an input sequence given a specified PAM sequence (default ‘NGG’ for spCas9) and desired gRNA length *w* (default 20 nt). The algorithm finds all matches to the PAM sequence in the forward and reverse directions and returns, for each match, *w* bases upstream of the PAM sequence. We then used the sgRNA designer from the Broad Institute (https://portals.broadinstitute.org/gpp/public/analysis-tools/

sgrna-design) to determine predicted on-target efficacy score and off-target scores (threat matrix) [17]. On-target predicted activity scores are in the range [0,1] with higher values predicting more active guides and a score of 1 indicating successful knockout in the experiments in [17, 23].

For each target site identified, we determined the number of exact matches found in an alignment of the region of interest (LTR, *gag*, or *pol*). We excluded all sites with close off-target matches to the human genome (> 3 matches in Match Bin I, i.e., CFD score = 1 [17]). For each region, we determined pairs and triplets of gRNAs by starting with the previously identified list of gRNAs and adding on guides that increase targeting when used in combination.

We computed target site conservation in terms of the frequency of occurrence of the target site (exact matches) within the alignment and also we used a measure of information content similar to what is used to generate sequence logo plots [49, 50]. We applied a moving window of size 23 (corresponding to the width of gRNA) and computed conservation from the relative frequencies of bases in the alignment using the method of Schneider et al. [50] incorporating small-sample correction. The result is a value between 0 and 2 bits with higher values indicating greater sequence conservation. All analyses were performed in R/Bioconductor, and code is available on GitHub (http://github.com/proychou/CRISPR).

Functional testing of gRNA activity

Starting with the list of target sites identified above in LTR, we selected gRNAs from a pool of the top 20 most conserved sites across group M overall, the top 10 most conserved sites in each subtype, and the top 20 pairs and triplets. As recommended by sgRNA designer, we excluded any gRNAs with on-target activity scores < 0.2.

We developed 4 LTR-GFP fusion reporter constructs using consensus sequences for all group M, subtype A, subtype B, and subtype C (further details in Additional file 5). Internal start codons and stop codons were identified within the sequence for each consensus LTR, and the reading frame with the fewest combined number of start codons and stop codons was identified. Reading frame 1 for group M contained 5 start and 4 stop codons, reading frame 1 for subtype A contained 3 start and 6 stop codons, reading frame 1 for subtype B contained 3 start and 6 stop codons, and reading frame 1 for subtype C contained 3 start and 5 stop codons. All the internal start and stop codons were modified for each consensus LTR sequence as follows: ATG to GTG - M to V; TGA to GGA - stop to G; TAG to GAG - stop to E; TAA to GAA - stop to E, so that one continuous open reading frame was generated. Each of the 4 modified consensus LTR sequences was then synthesized as a gBlock and cloned into a reporter plasmid vector (cloning details available upon request) as a fusion to the 5′ end of the eGFP ORF so that the MND promoter drove expression of a single continuous ORF (see Additional file 2: Figure S1A for amino acid sequences). The majority of the 59 gRNA target sites identified for analysis within the group M, subtype A, subtype B, and subtype C consensus LTRs were not changed by start or stop codon modification, with the exception of overlapping gRNA targets 1 and 2, and overlapping gRNA targets 18 and 19. A separate reporter construct was generated for gRNAs 1, 2, 18, and 19 by fusing their target sequences to the 5′ end of the eGFP ORF so that the MND promoter also drove expression of a single continuous ORF (cloning details available upon request).

Of the 59 LTR-specific gRNA target sites we elected to screen for activity, 23 were present in the group M reporter, 27 were present in the group A reporter, 20 were present in the group B reporter, 18 were present in the group C reporter, and gRNAs 1, 2, 18, and 19 were not present in any LTR reporter. Three of the gRNA targets were present in all 4 LTR-reporter constructs, 8 were present in 3 LTR-reporter constructs, and 8 were present in 2 LTR-reporter constructs. To screen the activity of individual LTR-specific gRNAs, they were cloned into the BbsI site of the plasmid pU6-(Bbs1) CBh-Cas9-T2A-mCherry (a gift from Ralf Kuehn; Addgene plasmid no. 64324) under the control of the U6 promoter. This plasmid expresses spCas9 and mCherry from the constitutive CBh promoter. Internal positive controls for GFP knockdown were used by also cloning gRNAs eGFP1 and eGFP2 targeting the sequences CAACTACAAGACCCGCGCCG and GTGAACCGCATCGAGCTGAA into pU6-(Bbs1) CBh-Cas9-T2A-mCherry. To assay gRNA activity 2×10^5, 293 cells were plated in 12-well plates and the following day individual wells were transfected by PEI transfection with 1000 ng of a Cas9/LTR-gRNA expressing plasmid and 250 ng of its corresponding LTR-reporter plasmid. At 24 h post-transfection, flow cytometry was performed and GFP fluorescence was analyzed in Cas9 expressing (mCherry positive) 293 cells to determine the level of GFP knockdown provided by each gRNA.

Analysis of flow cytometry data

Raw fcs files were gated using functions from the OpenCyto framework in R/Bioconductor [51] as described previously [37]. Flow data has been uploaded to FlowRepository (https://flowrepository.org/id/FR-FCM-ZYHR), and code is available at http://github.com/proychou/CRISPR.

Intra-host target site conservation

Focusing on the *pol* gene, we identified spCas9 gRNA target sites within the HIV consensus sequence for each patient using the script described above, excluding any sites containing degenerate bases. We also determined which of the target sites we had previously identified from group- and subtype-level consensus sequences for *pol* were present in the patient consensus sequence. Using the average number of reads overlapping all identified target sites, we excluded any patients with < 5000× target site depth since we were interested in variants that may escape targeting by candidate gRNAs. For each target site, we determined the number of reads in the alignment containing an exact match to the target site and excluded any sites where coverage was less than 5000×. We then used the total number of reads that completely overlap the target site to calculate the percentage of exact target site matches.

Statistical analysis of within-host conservation

To test whether there were differences in target site conservation measured by mean percentages of exact target site matches per total reads, a linear mixed model was fit with percentage as the outcome and the consensus sequence group (group M, subtypes A–C, and patient) as the predictors. A random intercept for each subject by consensus group was used to account for within subject and group variation across the repeated outcomes. An overall test was performed from ANOVA for mixed models using the lmerTest package in R [52]. Post-hoc pairwise tests were also performed comparing the patient-derived sequences, group M, and subtype B (the circulating strain in the patient population). To compare the conservation using patient target sites to the consensus groups, we pooled group M and subtypes A–C into a single group for comparison in the model, while the random effects specification remained the same. *P* values corrected for multiple testing were also reported using the Holm method [53]. Code and data are available at http://github.com/proychou/CRISPR.

Mathematical model of reservoir depletion with simultaneous suppressive cART and CRISPR therapy

We have used a mathematical model to describe natural clearance of the HIV reservoir on consistent cART previously [27]. That model assumed an HIV reservoir that exponentially cleared with previously measured rates. Here, we extended that model to consider simultaneous treatment with suppressive cART and CRISPR gene therapy. The reservoir is now conceived of as a population of different strains, and each strain is associated with some number of infected cells. cART is assumed to prevent ongoing replication, viral evolution, and/or increases of diversity. Additional CRISPR therapy targets some fraction of these strains, and depending on the coverage, or "proportion" (ρ), and the enzyme activity to those covered strains, or "efficacy" (ϵ), the reservoir is reduced accordingly with each successive CRISPR dose. Throughout the simulations, we use weekly doses τ = 7 days, but this choice is arbitrary and adjustable.

The natural clearance of the reservoir on suppressive cART was modeled as follows. For each strain, a clearance rate was randomly sampled so that the clearance of the entire reservoir agrees with previously measured population level statistics [25, 26] such that the half-life of latently infected cells is normally distributed with mean and standard deviation of 3.6 and 1.5 years, respectively, or $t_{\{1/2\}} \sim \mathcal{N}(3.6, 1.5)$. Of note, this half-life represents the natural clearance rate of the replication-competent reservoir as measured by viral outgrowth assays [25, 26]. In contrast, the half-life of HIV DNA is longer [54, 55]. We call the strain-specific clearance rate θ_s (per day). Each strain (indexed by s) is initialized with a number of infected cells $L_s(0)$ drawn from a log-normal distribution with average value μ_s and standard deviation $\sigma_s = \mu_s$ so that each strain has size $\log L_s(0) \sim \mathcal{N}(\mu_s, \sigma_s)$. Then, we denote the total number of strains $\mathcal{S}$ and the total initial reservoir size $L(0)$ that $\sum_{s=1}^{\mathcal{S}} L_s(0) = L(0)$. The total number of strains is constrained by the initial reservoir size as $\mathcal{S} \approx L(0)/\mu_s$.

We can write model for a single strain without CRISPR therapy using an ordinary differential equation (ODE) model as $\dot{L_s} = -\theta_s L_s$, where the over-dot denotes derivative in time. Such an equation is solved simply, $L_s(t) = L_s(0)\exp(-\theta_s t)$, and applies for strains not in the covered CRISPR set, $(s \notin \mathbb{C})$, where $\mathbb{C} = \{1, 2, 3, \ldots |\rho\mathcal{S}|\}$ and $|\cdot|$ denotes rounding to the nearest integer. For strains in the CRISPR set, the dynamics are governed by the additional reduction in reservoir due to CRISPR, $\eta(t, \tau)$, such that the CRISPR instantaneously removes a fraction of the reservoir $\epsilon L_s(t)$ after each dosing time τ. We solve these equations accordingly for strains in and not in the covered set and sum to find the total reservoir size $L(t) = \sum_s L_s(t)$. Stochastic simulations and deterministic simulations result in similar results (data not shown). All code is freely available at http://github.com/proychou/CRISPR.

Parameters relating to CRISPR (ϵ, ρ, μ_s) are varied throughout simulations. The reservoir initial size was held constant throughout simulations at ~ 1 million cells [25, 26, 56]. The clearance rate of each strain was sampled from a normal distribution with mean half-life 3.6 years and standard deviation 1.5 years as has been measured previously [26]. In the stochastic simulation, strains do sometimes increase over time on cART, a realistic phenomenon. However, simulations were also performed with clearance rates of zero to similar results. Indeed, based on the timeframe of the present analyses (less than a year of cART), natural clearance has a minimal impact compared to CRISPR intervention.

Additional files

Additional file 1: Table S1. (A) Correlation between predicted activity and target site conservation. (B) Correlation between measured and predicted activity. (C) Correlation between measured activity and target site prevalence. **Table S2.** List of highly conserved, subtype-specific triplet/paired gRNAs. **Table S3.** Analysis of the number of guides needed to target all available LANL sequences for LTR, gag, and pol for group M and subtypes A–C. **Table S4.** GFP knockdown with candidate guides tested using fluorescent reporter. **Table S5.** Sequences used in intra-host analysis. **Table S6.** Guides from globally conserved list (using LANL sequences) that have matches in patient sequence. (XLSX 59 kb)

Additional file 2: Figure S1. (A) gRNAs were selected for functional testing based on the number of sequences targeted in a global group- or subtype-level alignment either singly, in pairs or triplets (B) amino acid sequence for the N-terminus of each LTR-reporter GFP fusion construct. M group, subtype A, subtype B, and subtype C reporter amino acid sequences are aligned for each of the 4 reporter constructs. The sequence for eGFP begins with the sequence VSKGEELFT. (C) Transfection efficiency shown in terms of percentage of mCherry+ cells in each treatment. (D) Absolute numbers of mCherry+GFP+ cells in each treatment. (EPS 498 kb)

Additional file 3: Figure S2. (A) Flowchart showing processing steps for intra-host deep-sequence data. (B) Target site depth based on number of reads overlapping the target site in an alignment for 4 patients with deep-sequence data. Black dots indicate outlier target sites (outside 1.5 × IQR), and target sites are grouped and colored according to which consensus sequence they were identified from (the group- or subtype-level consensus from LANL alignments, or from the patient's HIV consensus sequence). (EPS 246 kb)

Additional file 4: Figure S3. (A) Three hypothetical distributions of quasispecies abundance in the HIV reservoir. In each case, the total size of the reservoir (number of infected cells) is the same ($L = 10^6$), but the average number of cells in a quasispecies, or "log10 clone size," is $\mu = 10^2$, 10^3, 10^4, respectively. Quasispecies abundances are drawn from a log-normal distribution with variance $\sigma_s = \mu_s$ in each case. The distributions match simulations in (B) by color. (B) Simulations of total reservoir clearance assuming suppressive cART and hypothetical CRISPR treatment of efficacy ϵ and coverage proportion ρ. Each colored line matches the respective distribution in (A). Simulations with smaller average clone sizes gave similar results. The dashed line represents a conservative HIV cure threshold (2000-fold decrease) taken from the literature. Coverage proportion is much more important that efficacy in reducing reservoir size—compare top right panels (low proportion covered, high efficacy) to bottom left panels (high proportion, low efficacy). Low efficacy can additionally be surmounted by more dosing, but HIV's large diversity remains the largest barrier to cure with this intervention. (EPS 561 kb)

Additional file 5: Supplementary methods: c reporter design. (DOCX 1641 kb)

Funding

This work was funded in part by an NIH/NIAID grant UM1 AI126623 (K. Jerome; HP Kiem, co-PIs) and NIH/NIAID University of Washington Center for AIDS Research grant P30 AI 027757-28 (K. Holmes, PI/K. Jerome Director, Co-investigator).

Availability of data and materials

The code used for analysis and visualization, along with supporting data, are available on Github at http://github.com/proychou/CRISPR and FlowRepository at https://flowrepository.org/id/FR-FCM-ZYHR. Additional data are presented in Supplementary Tables, and external data sources have been cited within the text.

Authors' contributions

PR, HDSF, DS, and KRJ conceptualized the project. PR, DR, BTM, and HDSF performed the data analysis. DR and JTS designed the mathematical model. HDSF and DS designed and performed the experiments. PR and HDSF drafted the manuscript with contributions from all other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Department of Laboratory Medicine, University of Washington, Seattle, USA. [2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, USA. [3]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, USA. [4]Department of Medicine, University of Washington, Seattle, USA.

Received: 26 May 2018 Accepted: 21 June 2018
Published online: 11 July 2018

References

1. Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, Pomerantz RJ. The challenge of finding a cure for HIV infection. Science (80-). 2009; 323:1304–7. https://doi.org/10.1126/science.1165706.
2. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med. 2009;15:893–900. https://doi.org/10.1038/nm.1972.
3. Soriano-Sarabia N, Archin NM, Bateson R, Dahl NP, Crooks AM, Kuruc JAD, et al. Peripheral Vγ9Vδ2 T cells are a novel reservoir of latent HIV infection. PLoS Pathog. 2015;11 https://doi.org/10.1371/journal.ppat.1005201.
4. Sarkar I, Hauber I, Hauber J, Buchholz F. HIV-1 proviral DNA excision using an evolved recombinase. Science (80-). 2007;316:1912–5. https://doi.org/10.1126/science.1141453.
5. Mariyanna L, Priyadarshini P, Hofmann-Sieber H, Krepstakies M, Walz N, Grundhoff A, et al. Excision of HIV-1 proviral DNA by recombinant cell permeable tre-recombinase. PLoS One. 2012;7 https://doi.org/10.1371/journal.pone.0031576.
6. Qu X, Wang P, Ding D, Li L, Wang H, Ma L, et al. Zinc-finger-nucleases mediate specific and efficient excision of HIV-1 proviral DNA from infected and latently infected human T cells. Nucleic Acids Res. 2013;41:7771–82. https://doi.org/10.1093/nar/gkt571.
7. Ebina H, Misawa N, Kanemura Y, Koyanagi Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. Sci Rep. 2013;3:2510. https://doi.org/10.1038/srep02510.
8. Hu W, Kaminski R, Yang F, Zhang Y, Cosentino L, Li F, et al. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc Natl Acad Sci U S A. 2014;111:11461–6. https://doi.org/10.1073/pnas.1405186111.
9. Zhu W, Lei R, Le Duff Y, Li J, Guo F, Wainberg MA, et al. The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. Retrovirology. 2015;12:22. https://doi.org/10.1186/s12977-015-0150-z.
10. Wang Z, Pan Q, Gendron P, Zhu W, Guo F, Cen S, et al. CRISPR/Cas9-derived mutations both inhibit HIV-1 replication and accelerate viral escape. Cell Rep. 2016;15:481–9. https://doi.org/10.1016/j.celrep.2016.03.042.
11. De Silva Feelixge HS, Stone D, Pietz HL, Roychoudhury P, Greninger AL, Schiffer JT, et al. Detection of treatment-resistant infectious HIV after genome-directed antiviral endonuclease therapy. Antivir Res. 2016;126:90–8. https://doi.org/10.1016/j.antiviral.2015.12.007.
12. 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 ElsevierCompany. 2016;17:2819–26. https://doi.org/10.1016/j.celrep.2016.11.057.
13. Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, 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. 2013;110:E4987–96. https://doi.org/10.1073/pnas.1308313110.

14. Dampier W, Nonnemacher MR, Mell J, Earl J, Ehrlich GD, Pirrone V, 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 https://doi.org/10.1371/journal.pone.0155382.
15. Hill AL, Rosenbloom DI, Fu F, Nowak MA, Siliciano RF. Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. Proc Natl Acad Sci U S A. 2014;111:13475–80. https://doi.org/10.1073/pnas.1406663111.
16. Pinkevych M, Cromer D, Tolstrup M, Grimm AJ, Cooper DA, Lewin SR, et al. HIV reactivation from latency after treatment interruption occurs on average every 5-8 days—implications for HIV remission. PLoS Pathog. 2015;11: e1005000. https://doi.org/10.1371/journal.ppat.1005000.
17. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34:184–91. https://doi.org/10.1038/nbt.3437. Nature Publishing Group
18. Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PLoS One. 2014;9:e100448. https://doi.org/10.1371/journal.pone.0100448. Khodursky AB, editor
19. Zhu LJ. Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. Front Biol (Beijing). 2015;10:289–96. https://doi.org/10.1007/s11515-015-1366-y.
20. Kaminski R, Bella R, Yin C, Otte J, Ferrante P, Gendelman HE, et al. Excision of HIV-1 DNA by gene editing: a proof-of-concept in vivo study. Gene Ther. 2016:1–6. https://doi.org/10.1038/gt.2016.41.
21. Yin C, Zhang T, Li F, Yang F, Putatunda R, Young W-B, et al. Functional screening of guide RNAs targeting the regulatory and structural HIV-1 viral genome for a cure of AIDS. AIDS. 2016;30:1163–74. https://doi.org/10.1097/QAD.0000000000001079.
22. Li G, Piampongsant S, Faria NR, Voet A, Pineda-Peña A-C, Khouri R, et al. An integrated map of HIV genome-wide variation from a population perspective. Retrovirology. 2015;12:18. https://doi.org/10.1186/s12977-015-0148-6.
23. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat Biotechnol. 2014;32:1262–7. https://doi.org/10.1038/nbt.3026. Nature Publishing Group
24. Pessôa R, Loureiro P, Esther Lopes M, Carneiro-Proietti ABF, Sabino EC, Busch MP, et al. Ultra-deep sequencing of HIV-1 near full-length and partial proviral genomes reveals high genetic diversity among Brazilian blood donors. PLoS One. 2016;11:e0152499. https://doi.org/10.1371/journal.pone.0152499. Kaderali L, editor
25. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat Med. 2003;9:727–8. https://doi.org/10.1038/nm880.
26. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JAD, et al. Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. J Infect Dis. 2015;212:1361–5. https://doi.org/10.1093/infdis/jiv218.
27. Reeves DB, Duke ER, Hughes SM, Prlic M, Hladik F, Schiffer JT. Anti-proliferative therapy for HIV cure: a compound interest approach. Sci Rep. 2017;7:4011. https://doi.org/10.1038/s41598-017-04160-3.
28. Spragg C, De Silva Feelixge H, Jerome KR. Cell and gene therapy strategies to eradicate HIV reservoirs. Curr Opin HIV AIDS. 2016;11:442–9. https://doi.org/10.1097/COH.0000000000000284.
29. Wang G, Zhao N, Berkhout B, Das AT. CRISPR-Cas9 can inhibit HIV-1 replication but NHEJ repair facilitates virus escape. Mol Ther. 2016;24:522–6. https://doi.org/10.1038/mt.2016.24.
30. Kaminski R, Chen Y, Fischer T, Tedaldi E, Napoli A, Zhang Y, et al. Elimination of HIV-1 genomes from human T-lymphoid cells by CRISPR/Cas9 gene editing. Sci Rep. 2016; https://doi.org/10.1038/srep22555.
31. Pinkevych M, Kent SJ, Tolstrup M, Lewin SR, Cooper DA, Søgaard OS, et al. Modeling of experimental data supports HIV reactivation from latency after treatment interruption on average once every 5–8 days. PLOS Pathog. 2016;12: e1005740. https://doi.org/10.1371/journal.ppat.1005740. Swanstrom R, editor
32. Hill AL, Rosenbloom DIS, Siliciano JD, Siliciano RF. Insufficient evidence for rare activation of latent HIV in the absence of reservoir-reducing interventions. PLOS Pathog. 2016;12:e1005679. https://doi.org/10.1371/journal.ppat.1005679. Swanstrom R, editor
33. Hernandez-Vargas EA. Modeling kick-kill strategies toward HIV cure. Front Immunol. 2017; https://doi.org/10.3389/fimmu.2017.00995.
34. Jerome KR. Disruption or excision of provirus as an approach to HIV cure. AIDS Patient Care STDs. 2016;30:551–5. https://doi.org/10.1089/apc.2016.0232.
35. Schiffer JT, Aubert M, Weber ND, Mintzer E, Stone D, Jerome KR. Targeted DNA mutagenesis for the cure of chronic viral infections. J Virol. 2012;86: 8920–36. https://doi.org/10.1128/JVI.00052-12.
36. Stone D, Kiem HP, Jerome KR. Targeted gene disruption to cure HIV. Curr Opin HIV AIDS. 2013;8:217–23. https://doi.org/10.1097/COH.0b013e32835f736c.
37. Roychoudhury P, De Silva Feelixge HS, Pietz HL, Stone D, Jerome KR, Schiffer JT. Pharmacodynamics of anti-HIV gene therapy using viral vectors and targeted endonucleases. J Antimicrob Chemother. 2016:dkw104. https://doi.org/10.1093/jac/dkw104.
38. Lebbink RJ, De Jong DCM, Wolters F, Kruse EM, Van Ham PM, Wiertz EJHJ, et al. A combinational CRISPR/Cas9 gene-editing approach can halt HIV replication and prevent viral escape. Sci Rep. 2017;7:1–10. https://doi.org/10.1038/srep41968. Nature Publishing Group
39. Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, et al. Establishment and stability of the latent HIV-1 DNA reservoir. elife. 2016;5 https://doi.org/10.7554/eLife.18889.
40. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, et al. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. PLoS Pathog. 2014;10 https://doi.org/10.1371/journal.ppat.1004010.
41. Kearney MF, Wiegand A, Shao W, McManus WR, Bale MJ, Luke B, et al. Ongoing HIV replication during ART reconsidered. Open Forum Infect Dis. 2017;4 https://doi.org/10.1093/ofid/ofx173.
42. Rosenbloom DIS, Hill AL, Rabi SA, Siliciano RF, Nowak MA. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. Nat Med. 2012;18:1378–85. https://doi.org/10.1038/nm.2892.
43. Lorenzo-Redondo R, Fryer HR, Bedford T, Kim EY, Archer J, Pond SLK, et al. Lorenzo-Redondo et al. reply. Nature. 2017;551:E10. https://doi.org/10.1038/nature24635.
44. Yin L, Hu S, Mei S, Sun H, Xu F, Li J, et al. CRISPR/Cas9 inhibits multiple steps of HIV-1 infection. Hum Gene Ther. 2018; https://doi.org/10.1089/hum.2018.018.
45. Yin C, Zhang T, Qu X, Zhang Y, Putatunda R, Xiao X, et al. In vivo excision of HIV-1 provirus by saCas9 and multiplex single-guide RNAs in animal models. Mol Ther. 2017;25:1168–86. https://doi.org/10.1016/j.ymthe.2017.03.012.
46. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–9. https://doi.org/10.1093/bioinformatics/bts199.
47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
49. Crooks GE. WebLogo: a sequence logo generator. Genome Res. 2004;14: 1188–90. https://doi.org/10.1101/gr.849004.
50. Schneider TD, Stormo GD, Gold L, Ehrenfeucht A. Information content of binding sites on nucleotide sequences. J Mol Biol. 1986;188:415–31. https://doi.org/10.1016/0022-2836(86)90165-8.
51. Finak G, Frelinger J, Jiang W, Newell EW, Ramey J, Davis MM, et al. OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. PLoS Comput Biol. 2014;10:e1003806. https://doi.org/10.1371/journal.pcbi.1003806.
52. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw. 2017;82 https://doi.org/10.18637/jss.v082.i13.
53. Holm SA. Simple sequentially Rejective multiple test procedure. Scand J Stat. 1979;6:65–70. https://doi.org/10.2307/4615733.
54. Jaafoura S, De Goër De Herve MG, Hernandez-Vargas EA, Hendel-Chavez H, Abdoh M, Mateo MC, et al. Progressive contraction of the latent HIV reservoir around a core of less-differentiated CD4+memory T cells. Nat Commun 2014;5. https://doi.org/10.1038/ncomms6407.
55. Besson GJ, Lalama CM, Bosch RJ, Gandhi RT, Bedison MA, Aga E, et al. HIV-1 DNA decay dynamics in blood during more than a decade of suppressive antiretroviral therapy. Clin Infect Dis. 2014;59:1312–21. https://doi.org/10.1093/cid/ciu585.
56. Ho Y-C, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DIS, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155:540–51. https://doi.org/10.1016/j.cell.2013.09.020. Elsevier Inc