

Design of an optimal combination therapy with broadly neutralizing antibodies to suppress HIV-1

Colin LaMont¹, Jakub Otwinowski^{1†}, Kanika Vanshylla², Henning Gruell², Florian Klein², Armita Nourmohammad^{1,3,4*}

¹Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany; ²Laboratory of Experimental Immunology, Institute of Virology Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany; ³Department of Physics, University of Washington, Seattle, United States; ⁴Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract Infusion of broadly neutralizing antibodies (bNAbs) has shown promise as an alternative to anti-retroviral therapy against HIV. A key challenge is to suppress viral escape, which is more effectively achieved with a combination of bNAbs. Here, we propose a computational approach to predict the efficacy of a bNAb therapy based on the population genetics of HIV escape, which we parametrize using high-throughput HIV sequence data from bNAb-naïve patients. By quantifying the mutational target size and the fitness cost of HIV-1 escape from bNAbs, we predict the distribution of rebound times in three clinical trials. We show that a cocktail of three bNAbs is necessary to effectively suppress viral escape, and predict the optimal composition of such bNAb cocktail. Our results offer a rational therapy design for HIV, and show how genetic data can be used to predict treatment outcomes and design new approaches to pathogenic control.

*For correspondence:
armita@uw.edu

Present address: [†]Dyno Therapeutics, Cambridge, United States

Competing interest: See page 30

Funding: See page 30

Preprinted: 05 December 2021

Received: 01 December 2021

Accepted: 04 July 2022

Published: 19 July 2022

Reviewing Editor: Kayla Sprenger, Massachusetts Institute of Technology, United States

© Copyright LaMont et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

This paper will be of interest to scientists within the fields of statistical and biological physics, immunology, and vaccinology. The mathematical/statistical framework is rigorously constructed based on key concepts from population genetics and high-throughput viral genetic sequence data. The results provide important insights into the failures of past treatment regimens with broadly neutralizing antibodies to suppress viral escape in clinical trial participants. The results also present exciting and highly testable predictions of improved treatment strategies for combatting HIV through passive bNAb immunization.

Introduction

Recent discoveries of highly potent broadly neutralizing antibodies (bNAbs) provide new opportunities to successfully prevent, treat, and potentially cure infections from evolving viruses such as HIV-1 (Walker et al., 2011; Walker et al., 2009; Liao et al., 2013; Mouquet and Nussenzweig, 2013; Klein et al., 2013; Kwong et al., 2013; Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018; Sok and Burton, 2018; Zwick et al., 2001; Burton et al., 2012, influenza Sparrow et al., 2016, and the Dengue virus Ekiert and Wilson, 2012; Durham et al., 2019). bNAbs target vulnerable regions of a virus, such as the CD4 binding site of HIV env protein, where escape mutations can be costly for the virus (Walker et al., 2009; Chen et al., 2009; Zhou et al., 2010; Walker et al., 2011; Liao et al., 2013; West et al., 2014; Burton and Hangartner, 2016). As a result, eliciting bNAbs is the goal of

a universal vaccine against the otherwise rapidly evolving HIV-1. Apart from vaccination, bNAbs can also offer significant advances in therapy against both HIV-1 and influenza (*West et al., 2014; Caskey et al., 2016; Gruell and Klein, 2018; Durham et al., 2019*). Specifically, augmenting current anti-retroviral therapy (ART) drugs with bNAbs may provide the next generation of HIV therapies (*Horwitz et al., 2013; Gruell and Klein, 2018*).

Recent studies have used bNAb therapies to curb infections by the Simian immunodeficiency virus (SHIV) in non-human primates (*Shingai et al., 2013; Barouch et al., 2013; Julg et al., 2017*), and HIV-1 infections in human clinical trials (*Caskey et al., 2015; Bar et al., 2016; Caskey et al., 2017; Baron et al., 2018*). Monotherapy trials with potent bNAbs, including 3BNC117 (*Caskey et al., 2015*), VRC01 (*Bar et al., 2016*), and 10–1074 (*Caskey et al., 2017*) indicate that administering bNAbs is safe and can suppress viral load in patients. Nonetheless, in each trial, escape mutants emerge resulting in a viral rebound after about 20 days past infusion of the bNAb. However, in trials that administered a combination of 10–1074 and 3BNC117, viral rebound was substantially suppressed (*Shingai et al., 2013; Baron et al., 2018*). The success of combination therapy is not surprising. For example, combinations of drugs has been repeatedly used against infectious agents, including current HIV ART cocktails and combination antibiotic treatments against Tuberculosis (*Lienhardt et al., 2012*).

The principle behind combination therapy with either drugs or antibodies is clear: It is harder for a pathogen population to acquire resistance against multiple treatment targets simultaneously than to acquiring resistance against each target separately. But deciding on combination therapy means navigating an enormous number of possible treatment options.

Experimental data from neutralization assays against pseudo-viruses together with modeling and machine learning techniques have been used to statistically characterize the efficacy of bNAbs and their combinations against different variants of HIV (*Wagh et al., 2016; Yu et al., 2019*). Using these neutralization models, an optimal combination therapy was proposed based on their breadth, potency of neutralization, and other relevant measures (*Wagh et al., 2016; Yu et al., 2019*). In another study, pharmacokinetic dynamics was coupled with drug-interaction models to determine an optimal dosing strategies (*Mayer et al., 2022*). These modeling approaches shed light on how combinations of bNAbs that can neutralize a panel of viruses. However, the key obstacle in bNAb therapy is the possibility of viral escape.

To characterize viral escape, mechanistic models, partly inspired by previous work on HIV escape from the anti-retro viral therapy (ART), have been developed to explain the dynamics of viremia in patients, following passive infusion of bNAbs. By making a fit to the trial data, these models are used to infer parameters related to the efficacy of a bNAb in clearing virions, reducing viral load and infectivity, and also to infer the characteristics of the HIV and the T-cell populations, such as the initial viral population size, the death rates of uninfected and infected T-cells, and the number of virions released by an infected T-cell (*Perelson et al., 1996; Ribeiro and Bonhoeffer, 2000; Rong et al., 2010; Rong et al., 2007; Tomaras et al., 2008; Lu et al., 2016; Reeves et al., 2020; Saha and Dixit, 2020; Cardozo-Ojeda and Perelson, 2021; Stephenson et al., 2021*). These detailed mechanistic models cannot easily generalize from one trial to another in order to predict the efficacy of a new bNAb mono- or combination therapy.

Evolution of the HIV-1 population is another key factor to consider in modeling the dynamics and escape of viruses in response to therapy. Studies on population genetics of HIV-1 have found rapid intra-patient evolution and turnover of the virus (*Lemey et al., 2006; Zanini et al., 2015*) and have indicated that the efficacy of drugs in ART can severely impact the mode of viral evolution and escape (*Feder et al., 2016*). Despite the complex evolutionary dynamics of HIV-1 within patients due to individualized immune pressure (*Nourmohammad et al., 2019*), genetic linkage (*Zanini et al., 2015*), recombination (*Neher and Leitner, 2010; Zanini et al., 2015*), and epistasis between loci (*Bonhoeffer et al., 2004; Zhang et al., 2020*), the genetic composition of a population can still provide valuable information about the evolutionary significance of specific mutations, especially in highly vulnerable regions of the virus. For example, analysis of genomic covariation in the Gag protein of HIV-1 has been successful in predicting fitness effect of mutations in relatively conserved regions of the virus, which could inform the design of rational T-cell therapies that target these vulnerable regions (*Ferguson et al., 2013*). In the context of bNAb therapy trials, an evolutionary model accounting for the intrinsic fitness cost associated with escape variants against a specific bNAb has been used to characterize HIV-1 dynamics and escape following bNAb infusion (*Meijers et al., 2021*).

In this study, we present a coarse-grained evolutionary model of viral response to bNAb infusion that uses genetic data of HIV-1 in untreated patient to predict bNAb therapy outcome by characterizing the chances of viral escape from a given bNAb in patients. Specifically, we develop a statistical inference framework that uses the high throughput longitudinal survey of viral sequences collected from 11 ART-naïve patients over about 10 years of infection (Zanini et al., 2015) to characterize the evolutionary fate of escape mutations and to predict patient outcomes in recent mono- and combination therapy trials with 10–1074 and 3BNC117 bNAbs (Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018), and a trial with PGT121 bNAb (Stephenson et al., 2021). Using the accumulated intra-patient genetic variation from deep sequencing of HIV-1 populations in ART-naïve patients (Zanini et al., 2015), we can estimate the diversity and the fitness effects of mutations at sites mediating escape. These variables parametrize our individual-based model for viral dynamics and characterize the expected path for a potential escape of HIV-1 populations in response to bNAb therapies in patients enrolled in the clinical trials. Although our coarse-grained model does not accurately reproduce the detailed dynamics of viremia in each patient and lacks the mechanistic insight of richer models proposed by Perelson et al., 1996; Rong et al., 2010; Rong et al., 2007; Tomaras et al., 2008; Cardozo-Ojeda and Perelson, 2021; Stephenson et al., 2021, it still accurately predicts the distribution of viral rebound times in response to passive bNAb infusions — a key measure the efficacy for a bNAb clinical trial.

Our prediction for the viral rebound time in response to a bNAb relies on only a few patient-specific parameters (i.e. the genetic diversity of patients prior to treatment), and is primarily done based on the inferred genetic parameters from the deep sequencing of HIV-1 populations in a separate cohort of ART-naïve patient. Therefore, our model could be used to guide therapy trial design with bNAbs against which viral escape variants are previously characterized. To this end, we use our approach to assess a broader panel of nine bNAbs, for which escape sites can be identified from prior deep mutational scanning experiments (Dingens et al., 2019), to characterize the therapeutic efficacy of each of these bNAbs and to propose optimal combination therapies that can efficiently curb an HIV-1 infection. Our results showcase how the wealth of genetic data can be leveraged to guide rational therapy approaches against HIV. Importantly, this approach is potentially applicable to therapy designs against other evolving pathogens, such as chronic viruses like HCV, resistant bacteria, or cancer tumor cells.

Model

HIV-1 response to therapy

After infusion of bNAbs in a patient, the antibodies bind and neutralize the susceptible strains of HIV. The neutralized subpopulation of HIV-1 no longer infects T-cells, and the plasma RNA copy-number associated with this neutralized population decays. The dynamics of viremia in HIV-1 patients off ART following a bNAb therapy with 3BNC117 (Caskey et al., 2015), 10–1074 (Caskey et al., 2017), and their combination (Baron et al., 2018) are shown in **Figure 3—figure supplements 1–3**. With competition of the neutralized strains removed, the resistant subpopulation grows until the viral load typically recovers to a level close to the pretreatment state (i.e. the carrying capacity); see **Figure 1A**. The time it takes for the viral load to recover is the *rebound time*—a key quantity that characterizes treatment efficacy within a patient. Although the details of the viremia dynamics, especially at beginning and at the end of the therapy, may be complex (Lu et al., 2016; Reeves et al., 2020; Saha and Dixit, 2020; Meijers et al., 2021), the rebound time can be approximately modeled using a logistic growth after bNAb infusion ($t > 0$),

$$N(t) = \begin{cases} N_k & t \leq 0 \\ (1-x)N_k e^{-rt} + \frac{N_k}{1 + \frac{1-x}{x} e^{-\gamma t}} & t > 0 \end{cases} \quad (1)$$

with the initial condition set for pre-treatment fraction of resistant subpopulation $x = N_r(0)/(N_r(0) + N_s(0))$, where $N_r(0)$ and $N_s(0)$ denote the size of resistant and susceptible subpopulations at time $t = 0$, respectively. Here, γ is the growth rate of the resistant population, r is the neutralization rate impacting the susceptible subpopulation, and N_k is the carrying capacity (**Figure 1A**, Methods). In our analysis, we set $\gamma = 1/3 \text{ days}^{-1}$ or a doubling time of ~ 2 days, the known HIV-1 growth rate in patients (Perelson et al., 1996). We infer the neutralization rate r as a global parameter

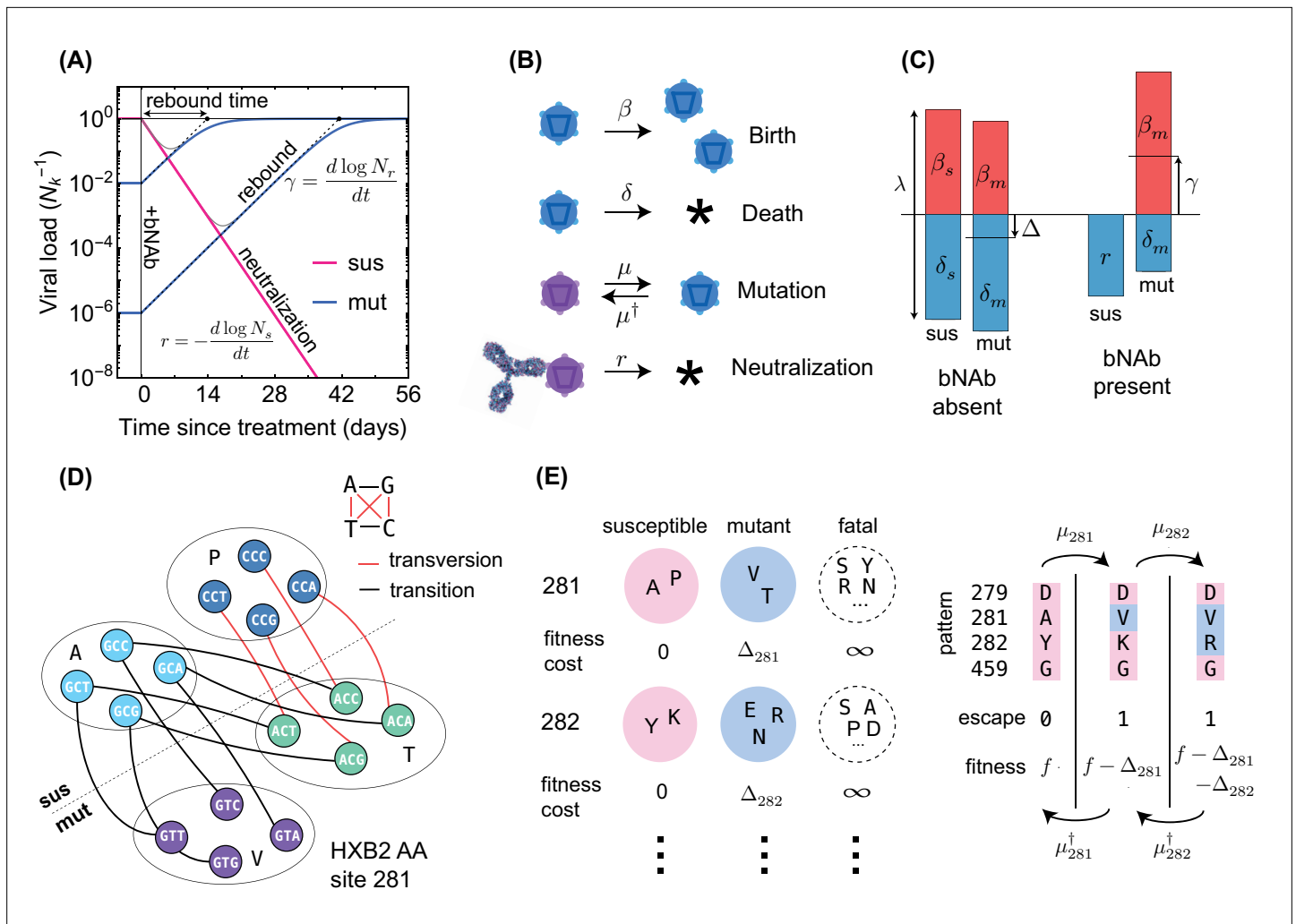


Figure 1. Schematics for the evolutionary dynamics of viral rebound. **(A)** The viral dynamics after the initiation of a treatment with bNAb infusion ($t = 0$) is determined by two competing processes. Susceptible strains (sus) undergo exponential decay (red line) with decay rate given by r , while the resistant mutants (mut) undergo logistic growth back up to the carrying capacity (N_k) of the patient. In the deterministic limit (**Equation 1**), the rebound time is linearly related to the log-frequency of the mutant fraction. **(B)** The schematic shows the four stochastic processes of birth, death, mutation, and neutralization with their respective rates for susceptible (purple) and resistant (blue) variants. These processes define the evolution of a viral population. Note that both the susceptible and the resistant variants are subject to birth and death with their respective rates. **(C)** The birth and death rates can be visualized as a region of size $\lambda = \beta + \delta$ which is partitioned into birth and death events. In the absence of antibodies, the susceptible population has balanced birth and death rates, $\beta_s = \delta_s$, while the resistant population has a negative net birth rate equal to the fitness difference $\Delta = \delta_m - \beta_m$. After introduction of the antibody, the susceptible population decays at rate r , and without competition from the susceptible population, the resistant population grows at the free growth-rate γ . **(D)** Mutational target size is inferred a priori from the genotype-phenotype mapping, which can be visualized as a bipartite graph. The nodes correspond to codons, while the edges are the mutations which link one codon to another, weighted according to the respective mutation rates. The average edge weight from codons of susceptible variants to the escape mutants determines the rate of escape mutations μ . Mutations can be divided into two types: transitions (black) are within-class, and transversions (red) are out of class nucleotide changes. Transitions occur at about 8 times higher rate than transversions (**Figure 2**). **(E)** A coarse grained fitness and mutation model for two of the escape sites (281 and 282) against antibody 3BNC117 are shown. Left: At each escape mediating site, amino acids fall into one of three groups: (i) susceptible (wild-type), (ii) escape mutant, and (iii) fatal. For an escape-class amino acid at site i the virus incurs a fitness cost Δ_i , and these costs are additive across sites. Right: Mutations at a given site i occur with (independent) forward μ_i and backward μ_i^\dagger rates which govern the substitution events between amino acid classes.

for each trial, since it depends on the neutralization efficacy of a bNAb at the concentration used in the trial. We will infer the patient-specific pre-treatment fraction of resistant subpopulation x , using a population genetics based approach based on which we characterize the mutational target size and selection cost of escape in the absence of a bNAb (see below). The maximum-likelihood fits of $N(t)$ to

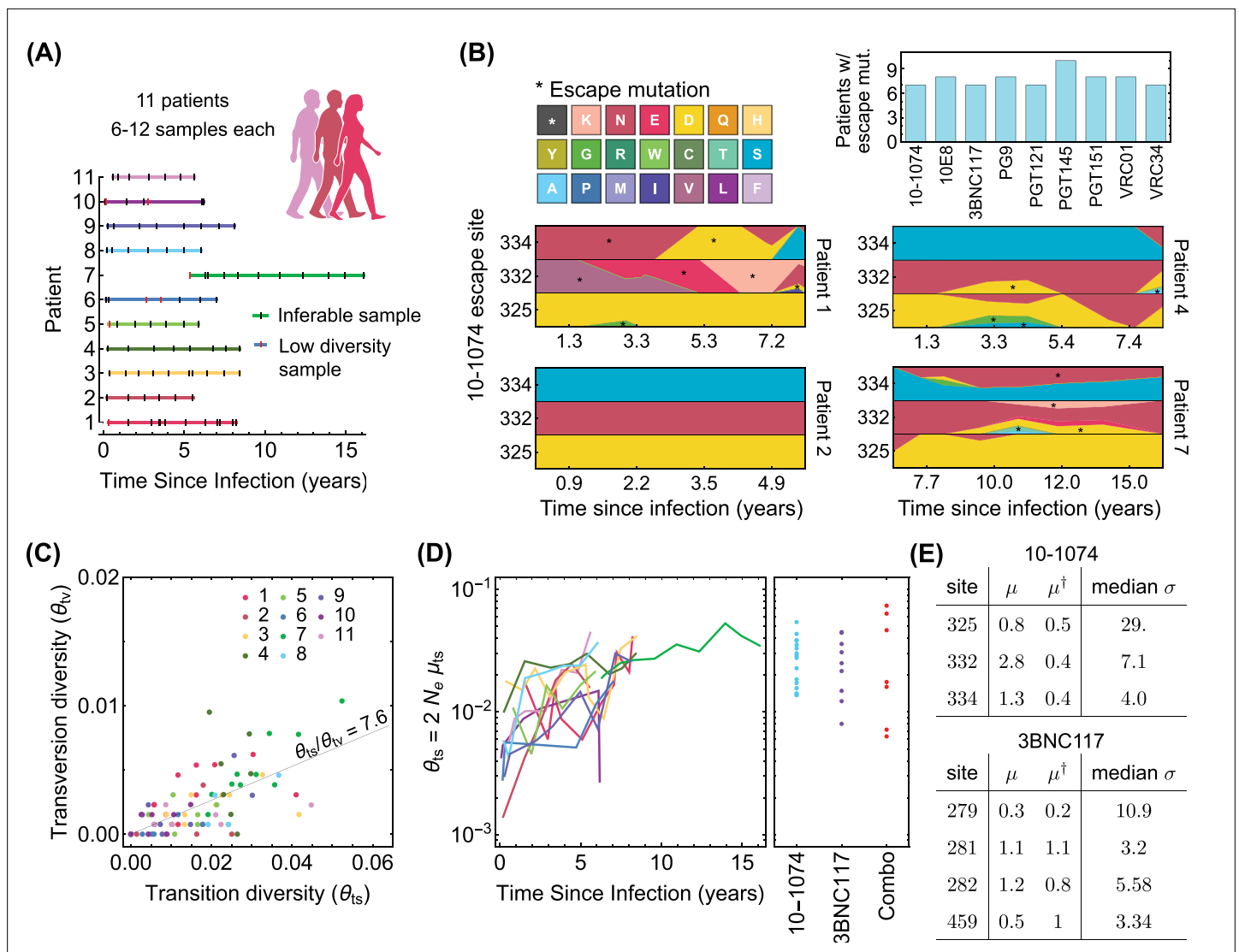


Figure 2. Statistics of viral genome sequences from bNAb-naive HIV-1 patients. **(A)** Statistics of the high-throughput longitudinal data collected from HIV-1 populations in 11 ART-naive patients from *Zanini et al., 2015*, is shown. Some of the have low diversity (vertical red lines) and were not usable for our study. Usable samples (vertical black lines) amount to 4–10 samples per patient, collected over 5–10 years of infection. **(B)** Lower panels show the relative frequencies (cube-root transformed for legibility) of different amino acids in four patients at the 3 escape sites against the 10–1074 bNAb, estimated from the polymorphism data at the nucleotide level in each patient over time. Despite 10–1074 being a broadly neutralizing antibody, mutations associated with escape (indicated by a *) are commonly observed in untreated patients. The upper right panel shows the number of individuals (out of the cohort of 11 bNAb-naive HIV-1 patients) that carry mutations associated with escape against the indicated bNAbs. **(C)** The nucleotide diversity associated with transversion $\theta_{tv} = 2N_e\mu_{tv}$ is shown against the transition diversity $\theta_{ts} = 2N_e\mu_{ts}$ for all patients (colors) and all time points. The covariance of these two diversity measurements yields an estimate for the transition/transversion ratio $\theta_{ts}/\theta_{tv} = 7.6$. **(D)** Left: The transition diversity is shown to grow as a function of time since infection in all the 11 patients (colors according to **(A)**). Right: The neutral diversity of viral populations in patients (points) from the three different clinical trials (*Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018*) analyzed in this study resemble the larger diversities of long-established viral populations in untreated patients. **(E)** The inferred forward and backward mutation rates (μ, μ^\dagger), relative to the transition rate, and the median selection strength $\sigma = 2N_e(f_{sus} - f_{res})$ at each escape site against the two bNAbs (10–1074, and 3BNC117) from the trial data used in this study are shown. Compared to the 10–1074 bNAb, escape from the 3BNC117 bNAb appears to be less costly, and is associated with a smaller mutational target.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Inference of neutral diversity from genetic data.

the viremia measurements in **Figures 2 and 3**, **Figure 3—figure supplements 1–3** specifies the initial resistant fraction x_p and the rebound time T_p in each patient p , which in this simple model, is given by $T_p = -\gamma^{-1} \log x_p$ (Methods).

The rebound times following passive infusion of 3BNC117 (**Caskey et al., 2015** and 10–1077 **Caskey et al., 2017**) bNAbs range from 1 to 4 weeks, with a small fraction of patients exceeding the monitoring time window in the studies (late rebounds past 56 days). The distribution of rebound times summarizes the escape response of the virus to a therapy and directly relates to the distribution for the pre-treatment fraction of resistant variants $P(x)$ across patients $P(T) \sim x^{-1}P(x)$.

Stochastic evolutionary dynamics of HIV-1 subject to bNAb therapy

The fate of an HIV-1 population subject to bNAb therapy depends on the composition of the pre-treatment population with resistant and susceptible variants, and the establishment of resistant variants following the treatment. To capture these effects, we construct an individual-based stochastic model for viral rebound (**Figure 1B**). We specify a coarse-grained phenotypic model, where a viral strain of type a is defined by a binary state vector $\vec{\rho}^a = [\rho_1^a, \dots, \rho_\ell^a]$, with ℓ entries for potentially escape-mediating epitope sites; the binary entry of the state vector at the epitope site represents the presence ($\rho_i^a = 1$) or absence ($\rho_i^a = 0$) of an escape mediating mutation against a specified bNAb at this site of variant a . We assume that a variant is resistant to a given antibody if at least one of the entries of its corresponding state vector is non-zero. For multivalent treatment, a virus must be resistant to all antibodies comprising the treatment to have a positive growth after infusion.

At each generation, a virus with phenotype a can undergo one of three processes: birth, death and mutation to another type b , with rates β_a , δ_a , and $\mu_{a \rightarrow b}$, respectively (**Figure 1B**). The net growth rate of the viral subpopulation with phenotype a is the birth rate minus the death rate, $\gamma_a = \beta_a - \delta_a$ (**Figure 1C**). The total rate of events (birth and death) per virion $\lambda = \beta_i + \delta_i$ modulates the amount of stochasticity in this birth-death process (Methods), which we assume to be constant across phenotypic variants. The continuous limit for this birth-death process results in a stochastic evolutionary dynamics for the sub-population size N_a ,

$$\frac{dN_a}{dt} = \begin{cases} \text{absence of bNAb or if } a \text{ is resistant :} \\ N_a (f_a - \phi) + \sum_b (N_b \mu_{b \rightarrow a} - N_a \mu_{a \rightarrow b}) \\ \quad + \sqrt{N_a \lambda} \eta(t) \\ \text{presence of bNAb and if } a \text{ is susceptible :} \\ -rN_a + \sum_b (N_b \mu_{b \rightarrow a} - N_a \mu_{a \rightarrow b}) + \sqrt{N_a \lambda} \eta(t) \end{cases} \quad (2)$$

where $\eta(t)$ is a Gaussian random variable with mean $\langle \eta(t) \rangle = 0$ and correlation $\langle \eta(t) \eta(t') \rangle = \delta(t - t')$ (Methods). Here, f_a denotes the intrinsic fitness of variant a and its net growth rate γ is mediated by a competitive pressure $\phi = \frac{1}{N_k} \sum_b N_b f_b$ with the rest of the population constrained by the carrying capacity N_k , such that $\gamma_a = f_a - \phi$. In the presence of a bNAb, birth is effectively halted for susceptible variants and their death rate is set by the neutralization rate of the antibody, resulting in a net growth rate, $\gamma_{\text{sus.}} = -r$. At the carrying capacity, the competitive pressure is the mean population fitness $\phi = \bar{f}$, making the net growth rate of the whole population zero (**Figure 1C**). When susceptible variants are neutralized by a bNAb, the competitive pressure ϕ drops, and as a result, the resistant variants can rebound to carrying capacity, at growth rates near their intrinsic fitness.

To connect the birth-death model (**Equation 2**) with data, we should relate the simulation parameters of a birth-death process to molecular observables. We have already made a connection between the birth and death rates of a variant and its intrinsic fitness in **Equation 2**. In addition, the neutral diversity θ of a population at steady-state can be expressed as $\theta = 2N_k \mu / \lambda$, where $\mu \approx 10^{-5}$ /day is the per-nucleotide mutation rate, which we infer from intra-patient longitudinal HIV-1 sequence data (**Zanini et al., 2015**) (Methods). For consistency, we set the total rate of events λ to be at least as large as the fastest process in the dynamics, which in this case is the growth rate of resistant viruses $\gamma \approx (3 \text{ days})^{-1}$, we choose $\lambda = (0.5 \text{ days})^{-1}$ (Methods). Therefore, the key parameters of the

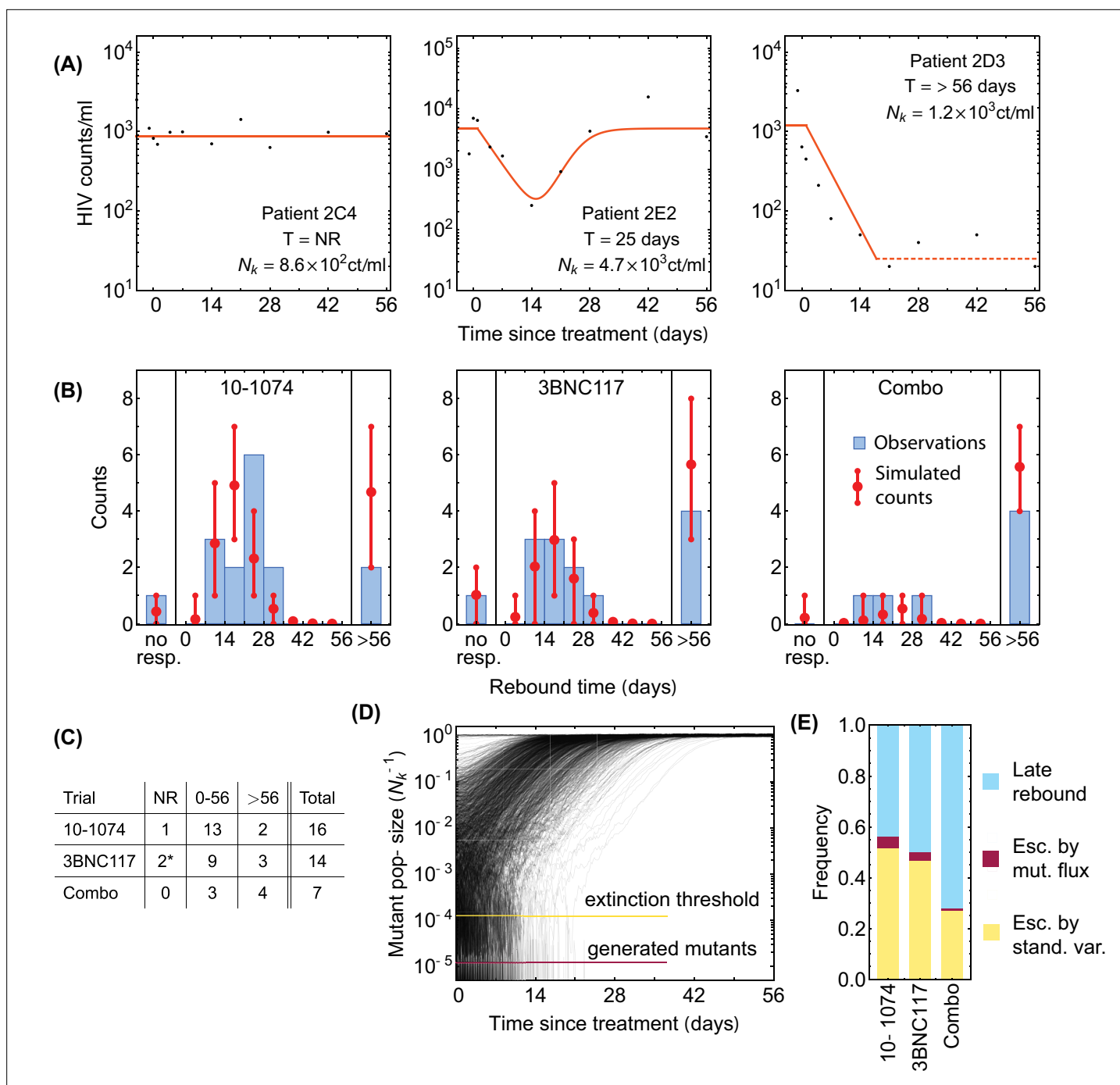


Figure 3. Statistics of viral rebound in clinical trials with bNAbs. **(A)** Panels show viremia of three patients from the 3BNC117 trial over time (black circles) and the fitted model of the viral decay and rebound processes from **Equation 1** (orange line). The viral rebound time T and the fitted carrying capacity N_k is shown in each panel. Shown are examples of a non-responder (NR; left), a rebound occurring during the trial window ($0 < T < 56$ days; center), and a late rebound ($T > 56$ days; right). **(B)** We compare the distribution of rebound times in patients from the three clinical trials with 10-1074 **Caskey et al., 2017**, 3BNC117 **Scheid et al., 2016**, and the combination of the two bNAbs **Baron et al., 2018** to the predictions from the simulations based on our evolutionary model (**Figure 1**, and Methods). The error bars show the inter decile range (0.1–0.9 quantiles) generated by the simulations for the corresponding trial. **(C)** The summary table shows the number of patients for whom the infecting HIV-1 population shows no response (NR), rebound during the trial window $0 < T < 56$, and a late rebound ($T > 56$ days) in each trial. Note that three patients were excluded from the 3BNC117 trial (*) because of insufficient dosage leading to weak viral response: 1mg/kg compared to the 3 – 30mg/kg in the other treatment groups. **(D)** Plotted are 1,200 trajectories of the mutant viral population simulated using our individual based model. Due to the individual birth-death events, fluctuations are larger when the population size is smaller. At a critical threshold, x_{ext} , fluctuations are large enough to lead to almost certain extinction in the existing

Figure 3 continued on next page

Figure 3 continued

viral population. The critical threshold (yellow line) is an order of magnitude larger than the post-treatment spontaneously-generated mutant fraction (red line). (E) The predicted fraction of escape events associated with post-treatment spontaneous mutations (red) and the pre-treatment standing variation (yellow) are shown for the three trials. Late rebound events are indicated in blue. Because the spontaneously-generated mutant fraction is smaller than the extinction threshold, these mutations contribute to less than 4% of escape events (red), and escape is likely primarily driven by standing variation (yellow), that is, pre-existing escape variants.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Dynamics of viremia in patients from the 10–1074 trial.

Figure supplement 2. Dynamics of viremia in patients from the 3BNC117 trial.

Figure supplement 3. Dynamics of viremia in patients from the combination therapy trial.

birth-death model, that is, β , δ , and N_k can be expressed in terms of the intrinsic fitness of the variants f_s and the neutral diversity θ , which we will infer from data.

Results

Population genetics of HIV-1 escape from bNAbs

HIV-1 escape from different bNAbs has been a subject of interest for vaccine and therapy design, and a number of escape variants against different bNAbs have been identified in clinical trials or in infected individuals (Lynch et al., 2015; Caskey et al., 2015; Scheid et al., 2016; Caskey et al., 2017; Baron et al., 2018). This in-vivo data is often complemented with information from co-crystallized structures of bNAbs with the HIV-1 envelope protein (Pancera et al., 2017), and in-vitro deep mutational scanning (DMS) experiments, in which the relative change in the growth rate of tens of thousands of viral mutants are measured in the presence of different bNAbs (Dingens et al., 2019; Dingens et al., 2017; Schommers et al., 2020). We identify escape mutations against each of the bNAbs in this study by using information from clinical trials, the characterized binding sites, and the DMS assays (Methods); the list of escape mutations against each bNAb is given in **Appendix 1—table 1**.

The rise and establishment of an escape variant against a specific bNAb depend on three key factors, (i) neutral genetic diversity of the viral population, (ii) the mutational target size for escape from the bNAb (i.e. the number of paths leading to escape, weighted by their respective probabilities), and (iii) the intrinsic fitness associated with such mutations. Although viremia traces in clinical trials and growth experiments can be used to model the escape dynamics (Haddox et al., 2018; Lynch et al., 2015; Lu et al., 2016; Reeves et al., 2020; Saha and Dixit, 2020; Meijers et al., 2021), they do not offer a comprehensive statistical description for HIV-1 escape as they are limited by the number of enrolled individuals. Alternatively, mutation and fitness characteristics of such escape-mediating variants can be inferred from a broader cohort of untreated and bNAb-naïve patients (Illingworth et al., 2020; Louie et al., 2018). We will infer statistical parameters for our coarse-grained fitness model (Figure 1E) from the large amount of high-throughput HIV-1 sequence data from Zanini et al., 2015 (see Figure 2A and B for details) and use them to parameterize the birth-death model (Figure 1B).

Diversity of the viral population

The neutral genetic diversity $\theta = 2N_k\mu/\lambda$ (i.e. the number of segregating alleles) is an observable that relates to key population genetics parameters, that is, the per-nucleotide mutation rate μ , the population carrying capacity N_k , and the total number of events per virus in the birth-death process λ , which determines the noise amplitude in the evolutionary dynamics (Methods). N_k and λ together determine the effective population size $N_e = N_k/\lambda$. We use synonymous changes as a proxy for diversity associated with the neutral variation in an HIV-1 population at a given time point within a patient. By developing a maximum-likelihood approach based on the multiplicities of different synonymous variants, we can accurately infer the neutral diversity of a population from the large survey of synonymous sites in the HIV-1 genome (Methods and Figure 2—figure supplement 1). Importantly, we infer the neutral diversity of transition θ_{ts} and transversion θ_{tv} mutations separately, and consistent with previous work (Feder et al., 2016; Zanini et al., 2017), find that transitions occur with a rate of about 8 times larger than transversions (Figure 2C, Figure 2—figure supplement 1B-E).

Our inference indicates that the neutral diversity grows over the course of an infection in untreated HIV-1 patients from *Zanini et al., 2015* (Figure 2D). The patients enrolled in the three bNAb trials (*Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018*) show a broad range of neutral diversity prior to bNAb therapy (Figure 2D). In addition to the circulating viruses in a patient's sera, the viral reservoir, which consists of replication-competent HIV-1 in latently infected cells or un-sampled tissue, can also contribute to a bNAb escape in a patient. Evidence that the latent reservoir can contribute to HIV escape from bNAbs is directly visible in trials as the failure of pre-trial sequencing to exclude patients who do not harbor escape variants. We model the effect of the reservoir as augmenting the neutral diversity by a constant multiplicative factor $r_{\text{resv},i}$ so that patients with more diverse sera, representing usually longer infections, are also expected to have correspondingly more diverse reservoir populations. By fitting the observed rebound data, we infer the reservoir factor $r_{\text{resv}} \simeq 2.07$ (Methods, Figure 4—figure supplement 1). We use the augmented genetic diversity of HIV-1 prior to the bNAb therapy in each trial to generate the rebound time and the probability of HIV-1 escape in patients.

Mutational target size for escape

We define the mutational target size for escape from a bNAb as the number of trajectories that connect the susceptible codon to codons associated with escape variants, weighed by their probability of occurrences (Methods). The connecting paths with only single nucleotide transitions or transversions dominate the escape and can be represented as connected graphs shown in Figure 1D. To characterize the target size of escape for each bNAb, we determine the forward mutation rate $\mu \equiv \mu_{\text{sus.} \rightarrow \text{res.}}$ from the susceptible codons to the resistant (escape) codons, and the reverse mutation rate $\mu^\dagger \equiv \mu_{\text{sus.} \leftarrow \text{res.}}$ back to the susceptible variant (Figure 1D, Methods). The mutational target sizes vary across bNAbs, with HIV-1 escape being most restricted from 10E8 ($\mu/\mu_{\text{ts}} = 1.8$ where μ_{ts} is the single-nucleotide transition substitution rate) and most accessible in the presence of 10–1074 ($\mu/\mu_{\text{ts}} = 4.9$); see Figure 4C and Appendix 1—table 1 for the list of mutational target size for escape against all bNAbs in this study.

Fitness effect of escape mutations

Since bNAbs target highly conserved regions of the virus, we expect HIV-1 escape mutations to be intrinsically deleterious for the virus (*Ferguson et al., 2013; Meijers et al., 2021*), and incur a fitness cost relative to pre-treatment baseline f_0 . We assume that fitness cost associated with escape mutations are additive and background-independent so the fitness of a variant a in the absence of bNAb follows, $f_a = f_0 - \sum_i \Delta_i \sigma_i^a$, where Δ_i is the cost associated with the presence of an escape mutation at site i (i.e., for $\sigma_i^a = 1$); see Figure 1D.

Interestingly, we observe the escape variants against different bNAbs to be circulating in the HIV-1 populations from the cohort of ART- and bNAb-naïve patients (*Zanini et al., 2015, Figure 2B*). We use this data (*Zanini et al., 2015*) and extract the multiplicity of susceptible and escape variants in HIV-1 populations at each sampled time point from a given patient. We use a single locus approximation under strong selection to represent the stationary distribution of the underlying frequency of escape alleles x in each patient from (*Zanini et al., 2015*), $P(x; \sigma, \theta, \theta^\dagger) \sim x^{-1+\theta} (1-x)^{-1+\theta^\dagger} \exp[-\sigma x]$, given the (scaled) fitness difference between the susceptible and the escape variants $\sigma = 2N_e(f_{\text{sus}} - f_{\text{mut}})$; see Methods.

Based on the statistics of escape and susceptible variants in all patients, we define a likelihood function that determines a Bayesian posterior for selection σ associated with escape at each site (Methods). We found that it is statistically more robust to infer the strength of selection relative to a reference diversity measure $\sigma/\theta_{\text{ts}} = (f_{\text{sus.}} - f_{\text{res.}})/\mu_{\text{ts}}$, for which we choose the transition rate (Methods). This approach generates unbiased selection estimates in simulations and is robust to effects of linkage and recombination (Methods and Model robustness Figure 4—figure supplement 1, Figure 4—figure supplement 2). The inferred values of the scaled fitness costs $\sigma/\theta_{\text{ts}}$ are shown for the escape-mediating sites of the trial bNAbs in Figure 2E, and are reported in Appendix 1—table 1.

Predicting the efficacy of bNAb therapy in clinical trials

Monotherapy trials with 10–1074 (*Caskey et al., 2017*, 3BNC117 *Caskey et al., 2015*, PGT121 *Stephenson et al., 2021*), and the combination therapy with 10–1074+3BNC117 in *Baron et al., 2018* have shown variable outcomes. In some patients, bNAb therapy did not suppress the viral load,

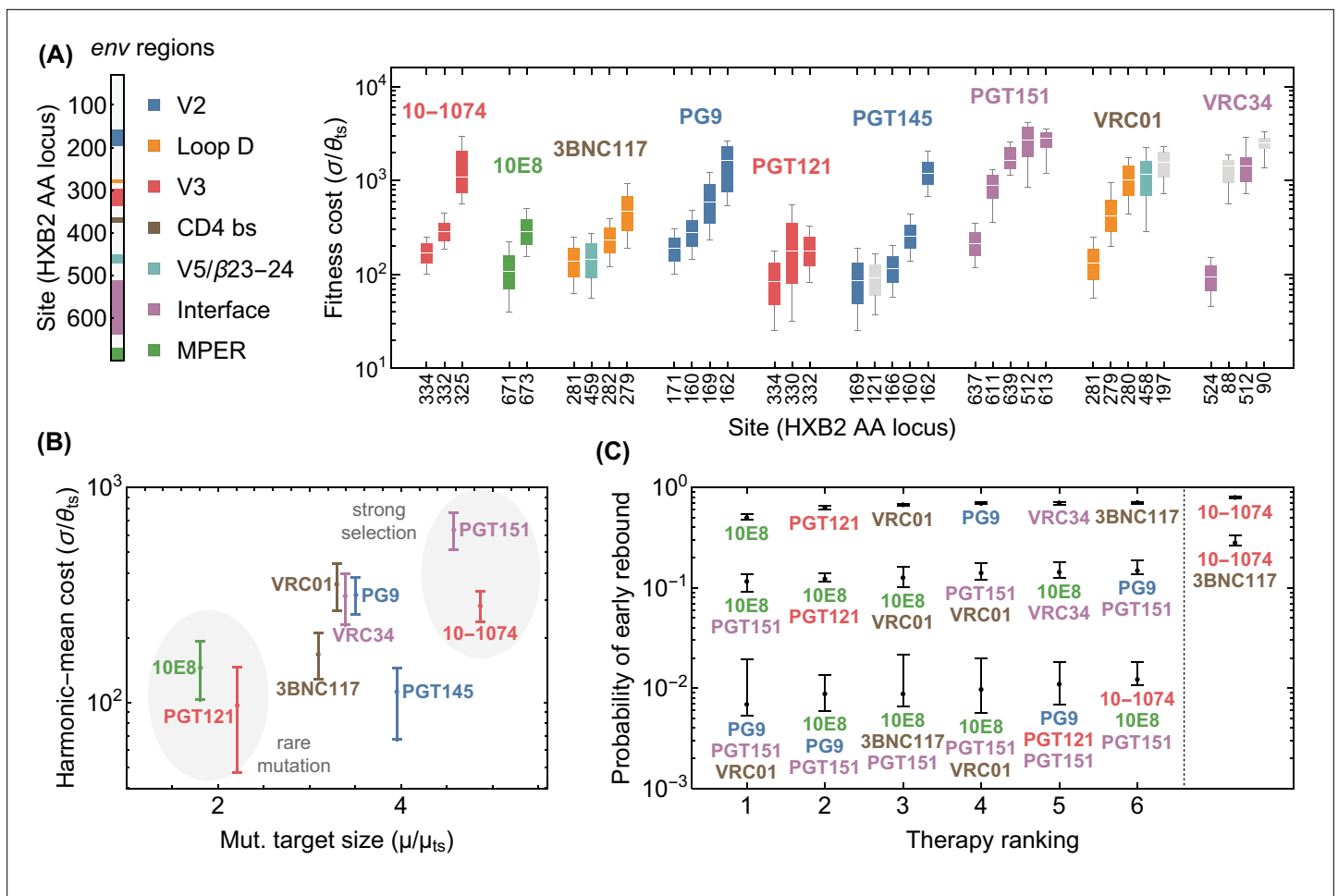


Figure 4. Statistics of viral escape for optimal combination therapy with bNAbs. **(A)** The posterior distribution for inferred selection strength on the escape-mediating sites associated with each of the 9 bNAbs in this study is shown (right); white line: median, box: 50% around the median, bar: 80% around median. Each escape site is color coded by its location on the *env* gene (left) and each antibody by its associated epitope location. **(B)** The harmonic mean of the selection strength σ associated with cost of escape (scaled by transition diversity θ_{ts}) is shown against the mutational target size for each bNAb; error bars indicate 50% around the median. For antibodies to be broadly neutralizing, it is sufficient that viral escape from them to be associated with a small mutational target size or a large fitness cost. The mutational target size is found to be weakly correlated with the average cost of escape from a given bNAb. We identify two distinct strategies for antibody breadth—selection limited and mutational-target-size limited escape pathways each highlighted in gray. **(C)** bNAb therapies with 1, 2, and 3 antibodies are ranked based on the predicted probability of early viral rebound, and in each case, six therapies with highest efficacies are shown; best ranked therapy is associated with the lowest probability of early rebound (It_{50} days); indicate 50% around the median. Also for reference, the probability of early viral rebound two therapies from the trials in this study (10–1074 and 10–1074+3BNC117) are shown.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Robustness of selection inference to intra-patient temporal correlations of HIV-1 alleles and clade-specific sampling.

Figure supplement 2. Robustness of selection inference to genetic linkage and hitchhiking.

Figure supplement 3. Minimum disparity estimation for adjustment of diversity.

Figure supplement 4. Disparity analysis for robustness of selection inference.

whereas in others suppression was efficient and no rebound was observed up to 56 days after infusion (end of surveillance in these trials); see **Figure 3A** for examples of patients with different rebound times, **Figure 3—figure supplements 1–3** for the viremia traces in all patients, and **Figure 3B and C** for the distributions and the summary statistics of the rebound times in patients in different trials.

Although we infer a large intrinsic fitness cost for a virus to harbor an escape allele (**Figure 2E**), these variants can emerge or already be present due to the large intra-patient diversity of HIV-1 populations (**Figure 2C**), or a larger mutational target size for these escape variants. Deep sequencing

data in untreated (likely bNAb-naïve) patients shows circulation of resistant variants against a panel of bNAbs in the majority of patients (**Figure 2B**). Our goal is to predict the efficacy of a bNAb trial, using the fitness effect and the mutational target size for escape from a given bNAb, both of which we infer from the high-throughput HIV-1 sequence data collected from bNAb-naïve patients in **Zanini et al., 2015 (Figure 2E, Appendix 1—table 1)**. In addition, we modulate these measures with the patient-specific neutral diversity θ inferred from whole genome sequencing of HIV-1 populations in each patient prior to bNAb therapy (**Figure 2D**). These quantities parametrize the birth-death process for viral escape in a bNAb therapy (**Figure 1A**), which we use to characterize the distribution of rebound times in a given trial (Methods).

For both the 3BNC117 and the 10–1074 trials (**Caskey et al., 2015; Caskey et al., 2017**), we see an excellent agreement between our predictions of the rebound time distribution and data; see **Figure 3B**, and Methods and **Figure 4—figure supplement 3** for statistical accuracy of this comparison. For the PGT121 trial, the genomic data from patients' HIV-1 populations are insufficient and therefore, we used the neutral diversity θ_{ts} estimated from the other three trials to predict the associated rebound time distribution (Appendix 4). Still, we see a good agreement between our predictions of the rebound time distribution and data; see **Appendix 4—figure 1**.

By assuming an additive fitness effect for escape from 10 to 1074 and 3BNC117, we also accurately predict the distribution of rebound times in the combination therapy (**Baron et al., 2018, Figure 3B**). The agreement of our results with data for combination therapy is consistent with the fact that the escape mediating sites from 10 to 1074 and 3BNC117 are spaced farther apart on the genome than 100 bp, beyond which linkage disequilibrium diminishes due to frequent recombination in HIV (**Zanini et al., 2015**). Importantly, in all the trials, our evolutionary model accurately predicts the fraction of participants for whom we should expect a late viral rebound (more than 56 days past bNAb infusion)—the quantity that determines the efficacy of a treatment.

Apart from the overall statistics of the rebound times, our stochastic model also enables us to characterize the relative contributions of the pre-treatment standing variation of the HIV-1 population versus the spontaneous mutations emerging during a trial to viral escape from a given bNAb. Given the large population size of HIV-1 and a high mutation rate ($\mu = 10^{-5}$ per generation), spontaneous mutations generate a fraction $x^{(\mu)}$ of resistant variants during a trial, which we can express as,

$$x^{(\mu)} = \int_0^{56 \text{ wks}} (1 - x(0))e^{-rt} \mu \gamma dt \quad (3)$$

In the best case scenario, there are no resistant virions prior to treatment i.e., $x(0) = 0$. Since the neutralization rate r and the growth rate γ are comparable, this deterministic approach predicts that mutations can generate a resistant fraction of $x^{(\mu)} \approx 10^{-5}$ during a trial. However, stochastic effects from random birth and death events play an important role in the fate and establishment of these resistant variants. The probability of extinction for a variant at frequency x can be approximated as $p(\text{extinct}) \approx 1 - e^{-x/x_{\text{ext}}}$ (see Extinction Probability); here $x_{\text{ext}} = \frac{\mu_{ts}}{\gamma\theta_{ts}} \approx 10^{-4}$ and a variant with fraction x that falls below this critical value is likely to go extinct (**Figure 3D**). Since the total integrated mutational flux fraction during a trial is $x^{(\mu)} \sim 10^{-5}$, mutational flux rarely decides the outcome of patient treatment. Indeed, we infer that spontaneous mutations contribute to less than 4% of escape events in all the three trials and escape is primarily attributed to the standing variation from the serum or the reservoirs prior to therapy (**Figure 3E**). A similar conclusion was previously drawn based on a mechanistic model of escape in VRC01 therapy trials (**Saha and Dixit, 2020**).

Devising optimal bNAb therapy cocktails

Clinical trials with bNAbs have been instrumental in demonstrating the potential role of bNAbs as therapy agents and in measuring the efficacy of each bNAb to suppress HIV. Still, these clinical trials can only test a small fraction of the potential therapies that can be devised. It is therefore important that trials test therapies that have been optimized based on surrogate estimates of treatment efficacy. The accuracy of our predictions for the rebound time of a HIV-1 population subject to bNAb therapy suggests a promising approach to the rational design of therapies based on genetic data of HIV populations collected from bNAb-naïve patients.

Here, we use viral genome sequences to infer the efficacy of therapies with bNAbs, for which clinical trials are not yet performed. To do so, we first need to identify the routes of HIV escape from these bNAbs. We use deep mutational scanning data (DMS) on HIV-1 subject to 9 different bNAbs

from *Dingens et al., 2019* together with information from literature to identify the escape mediating variants from each of these bNAbs (Methods and Table S1). We then determine the mutational target size and the fitness effect of these escape variants using high-throughput sequences of HIV-1 in bNAb-naive patients from *Zanini et al., 2015*; these inferred values are reported in **Figure 4A and B** and Table S1. Using the inferred fitness and mutational parameters and by setting the pre-treatment neutral diversity θ to be comparable to that of the patients in the previous three trials with 10–1074 and 3BNC117 (**Figure 2C**), we simulate treatment outcomes for these 9 bNAbs and their twofold and threefold combinations.

Interestingly, we infer that escape from mono-therapies is almost certain and a combination of at least three antibodies is necessary to limit the probability of early rebound (<56 days) to below 1% (**Figure 4C**). When considering each of the nine antibodies, interesting patterns emerge. We find that the mutational target size and the fitness cost of escape, estimated as the harmonic-mean selection cost of individual sites, obey a roughly linear relationship (**Figure 4B**). As all these bNAbs have similar overall breadth (i.e. they neutralize over 70% of panel strains), this result suggests that for an antibody to be broad, its escape mediating variants should either be rare (i.e. small mutational target size) or intrinsically costly (i.e. incurring a high fitness cost), but it is not necessary to satisfy both of these requirements. For instance, we find that resistant variants against 10E8 weaker negative selection but escape target size is small, while PGT151 has a larger escape target size but makes up for it by having resistant variants with unavoidably high fitness cost.

The fitness-limited versus the mutation-limited strategies have different implications for the design of combination cocktails. The small mutational target size of 10E8 makes it the best candidate antibody for mono-therapy among the antibodies we consider because the escape variants against this antibody are less likely to circulate in a patient's serum prior to treatment. However, in combination, 10E8 appears less often in top ranked therapies than PGT151. PGT151 is unremarkable on its own because of a relatively large target size, but the high cost of escape makes it especially promising in combination therapies. Overall, fitness-limited bNAbs like PGT151 are more effective against high diversity viral populations, while mutation-limited bNAbs such as 10E8 are more effective against low diversity viral populations. Indeed, the best ranked therapy, namely the combination of PG9, PGT151, and VRC01, combines antibodies that target different regions of the virus and also have both types of fitness- and mutation-limited strategies for coverage against the full variability of viral diversities found in pre-treatment individuals (**Figure 4C**) participating in the clinical trials.

Escape from bNAbs in multivalent therapy is also influenced by the non-independence of escape pathways. Several of the bNAbs (e.g. 10–1074 and PG121) in this study target the same (or structurally adjacent) epitopes and escape from them is mediated by the same mutations. The sharing of mutational pathways invalidates the assumption of independence and may make our predictions too optimistic. However, the best performing antibody combinations in **Figure 4C** target multiple epitopes to reduce the chances of collective escape. Therefore, the assumptions of independent fitness effects and independent mutational pathways in this case are not consequential for the main predictions of our model (i.e., the choice of bNAb combinations). In vitro data shown in *Kong et al., 2015* suggest that additive and independent effects are the norm, but that small but consistent synergistic effects may imply that our assumption of site-wise independence is conservative. More data on HIV escape from combinations of bNAbs would be informative for further modeling efforts and relevant for long-term therapy design.

Discussion

HIV therapy with passive bNAb infusion has become a promising alternative to anti-retroviral drugs for suppressing and preventing the disease in patients without a need for daily administration. The current obstacle is the frequent escape of the virus seen in mono- and even combination bNAb therapy trials (*Caskey et al., 2015; Bar et al., 2016; Caskey et al., 2017; Baron et al., 2018*). The key is to identify bNAb cocktails that can target multiple vulnerable regions on the virus in order to reduce the likelihood for the rise of resistant variants with escape-mediating mutations in all these regions. Identifying an optimal bNAb cocktail can be a combinatorially difficult problem, and designing patient trials for all the potential combinations is a costly pursuit.

Here, we have proposed a computational approach to predict the efficacy of a bNAb therapy trial based on population genetics of HIV escape, which we parametrize using high-throughput HIV-1

sequence data collected from a separate cohort of bNAb-naive patients (Zanini *et al.*, 2015). Specifically, we infer the mutational target size for escape and the fitness cost associated with escape-mediating mutations in the absence of a given bNAb. These quantities together with the neutral diversity of HIV-1 within a patient parametrize our stochastic model for HIV dynamics subject to bNAb infusion, based on which we can accurately predict the distribution of rebound times for HIV in therapy trials with 10–1074, 3BNC117 and their combination, as well as a trial with PGT121 bNAb. Consistent with previous work on VRC01 (Saha and Dixit, 2020), we found that viral rebounds in bNAb trials are primarily mediated by the escape variants present either in the patients' sera or their latent reservoirs prior to treatment, and that the escape is not likely to be driven by the emergence of spontaneous mutations that establish during the therapy.

One key measure of success for a bNAb trial is the suppression of early viral rebound. Our model can accurately predict the rebound times of HIV-1 subject to three distinct therapies (Caskey *et al.*, 2015; Caskey *et al.*, 2017; Baron *et al.*, 2018), based on the fitness and the mutational characteristics of escape variants inferred from high-throughput HIV-1 sequence data. This approach enables us to characterize routes of HIV-1 escape from other bNAbs, for which therapy trials are not available, and to design optimal therapies. We used deep mutational scanning data (Dingens *et al.*, 2019) to identify escape-mediating variants against 9 different bNAbs for HIV. Our genetic analysis shows that bNAbs gain breadth and limit viral escape either due to their small mutational target size for escape or because of the large intrinsic fitness cost incurred by escape mutations. bNAbs with mutation-limited strategy are more effective at preventing escape in patients with low viral genetic diversity, while bNAbs with selection-limited strategy are more effective at high viral diversity.

Combination therapy with more than two bNAbs (or drugs in ART) has long been shown to be more effective in suppressing early viral rebound, both in theory and practice (Perelson *et al.*, 1996; Feder *et al.*, 2016; Wagh *et al.*, 2016; Klein *et al.*, 2012; Mendoza *et al.*, 2018; Yu *et al.*, 2019). In addition to corroborating this conclusion quantitatively, we provide a method for assessing new bNAbs for which escape mutations are known. Our method can be understood as a tool to navigate the combinatorial explosion of higher order cocktails for which we cannot possibly test all combinations. By assessing the evolvability of resistance against different combinations we can identify the best therapies to target for clinical trial. Specifically, we show that to suppress the chance of viral rebound to below 1%, we show that a combo-therapy with 3 bNAbs with a mixture of mutation- and selection-limited strategies that target different regions of the viral envelope is necessary. Such combination can counter the full variation of viral diversity observed in patients. We found that PG9, PG151, and VRC01, which respectively target V2 loop, Interface, and CD4 binding site of HIV envelope, form an optimal combination for a 3-bNAb therapy to limit HIV escape in patients infected with clade B of the virus.

The statistical agreement between our coarse-grained model and the observed distribution of the viral rebound times in trials (Figure 3B) implies that many of the mechanistic details are of secondary importance in predicting viral escape. Nonetheless, our approach falls short of reproducing the detailed characteristics of viremia traces in patients, especially at very short or very long times, during which the dynamics of T-cell response or the decay of bNAbs could play a role (Lu *et al.*, 2016; Reeves *et al.*, 2020; Saha and Dixit, 2020). The relationship between the short-term suppression of the virus, which is the focus of this analysis, and the long-term treatment success is complicated by reestablishment of HIV from the latent reservoirs and different modes of intra-host HIV evolution (Liu *et al.*, 2019; Margolis *et al.*, 2017).

One strategy to achieve a longer term treatment success is by combining bNAb therapy with ART. One main advantage of bNAb therapy is the fact that it can be administered once every few months, in contrast to ART, which should be taken daily and missing a dose could lead to viral rebound. Although multivalent bNAb therapy reduces the chances of short-term viral escape, viral escape remains a real obstacle for longer term success of a treatment with bNAbs. Alternatively, (fewer) bNAbs can be administered in combination with ART (Horwitz *et al.*, 2013; Gruell and Klein, 2018), whereby ART could lower the replication rate of the HIV population, reducing the viral diversity and the chances of viral escape. Specifically, we can expect that emergence and establishment of rare (i.e., strongly deleterious) escape variants against bNAbs to be less likely in ART+ patients, which suggests that fitness-limited bNAbs should be more effective in conjunction with ART. More data would be necessary to understand the long-term efficacy of such augmented therapy, and specifically the role of viral

reservoirs in this context. A modeling approach could then shed light on how ART administration and bNAb therapy could be combined to efficiently achieve viral suppression.

Limitations

We rest our analysis primarily on the predictive power of the observed variant frequencies in the untreated patients. Our model weighs these frequencies with respect to the viral diversity in a mathematically and biologically consistent way. However, we ignore the dynamics of antibody concentration and IC_{50} neutralization during treatments, the details of T-cell dynamics during infection (Perelson, 2002), and also the evolutionary features of the genetic data, such as epistasis between loci (Bonhoeffer et al., 2004; Zhang et al., 2020, genetic linkage Zanini et al., 2015, and codon usage bias Meintjes and Rodrigo, 2005).

In our model of viral escape, we neglect the possibility of incomplete escape of the virus due to the reduced neutralization efficacy of bNAbs as their concentrations decay during trials. In Appendix 3, we show that this simplifying assumption is valid as long as the IC_{50} is not the same order of magnitude as the initial dosage concentration of the infused bNAb. Notably, the data from therapy trials used in this study fall into the regime for which we can neglect the impact of incomplete neutralization (Appendix 3—figure 2). However, taking into account the dependence of viral fitness on bNAb concentration and its neutralization efficacy, as in the model proposed by Meijers et al., 2021, could improve the long-term predictive power of our approach.

Our predictions are limited by our ability to identify the escape variants for each bNAb, either based on the in vivo trial data and patient surveillance, or in-vitro assays such as DMS experiments. In Appendix 4, we compare the accuracy of our predictions for the rebound time distributions of HIV using DMS-inferred versus trial-inferred escape variants against the 10–1074 and the PGT121 bNAbs, and we find good agreements between the two approaches in both cases. However, it should be noted that identifying escape variants from DMS experiments lead to a more optimistic prediction for treatment success during the first 8 weeks of trials (i.e. they suggest a later rebound). One reason for this discrepancy may be related to the distinct genetic composition of viruses in the two approaches. DMS experiments characterize HIV escape by introducing mutations on a single genetic background (Dingens et al., 2019; Dingens et al., 2017; Schommers et al., 2020) (e.g. the HIV-1 strain BF520.W14M.C2 in Dingens et al., 2017), whereas clinical data contain diverse populations of viruses between and within individuals. It is more likely for diverse viral populations to contain variants in which positive epistasis between escape mutations and the background genome is present. Therefore, it is reasonable to expect that genetic data originating in clinical trials show more pathways of escape, resulting in a faster viral rebound following bNAb infusion. Nevertheless, DMS experiments are less costly for identifying escape variants compared to trials, and they provide a baseline to assess the efficacy of different bNAbs for therapy and further in-vivo investigations. More experiments would be necessary for a more systematic understanding of the limitations of each approach.

It should be noted that our analysis in Figure 4C only focuses on one aspect of therapy optimization, that is, the suppression of escape. Other factors, including potency (neutralization efficacy) and half-life of the bNAb, or the patient's toleration to bNAbs at different dosages should also be taken into account for therapy design. For example, the bNAb 10E8, which we identified as of the most promising mono-therapy candidates in Figure 4C, is shown to be poorly tolerated by patients with short half-life (Kwon et al., 2016), making it undesirable for therapy purposes. Thus, the bNAb candidates shown in Figure 4C should be taken as a guideline to be complemented with further assessment of efficacy and safety for therapy design.

Outlook

Our approach showcases that, when feasible, combining high-throughput genetic data with ecological and population genetics models can have surprisingly broad applicability, and their interpretability can shed light into the complex dynamics of pathogens subject to therapy. Application of similar methods to therapy design to curb the escape of cancer tumors against immune- or chemo-therapy, the resistance in bacteria against antibiotics, or the escape of seasonal influenza against vaccination is a promising avenue for future work. However, we expect that more sophisticated methods for inferring fitness from evolutionary trajectories may be necessary to capture the dynamical response of these populations.

Methods

Data and code accessibility

The code for the algorithms used in this work and the data are available on GitHub at <https://github.com/StatPhysBio/HIVTreatmentOptimization> (copy archived at [swh:1:rev:0194cf6e554996a066633e99d-d53cd5901da552e](https://doi.org/10.1101/2022.01.19.476633), *Chelate*, 2022a) and in the Julia package <https://github.com/StatPhysBio/EscapeSimulator> (copy archived at [swh:1:rev:9a343f598820bafddfc7ea4547cefa90bf96fd6e](https://doi.org/10.1101/2022.01.19.476633), *Chelate*, 2022b).

Description of molecular data

Data from bNAb trials

In this study, we considered four clinical trials for passive therapy with bNAbs:

- Monoclonal therapy with 3BNC117 bNAb *Caskey et al., 2015* and sequences reported in *Schoofs et al., 2016*. There were 16 patients enrolled, 13 of whom were off anti-retroviral therapy (ART).
- Monoclonal therapy with 10–1074 bNAb *Caskey et al., 2017*, with 19 patients enrolled, 16 of whom were off ART.
- Combination therapy with 10–1074+3BNC117 bNAb *Baron et al., 2018*, with 7 patients off ART.
- Monoclonal therapy with PGT121 with 17 patients off ART. *Stephenson et al., 2021*. Far fewer viral sequences were available for this trial, and therefore, we could not reliably infer the neural diversity of viruses within patients, as we did for other trials. We only used this trial to compare our predictions based on the DMS-inferred with the trial-inferred escape pathways in Appendix 4.

All sequenced patients across all trials were infected with distinct HIV-1 clade B viral strains. We limited our analyses to those patients not on ART at the time of treatment initiation. In these studies, the injected bNAb level falls off over time within patients and therefore, we only considered dynamics within an 8-week window since infusion. This assures that rebound is not confounded by a drop in bNAb below sensitive-strain neutralizing levels of $IC_{50_s} < 2\mu\text{g}/\text{m}$.

We used single-genome sequence data of *env* collected from all patients in each trial to characterize the diversity of HIV-1 population within each patient shown in *Figure 2* available from *Zanini et al., 2015* and accessible through European Nucleotide Archive (Accession no: PRJEB9618). The patient sequence data for each trial is available through *Caskey et al., 2015* GeneBank accession number: KX016803, *Caskey et al., 2017* GenBank accession numbers KY323724.1 - KY324834.1, and *Baron et al., 2018*, GenBank accession numbers MH632763 - MH633255.

Longitudinal HIV-1 sequence data from untreated patients

Single nucleotide polymorphism (SNP) data was obtained from *Zanini et al., 2015* and aligned to the HXB2 reference using HIV-align tool (*Gaschen et al., 2001*). The dataset includes 11 patients observed for 5–8 years of infection, with HIV-1 sequence data sampled over 6–12 time points per patient (*Figure 2*).

Patient 4 and patient 7 were excluded from the original analysis done in *Zanini et al., 2015* because of suspected superinfection and failure to amplify early samples, respectively. These patients were included in our analysis, since (i) super-infection poses no additional difficulties for our tree-free procedure, and (ii) only time points with measurable viral diversity entered into our selection likelihood, which automatically limits our analysis to samples with high-quality sequences.

All patients were infected with clade B of HIV, except for patient 6 (clade C), and patient 1 (clade O1_AE). We assessed the robustness of our inference to exclusion of these patients from our analysis in Effect of genomic linkage on the inference of selection. Overall, our inference was not strongly affected by this choice (Model robustness *Figure 4—figure supplement 1*), and therefore, we included these patients in our main analyses to enhance the statistics with larger data.

For our analysis we considered only data reported in single nucleotide polymorphism (SNP) counts. The number of raw SNP counts are the result of amplification and must be converted into estimates for the number of pre-amplification template fragments. Zanini et. al. reported that on average about 10^2 templates of amplicon were associated with the fragments of the envelope (*env*) protein *Zanini*

et al., 2015. We converted raw SNP counts into templates by normalizing to 120 counts and rounding the resulting number to the nearest integer.

Identifying escape-mediating variants against bNAbs

The starting point for our analysis of HIV response to a given a bNAb is the description of the escape-mediating amino acids in the HIV *env* protein. We use a combination of methods to identify the escape variants for a given bNAb. First, we use deep mutational scanning (DMS) data of HIV-1 in the presence of a bNAb from *Dingens et al., 2019* to identify these mutations. These DMS experiments have created libraries of all single mutations from a given genomic background of HIV-1 and tested the fitness of these variants (i.e. growth on T-cell culture) in the absence and presence of nine different bNAbs, including the 10–1074 and 3BNC117 *Dingens et al., 2019*.

Escape variants in DMS data are identified as those which are strongly selected for only in the presence of a bNAb. Specifically, we identify escape variants as those which show 3-logs-change in their frequency in the presence versus absence of a bNAb. DMS data reflects in vitro escape in cell culture. However, some of these variants may not be viable in vivo. To identify the reasonable candidates of escape in vivo, we limit our set to the variants that are also observed in the circulating viral strains of untreated HIV-1 patients from *Zanini et al., 2015*. It should be noted that since we use HIV-1 sequence data from *Zanini et al., 2015* to infer selection on escape mediating variants in the absence of a bNAb, the candidates of escape that are not observed in the dataset *Zanini et al., 2015* would be inferred to be strongly deleterious, and hence, unlikely to contribute to our predictions of viral rebound.

Our analysis of DMS data results in a set of escape mediating amino acids for 10–1074 that is consistent with the escape variants that emerge in response to the bNAb trial (*Caskey et al., 2017*; *Stephenson et al., 2021* Appendix 4). However, the DMS data is very noisy for bNAbs that target CD4 binding site of HIV, that is, 3BNC117 and VRC01 (*Dingens et al., 2019*). One reason for this observation may be that the CD4 binding site is crucial for the entry of HIV to the host's T-cells and mutations in this region are highly deleterious. As a result, only a small number of variants with mutations in this region can survive in the absence of a bNAb in a DMS experiment. Growth in the absence of a bNAb is the first step in the DMS experiments, which is then followed by exposure of the replicated variants to a bNAb. Therefore, a low multiplicity of variants in the absence of a bNAb could result in a noisy pattern of growth of the small subpopulation in the next stage of the experiment, in which growth is subject to a CD4-targeting bNAb.

For the CD4 binding site antibodies 3BNC117 and VRC01, we used additional data to call the escape variants. For 3BNC117, we used a combination of trial-patient sequences (*Caskey et al., 2015*; *Scheid et al., 2016*), that is post-treatment enrichment, along with contact site information compiled in the crystallographic studies to narrow down candidate sites (*Zhou et al., 2015*; *LaBranche et al., 2018*). For VRC01, we assumed a similar escape pattern to 3BNC117 but included sites known from other studies (*Lynch et al., 2015*) and the clear DMS signal at HXB2 site 197. The sites we called were similar to those identified using humanized-mouse models of HIV infection (*Horwitz et al., 2013*), although more complex mutational patterns were seen in the soft-randomization scanning of *Otsuka et al., 2018*. Although the complete list of escape substitutions are unknown and background-dependent (*Otsuka et al., 2018*), the escape profiles which are most important are those that are most likely to be seen consistently in data and to be correctly identified. The list of substitutions are shown in **Appendix 1—table 1**.

Statistics and dynamics of viral rebound

Inference of growth parameters from dynamics of viremia

The concentration of viral RNA copies in blood serum is a delayed reflection of the total viral population size $N(t)$, containing a resistant and susceptible subpopulations, with respective sizes $N_r(t)$ and $N_s(t)$. After infusion of bNAbs in a patient, the susceptible sub-population decays due to neutralization by bNAbs and the resistant sub-population grows and approaches the carrying capacity N_k , with the dynamics,

$$\begin{aligned}\frac{dN_r}{dt} &= \gamma N_r (1 - N_r/N_k) \\ \frac{dN_s}{dt} &= -r N_s\end{aligned}\tag{4}$$

Here, γ is the growth rate of the resistant population, and r is the neutralization rate impacting the susceptible subpopulation. By setting the initial condition for fraction of resistant subpopulation prior to treatment (at time $t = 0$) $x = N_r(0)/(N_r(0) + N_s(0))$, we can characterize the evolution of the total viremia in a patient. This dynamics is governed by the combined processes of neutralization by the infused bNAbs and the viral rebound (Figure 1A), which entails,

$$N(t) = \begin{cases} N_k & t \leq 0 \\ (1-x)N_k e^{-rt} + \frac{N_k}{1 + \frac{1-x}{x} e^{-\gamma t}} & t > 0 \end{cases} \quad (5)$$

We use Equation 5 to define the evolution of blood concentration of viral RNA sequences which is observed indirectly via noisy viremia measurement data from Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018. To connect the data with the simple model of viral dynamics in Equation 5, we fit the initial frequency of resistant mutants x for each patient separately, and fit a global estimate for the decay rate of susceptible variants r shared across all patients in a trial, using a joint maximum-likelihood procedure. In addition, we fix the growth rate γ to 1/3 days, corresponding to a doubling time of approximately 2 days Garcia et al., 1999. Our analyses indicate that the initial viremia decline lagged treatment by about 1 day (Figure 3—figure supplements 1–3), consistent with previous findings (Ioannidis et al., 2000), and therefore, we included a 1 day lag between the fitted viremia response model and the treatment.

$k = n_p(t)$ The number of viral RNA copies in a blood sample is subject to count fluctuations with respect to the true number of circulating virions in a given volume of the blood. We use a Poisson sampling model to define a likelihood for our model of viral population. The likelihood of observing viral counts in a sample collected from patient at time is given by a Poisson distribution,

$$p(k = n_p(t) | \eta = N_p(t)) = e^{-\eta} \frac{\eta^k}{k!} \quad (6)$$

with rate parameter set by the model value of the viral multiplicity $N_p(t)$ (Equation 5). We use the Poisson likelihood in Equation 6 to characterize an error model to fit the parameters of the viral dynamics in Equation 5. However, since the mean and variance of the Poisson distribution are related, combining data with different mean values $N_p(t)$ at different times and from different patients can cause inconsistencies in evaluations of errors in our fits. To overcome this problem, we use a variance stabilizing transformation (McCullagh and Nelder, 2019) and define a change in variable $\hat{n}_p(t) = \sqrt{n_p(t)}$. This transformed variable has a constant variance, and in the limit of large-sample size, it is Gaussian distributed with a mean and variance given by, $\hat{n}_p(t) \sim \mathcal{N}(\sqrt{\lambda}, 1/4)$. The constant variance of the transformed variable enables us to combine data from all patients and time points, irrespective of the sample's viral loads, and fit the model parameters (r, x_p) using (non-linear) least-squares fitting of the function

$$R(r, \{x_p\}) = \sum_{p: \text{patients}, t} (\sqrt{N_p(t|r, x_p)} - \sqrt{n_p(t)})^2 \quad (7)$$

Here, $N_p(t|r, x_p)$ is the model estimate of viremia in patient p at time t (Equation 5), given the pre-treatment fraction of resistant variants x_p , and the decay rate r .

Note that the viremia measurements have a minimum sensitivity threshold of 20 RNA copies per ml. We treat the data points below the threshold of detection as missing data and if $n_p(t)$ is below the threshold of detection we impute $n_p(t) = \min(20, N_t)$.

The fitted viremia curves for patients enrolled in the three bNAbs trials under consideration are shown in Figure 3—figure supplements 1–3, and the respective decay rates r for each experiment are,

trial	10-1074	3BNC117	Combination	Avg
fitted r (days ⁻¹)	0.36	0.23	0.33	0.31

(8)

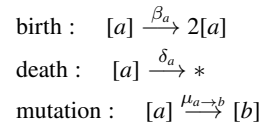
Individual-based model for viral population dynamics

To encode for different viral variants, we specify a coarse-grained phenotypic model, where a viral strain of type a is defined by a binary state vector $\vec{\rho}^a = [\rho_1^a, \dots, \rho_\ell^a]$, with ℓ entries for potentially

escape-mediating epitope sites; the binary entry of the state vector at the epitope site represents the presence ($\rho_i^a = 1$) or absence ($\rho_i^a = 0$) of an escape mediating mutation against a specified bNAb at site i of variant a . We assume that a variant is resistant to a given antibody if at least one of the entries of its corresponding state vector is non-zero.

We define an individual-based stochastic birth-death model (Wilkinson, 2019) to capture the competitive dynamics of different HIV variants within a population. This dynamic model will allow us to predict the distribution of rebound times under any combination of antibodies.

We assume that a viral strain of type a can undergo one of three processes: birth, death and mutation to another type b with rates β_a , δ_a , and $\mu_{a \rightarrow b}$, respectively:



We specify an intrinsic fitness f_a for a given variant a , defined as the growth rate of the virus in the absence of neutralizing antibody or competition. Since bNAbs target highly vulnerable regions of the virus, we expect that HIV-1 escape mutations to be intrinsically deleterious for the virus and to confer a fitness cost relative to the susceptible viral variants prior to the infusion of bNAbs. Assuming that fitness cost of escape is additive across sites and background-independent, we can express the fitness of a variant as $f_a = \bar{f} - \Delta$, where Δ is the cost associated with the presence of an escape mutation at site i of variant a (i.e., $\Delta_i = 1$).

We assume that growth is self-limiting via a competition for host T-cells. This competition enforces a carrying capacity, which sets the steady-state population size N_k . Competition is mediated through a competitive pressure term $\phi = \frac{\sum_a N_a f_a}{N_k}$ which attenuates the net growth rate γ_a so that $\gamma_a = f_a - \phi$. At the carrying capacity, the competitive pressure equals the mean population fitness $\phi = \bar{f}$, making the net growth rate of the population zero.

The net growth rate of a variant a is given by its birth rate minus the death rate: $\gamma_a = \beta_a - \delta_a$. We assume that the total rate of events (i.e. the sum of birth and death events) is equal for all types, that is, $\lambda = \beta_a + \delta_a, \forall a$. Assuming that λ is constant is to be agnostic about the mechanism of a fitness decrease, attributing fitness loss equally to (i) an increase in the death rate and (ii) a decrease in the birth rate.

Because the absolute magnitude of β and δ asymptotically converge in the continuum limit for a surviving population, that is, $\lim_{N \rightarrow \infty} \beta/\delta = 1$, it is impossible to distinguish between (i) and (ii) in the continuous limit. Choosing constant λ simplifies both theoretical calculations and the simulation algorithm.

This leads to the following equations for the birth and the death rates:

$$\beta_i = \frac{\lambda + (f_i - \phi)}{2} \quad \delta_i = \frac{\lambda - (f_i - \phi)}{2} \tag{9}$$

In the presence of an antibody, birth is effectively halted for susceptible variants, resulting in birth and death rate for a susceptible variant s ,

$$\beta_s = 0 \quad \delta_s = r \tag{10}$$

so that the susceptible phenotype decays at rate r .

We assume that mutations occur independently at each site,

$$\mu_{a \rightarrow b} = \begin{cases} \mu_i & \text{if } \rho_i^a - \rho_i^b = 1_i \\ \mu_i^\dagger & \text{if } \rho_i^a - \rho_i^b = -1_s \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where 1_s is the vector which has only one non-zero entry at site s , and μ_i and μ_i^\dagger are the forward the backward mutation rates at site i , respectively. We characterize the state of a population by vector $\mathbf{n} = (n_1, \dots, n_M)$, where n_a is the number of type a variants within the population. We approximate the evolution of the population state distribution using a Fokker-Planck equation (see Appendix 2).

Of special interest for our analysis is the steady-state density of frequencies, which we use to describe the initial state of the population (before treatment) and to infer selection intensity. In the steady state, the population is fluctuating around carrying capacity $\sum_a n_a \approx N_k$ and we can represent the population state via allele frequencies $x_a = n_a/N_k$. In the simple case of a bi-allelic problem, the equilibrium allele frequency distribution $P_{eq}(x)$ follows the Wright-equilibrium distribution (**Crow and Kimura, 2010**) with modified rates,

$$P_{eq}(x) = \frac{1}{Z} \frac{e^{\frac{2N_k}{\lambda}(f_1 - f_0)x} (1-x)^{\frac{2N_k}{\lambda}\mu^\dagger} x^{\frac{2N_k}{\lambda}\mu}}{(1-x)^x} \equiv \frac{1}{Z(\sigma, \theta, \theta^\dagger)} \frac{e^{-\sigma x(1-x)^{\theta^\dagger}} x^\theta}{(1-x)^x} \tag{12}$$

where Z is the normalization factor, f_1 is the intrinsic fitness of the variant of interest, f_0 is the fitness of the competing variant, μ and μ^\dagger are the forward and backward mutation rates, N_k is the carrying capacity, and λ is the total rate of events in the birth-death process, which sets a characteristic time scale over which the impact of selection and mutations can be measured. In this case, we can define an ‘effective population size’ that sets the effective size of a bottleneck and the natural time scale of evolution as $N_e = N_k/\lambda$, and specify a scaled selection factor $\sigma = N_e s = N_e(f_0 - f_1)$, and scaled forward mutation and backward mutation rates (diversity) $\theta = 2N_e\mu$, and $\theta^\dagger = 2N_e\mu^\dagger$. The normalization factor is given by,

$$Z \equiv Z(\sigma, \theta, \theta^\dagger) = \mathcal{B}(\theta, \theta^\dagger) {}_1F_1(\theta, \theta + \theta^\dagger, -\sigma) \tag{13}$$

where ${}_1F_1(\cdot)$ denotes a Kummer confluent hypergeometric function and $\mathcal{B}(\theta, \theta^\dagger) = \frac{\Gamma[\theta]\Gamma[\theta^\dagger]}{\Gamma[\theta + \theta^\dagger]}$ is the Euler beta function.

Extinction Probability

The logistic dynamics describing a patient’s viremia over time in **Equation 5** is the deterministic approximation to the underlying birth-death process. However, the resistant population can also go extinct due stochastic effects, which in turn contribute to the probability of late rebound in a population. To capture this effect, we derive an approximate closed form expression for the probability of extinction.

Using the standard birth-death process generating function theory **Allen, 2010** the probability $P(\text{extinct}|n_i)$ that a population consisting of n_i resistant variants of type i go extinct can be expressed as,

$$P(\text{extinct}|n_i) = \left(\frac{\delta_i}{\beta_i}\right)^{n_i} \tag{14}$$

To characterize the probability of extinction for a population of size N_k with pre-treatment fraction of i^{th} resistant variants x_i , we can convolve the extinction probability in **Equation 14** with a Binomial probability density for sampling n_i resistant variants from N_k trials. Given that the pre-treatment fraction of resistant variants is small $x_i \ll 1$ and N_k is large, this Binomial distribution can be well approximated by a Poisson distribution, $\text{Pois}(n_i; N_k x_i)$ with rate $N_k x_i$, resulting in an extinction probability,

$$\begin{aligned} P(\text{extinct}|x_i) &= \sum_{n_i} \text{Pois}(n_i; N_k x_i) \left(\frac{\delta_i}{\beta_i}\right)^{n_i} \\ &= \exp(-N_k x_i) \sum_{n_i} \frac{(N_k x_i)^{n_i}}{n_i!} \left(\frac{\delta_i}{\beta_i}\right)^{n_i} \\ &= \exp\left(-N_k \frac{\beta_i - \delta_i}{\beta_i} x_i\right) \end{aligned} \tag{15}$$

Using the expressions for the growth in the absence of competition, $\beta_i - \delta_i = \gamma_i = f_i$ (since $\phi = 0$), and assuming that fitness is small relative to the total rate of birth and death events $f_i \ll \lambda$, we can use the approximation $\beta_i = (\lambda + f_i)/2 \approx \lambda/2$, to arrive at,

$$P(\text{extinct}|x_i) \approx \exp\left(\frac{-2N_k f_i}{\lambda} x_i\right) = \exp\left(-\frac{x_i}{x_{\text{ext}}}\right) \tag{16}$$

x_{ext} where the characteristic escape threshold can be written in terms of concrete genetic observables,

$$x_{\text{ext}} \equiv \frac{\lambda}{2N_k f_i} = \frac{\mu_{\text{ts}}}{f_i} \theta_{\text{ts}}^{-1} \tag{17}$$

Figure 3 shows that this threshold can well separate the fate of stochastic evolutionary trajectories, simulated with relevant parameters for intra-patient HIV evolution.

Numerical simulations of the birth-death process

To treat the full viral dynamics including mutations, and transient competition effects, we can exactly simulate the viral dynamics defined by our individual based model. Below are the key steps in this simulation.

Population initialization

At the starting point, we set the population size (i.e. the carrying capacity in the simulations) N_k as a free parameter chosen to be large enough to make discretization effects small. The population is then evolved through time using an exact stochastic sampling procedure (the Gillespie algorithm Gillespie (1977)). Simulating the outcome of this stochastic evolution generates the distribution of rebound times and the probability of late rebound—the key quantities related to treatment efficacy.

The input to our procedure is a list of antibodies for which we specify (i) the escape mediating sites for each antibody, and the (invariant) quantities describing (ii) the site-specific cost of escape $\frac{\sigma}{\theta_{ts}}$, and (iii) the forward and backward mutation rates $(\frac{\theta}{\theta_{ts}}, \frac{\theta^\dagger}{\theta_{ts}})$. To simulate the trial outcome for each patient, we use the neutral population diversity θ_{ts} directly inferred from the patients; see Inference of mutation rates and the neutral diversity within a population. From this, we construct the list of L site parameters (concatenated across all antibodies) for selection and diversity: $\sigma_{1:L}, \theta_{1:L}^\dagger$.

We assume that at the start of the simulation, populations are in the steady state and that the potential escape sites are at linkage equilibrium. The approximate linkage equilibrium assumption is justified since the distance between these escape sites along the HIV-1 genome is greater than the characteristic recombination length scale $\approx 100\text{bp}$ of the virus (Zanini et al., 2015). As a result, we draw an independent frequency x_i from the stationary distribution $P_{\text{eq}}(x|\sigma_i, \theta_i, \theta_i^\dagger)$ in Equation 12 to describe the state of a give site i , and use these frequencies to construct the initial viral genotypes ρ^v for each virus v in our initial population; see Algorithm 1 (Appendix 5). In simulations, we show that this assumption does not bias our results even when θ_{ts} is fluctuating and recombination is absent (Section 6.1 and Figure 4—figure supplement 2).

To sample from the stationary distribution itself, we define a novel Gibbs-sampling procedure (Geman and Geman, 1984) for generating the allele frequencies of the escape variants for the initial state of the population $x \sim P_{\text{eq}}(x|\sigma, \theta, \theta^\dagger)$ (Equation 12). To characterize this procedure, we expand the exponential selection factor $e^{\sigma(1-x)}$ in the original distribution, which results in,

$$\begin{aligned} P_{\text{eq}}(x|\sigma, \theta, \theta^\dagger) &= \frac{e^{-\sigma}}{Z(\sigma, \theta, \theta^\dagger)} e^{\sigma(1-x)} \frac{x^\theta (1-x)^{\theta^\dagger}}{x(1-x)} \\ &= \sum_{k=0}^{\infty} \frac{e^{-\sigma}}{Z(\sigma, \theta, \theta^\dagger)} \frac{\sigma^k}{k!} x^\theta (1-x)^{\theta^\dagger+k} \\ &\equiv \sum_{k=0}^{\infty} Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger) \end{aligned} \tag{18}$$

Here, $Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)$ is a joint distribution over (x, k) , and the desired distribution over the allele frequency x can be achieved by marginalizing the joint distribution over the discrete variable k . We can also express the conditional distributions for x and k as,

$$Q_{\text{eq}}(x|k, \sigma, \theta, \theta^\dagger) = \frac{Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)}{\int dx Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)} = \text{Beta}(x; \theta, \theta^\dagger + k) \tag{19}$$

$$Q_{\text{eq}}(k|x, \sigma, \theta, \theta^\dagger) = \frac{Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)}{\sum_k Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)} = \text{Poisson}(k; (1-x)\sigma) \tag{20}$$

We use these conditional distributions to define a joint Gibbs sampler for Q_{eq} . We summarize the resulting $(x, k) \sim Q_{\text{eq}}(x, k|\sigma, \theta, \theta^\dagger)$ in the joint Gibbs sampler in Algorithm 2 (Appendix 5). This chain mixes extremely quickly and avoids calculation of the hypergeometric function for the normalization factor (Equation 20), which is computationally costly; see Algorithm 2 (Appendix 5).

Simulation of the evolutionary process

We use a Gillespie algorithm to simulate the evolutionary process, where we break up the reaction calculation into two parts: randomly choosing a viral strain ρ_i from the population and then determining whether it reproduces or dies based on its fitness f_i and escape status; see Algorithm 3 (Appendix 5).

Determining the simulation parameters of the birth-death process from genetic data

We set the intrinsic growth rate (fitness) of the wild-type virus, in the absence of competition to be $\gamma = (3 \text{ days})^{-1}$, consistent with intra-patient doubling time of the virus (García et al., 2001; Ioannidis et al., 2000; García et al., 1999). We infer the neutralization rate r by fitting the viremia curves (Figure 3) in the trials under study, and use the averaged decay rate $r = 0.31$ for simulations, fitted using Equation 8. For the absolute mutation rate μ_{ts} (per nucleotide per day), we use 1.1×10^{-5} which is the average of the reported values for transitions per site per day from Zanini et al., 2017. Using the covariance of neutral diversity in twofold and four-fold synonymous sites, we determine the transition/transversion diversity ratio to be $\theta_{ts}/\theta_{tv} = 7.8$ (Figure 2—figure supplement 1). We use these values to determine the forward and backward mutation rates μ_s and μ_s^\dagger for each site (see Inference of mutational target size for each bNAb).

Generally, the trial patients show viral populations with larger neutral diversity at the start of the trial compared to the patients enrolled in the high-throughput study of Zanini et al., 2015 (Figure 2). We account for differences in the genetic makeup of the patients enrolled in the trial by directly estimating the neutral diversity θ_{ts} from the synonymous site-frequency data of patients, before the start of the trial. The estimated viral diversity θ_{ts} , coupled with the mutation rate $\mu = 1.1 \times 10^{-5} / \text{day} / \text{nt}$, and the total rate of birth and death events in the viral population λ , set the carrying capacity N_k for a given individual,

$$N_k = \frac{\lambda}{2\mu_{ts}} \theta_{ts}. \quad (21)$$

It should be noted that the value of N_k , as the number of viruses in our individual-based simulations, is not related to the maximal viral load in the viremia measurements (i.e. steady state copy number per ml) as this relationship depends on the microscopic details of the population dynamics.

The total rate of birth and death events $\lambda = \beta + \delta$ tunes the amount of stochasticity, that is, more events cause noisier dynamics. Notably, stochasticity can be linked to the size of the population N_k , which is directly coupled to λ (Equation 21). We set the value of λ self-consistently by requiring the minimum frequency of a variant in our simulations $x_{\min} = 1/N_k = \frac{2\mu_{ts}}{\lambda} \theta_{ts}^{-1}$ to be smaller than the escape threshold $x_{\text{ext}} = \frac{\mu_{ts}}{\gamma} \theta_{ts}^{-1}$ due to stochasticity (Equation 17). We set $\lambda = 2 \text{day}^{-1}$ so that $x_{\min} = \frac{1}{3} x_{\text{ext}}$. Increasing λ results in an increase in the size of population N_k in our simulations, which is computationally costly, without qualitatively changing the statistics of the rebound trajectories.

Inference of mutation rates and the neutral diversity within a population

Previous work has indicated an order of magnitude difference between the rate of transitions (mutations within a nucleotide class) and transversions (out-class mutations) in HIV (Nielsen, 2006; Zanini et al., 2017; Theys et al., 2018; Feder et al., 2017). Therefore, to infer the neutral diversity parameter θ_{ts} , we also account for the differences between transition and transversion rates.

Consider the set of sequences sampled from a patient's viral population at a particular time. Two neutral alleles that are linked by a symmetric mutational process $\mu_{1 \rightarrow 2} = \mu_{2 \rightarrow 1}$ have a simple count likelihood. The probability to see allele 1 with multiplicity n and allele 2 with multiplicity m is given by a binomial distribution $\text{Binom}(n, mx)$ with parameter x denoting the probability for occurrence of allele 1, convolved with the neutral biallelic frequency distribution $P_{\text{eq}}(x|\sigma = 0, \theta)$ from Equation 12. Using this probability distribution, we can evaluate the log-likelihood $\mathcal{L}(\theta|n, m)$ for the neutral diversity θ given the observations (n, m) for the multiplicities of the two alleles in the population,

$$\mathcal{L}(\theta|n, m) = \log \int dx \text{Binom}(n, mx) \times P_{\text{eq}}(x|\sigma = 0, \theta) = \log \int dx \binom{n+m}{n} x^n (1-x)^m \times \frac{x^{\theta-1} (1-x)^{\theta-1}}{Z(\theta)} \quad (22)$$

Inference of netral diversity, mutational target size, and selection from genetic data

To estimate the transition diversity, we only use twofold synonymous sites, and treat each site independently but with a shared diversity parameter θ_{ts} . For example, consider neutral variations for two amino acids glutamine and phenylalanine. The third position in a codon for both of these amino acids are twofold synonymous, as the two possible codons for glutamine are CAG and CAA, and for phenylalanine are TTT, and TTC. Now consider that in the data a conserved glutamine has $n = 3$ G's and $m = 97$ A's in the third codon position, and a conserved phenylalanine has $n = 10$ T's and $m = 90$ C's at its third codon position. In this case, the combined log-likelihood for the shared diversity parameter is $\mathcal{L}(\theta_{data}) = \mathcal{L}(\theta_{ts}|3, 97) + \mathcal{L}(\theta_{ts}|10, 90)$. Extending to all of sites in the *env* protein, the maximum-likelihood estimator for the transition diversity θ_{ts} can be evaluated by maximizing the likelihood summed over all conserved twofold synonymous sites,

$$\theta_{ts}^* = \arg \max_{\theta_{ts}} \sum_{\text{two-fold sites}} \mathcal{L}(\theta_{ts}|n = n_A + n_T; m = n_G + n_C) \quad (23)$$

In **Figure 2—figure supplement 1** (panel A), we show that the maximum likelihood estimation method described above has better properties than the more commonly used estimator of the variance $\overline{x(1-x)}$ [Stoddart and Taylor, 1988](#).

In a similar way, the likelihood for the transversion θ_{tv} is determined from polymorphic data at all conserved four-fold synonymous sites. One such example is the third position in a glycine codon, where (GGT, GGC, GGA, GGG) translate to the same amino acid. The maximum-likelihood estimator for the transversions is

$$\theta_{tv}^* = \arg \max_{\theta_{tv}} \sum_{\text{four-fold sites}} \mathcal{L}(2\theta_{tv}|n = n_G + n_T; m = n_C + n_A) \quad (24)$$

The factor of 2 in the argument of the likelihood accounts for the multiplicity of mutational pathways, e.g. from a *G* nucleotide there are two transversion possibilities, $G \rightarrow C$ and $G \rightarrow A$ for moving from one allele to the other [Kimura, 1981](#).

Using this likelihood approach, we can infer the neutral diversities θ_{ts}^* and θ_{tv}^* for each patient at each time point from the polymorphism in twofold and fourfold synonymous sites. To characterize the ratio of transition to transversion rates, we use linear regression on the entire patient population and sample history and infer a constant ratio $\mu_{ts}/\mu_{tv} = \theta_{ts}/\theta_{tv} = 7.8$ (**Figure 2—figure supplement 1B**). In **Figure 2—figure supplement 1**, we also show that the estimate for this ratio is relatively consistent across different data sources, produced even by different sequencing technologies. The previously reported relative rate of transitions to transversions, based on the estimates of sequence divergence along phylogenetic trees of HIV-1 is $\mu_{ts}/\mu_{tv} = 5.6$ ([Zanini et al., 2017](#)), which is similar to our maximum likelihood estimate.

Inference of mutational target size for each bNAb

The nucleotide triplets which encode for amino acids at an escape site undergo substitutions which can change the amino acid type and create an escape variant. The changes in the state of an amino acid codon can be modeled as a Markov jump process and can be visualized as a weighted graph where the nodes represent codon states, and edges represent single nucleotide substitutions linking two codon states (**Figure 1D**). In our mutational model, these edges have weights associated with either the mutation rates for transitions μ_{ts} or transversions μ_{tv} . We call this the *codon substitution graph*.

The codon states can be clustered into three distinct classes: (i) codons which are fatal *F*, (ii) wild-type (i.e. susceptible to neutralization by the bNAb) *W*, and escape mutants (i.e. resistant to the bNAb) *M*. We expect the escape mutants to be at a selective disadvantage compared to the resistant wild-type, and that the most common escape codons to be those which are adjacent to wild-type states.

The mutational target size is determined by the density of paths from the wild-type *W* to the escape mutants *M*.

$$\mu = \frac{1}{|W|} \sum_{c \in W; d \in M} [c - d = ts] \mu_{ts} + [c - d = tv] \mu_{tv} \quad (25)$$

$$\mu^\dagger = \frac{1}{|M|} \sum_{c \in W; d \in M} [c - d = ts]\mu_{ts} + [c - d = tv]\mu_{tv} \tag{26}$$

The functions $[c - d = ts]$ or $[c - d = tv]$ are 1 when the two codons are separated by a transition or transversion, and are zero otherwise. Note that since we only have an estimate for the ratio of the transition to transversion rates μ_{tv}/μ_{ts} , we can only determine the scaled mutational target sizes, $\hat{\mu} = \mu/\mu_{ts}$ and $\hat{\mu}^\dagger = \mu^\dagger/\mu_{ts}$, which are sufficient for inference of selection in the next section. The full list of mutational target sizes inferred for the bNABs in this study are shown in **Appendix 1—table 1**.

When discussing the mutational target size of escape from a given bNAB, we refer to the total mutation rate from the susceptible (wild type) to the escape variant as,

$$\mu = \sum_{i \in \text{esc. sites}} \mu_{s \rightarrow i} \tag{27}$$

where the sum runs over all the mediating escape sites, and can be interpreted as the average number of accessible escape variants. In the strong selection regime, we can write the frequency of escape mutants as,

$$x_{\text{mut}} \approx \sum_i x_i = \sum_i \frac{\mu_{s \rightarrow i}}{\Delta_i} \approx \frac{\mu}{\Delta_{\text{hm}}} \tag{28}$$

Inference of selection for escape mutations against each bNAB

Here, we develop an approximate likelihood approach to infer the selection ratio $\hat{\sigma} = \frac{\sigma}{\theta_{ts}}$, using the high-throughput sequence data from bNAB-naive HIV-1 patients from **Zanini et al., 2017**. The quantity $\hat{\sigma}$ is a dimensionless ratio which is independent of the coalescence timescale N_e , and therefore, represents a stable target for inference.

We assume that the probability to sample m escape mutants and s susceptible (wild-type) alleles at a given site in the genome of HIV in a population sampled from a patient at a given time point follows a binomial distribution $\text{Binom}(m, sx)$, governed by the underlying frequency x of the mutant allele. In addition, the frequency x of the allele of interest itself is drawn from the equilibrium distribution $P_{\text{eq}}(x|\sigma, \theta, \theta^\dagger)$ (**Equation 12**), governed by the diversity θ_{ts} inferred from the neutral sites, the estimated mutational target sizes $\hat{\mu} = \mu/\mu_{ts}$, $\hat{\mu}^\dagger = \mu^\dagger/\mu_{ts}$, and the unknown selection ratio $\hat{\sigma} = \sigma/\theta_{ts}$. As a result, we can characterize the probability $P(m, s|\theta_{ts}, \hat{\mu}, \hat{\mu}^\dagger, \hat{\sigma})$ to sample m escape mutants and s susceptible-type alleles, given the scaled selection and diversity parameters as,

$$\begin{aligned} P(m, s|\theta_{ts}, \hat{\sigma}) &= P(m, s|\sigma = \hat{\sigma}\theta_{ts}, \theta = \hat{\mu}\theta_{ts}, \theta^\dagger = \hat{\mu}^\dagger\theta_{ts}) \\ &= \int dx \text{Binom}(m, sx)P_{\text{eq}}(x|\sigma, \theta, \theta^\dagger) \\ &= \frac{1}{Z(\sigma, \theta, \theta^\dagger)} \binom{s+m}{s} \int dx \frac{e^{-\sigma x} x^{\theta+m} (1-x)^{\theta^\dagger+s}}{x(1-x)} \end{aligned} \tag{29}$$

Here, $Z(\sigma, \theta, \theta^\dagger)$ is a confluent hypergeometric function of the model parameters that sets the normalization factor for the allele frequency distribution $P_{\text{eq}}(x)$ (**Equation 13**). It should be noted that the viral population is in fact out of equilibrium, due to constant changes in immune pressure evolution from the B-cell and T-cell populations (**Nourmohammad et al., 2016**). Although we are ignoring these significant complications, we later use the same equilibrium distribution in a consistent way to generate standing variation in simulations. For the model to make accurate predictions, it is not necessary that the equilibrium model be exactly correct, but only that it is rich enough to provide a consistent description for the distribution of mutant frequencies observed across viral populations.

We will use the probability density in **Equation 29** to define a log-likelihood function in order to infer the scaled selection $\hat{\sigma} = \sigma/\theta_{ts}$ from data. To do so, we first express the logarithm of this probability density as,

$$\begin{aligned} \log P(m, s|\theta_{ts}, \hat{\sigma}) &= \log P(m, s|\sigma = \hat{\sigma}\theta_{ts}, \theta = \hat{\mu}\theta_{ts}, \theta^\dagger = \hat{\mu}^\dagger\theta_{ts}) \\ &= \log Z(\sigma, \theta + m, \theta^\dagger + s) - \log Z(\sigma, \theta, \theta^\dagger) + \text{const.} \\ &= \log \mathbb{E} [e^{-\sigma x}]_{\text{Beta}(\theta+m, \theta^\dagger+s)} - \log \mathbb{E} [e^{-\sigma x}]_{\text{Beta}(\theta, \theta^\dagger)} + \text{const.} \end{aligned} \tag{30}$$

where the constant factors (const.) are independent of selection, and $\mathbb{E}[\cdot]_{\text{Beta}(\cdot)}$ denotes the expectation of the argument over a Beta distribution with parameters specified in the subscript. The expression in **Equation 30** implies that we can evaluate the likelihood of selection strength by computing the difference between the logarithms of the expectation for $e^{-\sigma x}$ over allele frequencies drawn from two neutral distributions (Beta distributions), with parameters (θ, θ^\dagger) and $(\theta + m, \theta^\dagger + s)$, respectively. This approach is more attractive as it would not require direct evaluation of the confluent hypergeometric functions for the normalization factors in **Equation 29**. Estimating these normalization factors is computationally intensive for large values of σ , since many terms in the underlying hypergeometric series should be taken into account to stably compute them. However, evaluating the expectations via sampling from these two neutral distributions has the disadvantage that it is subject to variations across simulations. We reduce the variance of our estimate of $\log P(m, s|\sigma, \theta, \theta^\dagger)$ in **Equation 30** by using a mixture-importance sampling scheme (**Owen and Zhou, 2000**) with the details shown in Algorithm 4 (Appendix 5).

We use data collected across all time points and from all patients to infer reliable estimates for selection strengths. However, allele frequencies are correlated across time points within patients (**Figure 2B**), and thus, sequential measurements are not independent data points. Our estimates indicate a coalescence time of about $N_e \sim 10^3$ days based on the estimates for the mutation rate $\mu_{\text{ts}} = 10^{-5}$ /nt/day, and the neutral diversity $\theta_{\text{ts}} = 2N_e\mu_{\text{ts}} = 0.01$. This coalescence time is much longer than the typical separation between sampled time points within a patient ($\sim 10^2$ days), suggesting that sequential samples collected from each individual in this data are correlated. Therefore, we treat each patient as effectively a single observation, using the time-averaged likelihood for the (scaled) selection factor $\hat{\sigma}$:

$$\mathcal{L}(\hat{\sigma}) = \sum_p \frac{1}{T_p} \sum_t \log P(m_t, s_t|\theta_{\text{ts}}(t), \hat{\sigma}) \quad (31)$$

where p and t denote patient identity and sampled time points, respectively, and T_p is the total number of time points sampled in patient p . We use the likelihood in **Equation 31** to generate samples from the posterior distribution for selection strengths under a flat prior, with a standard Metropolis-Hastings algorithm (**Hastings, 1970**). Since the prior is constant, this procedure amounts to simply accepting or rejecting samples based on the likelihood ratio of **Equation 31**. We used the centered-normal distribution with standard deviation of 50 ($\times \mu_{\text{ts}}$ in absolute units) as the proposal density for the jumps in the Markov chain.

Prior work has also inferred the fitness effect of mutations in HIV, but our approach differs in important aspects. For instance, maximum entropy models have been used to infer the preference of different amino acids in the Gag and the env proteins of HIV-1 from their prevalences across sequences sampled from different patients (**Louie et al., 2018; Ferguson et al., 2013**). The inferred preference values can explain the in-vitro growth rate (fitness) of the associated viral strains, especially for sites that are relatively conserved and are not the drivers of antigenic evolution in HIV-1. However, further modeling would be needed to quantitatively map these inferred amino acid preferences onto population genetics measures that can be used to characterize the evolutionary dynamics of an HIV population. Other work has used longitudinal HIV-1 sequence data to infer selection and to characterize the role of selective sweeps due to immune pressure, genetic hitchhiking and recombination in the turnover of HIV-1 populations within patients (**Zanini et al., 2017; Neher and Leitner, 2010; Illingworth et al., 2020; Haddox et al., 2018**). In contrast, our work focuses on the expected composition of the population in a viremic patient prior to the start of treatment as opposed to the history of a viral population. Our approach uses a self-consistent formalism for inference of the population genetics parameters (e.g. population diversity and selection strength) and for the evolutionary simulations used to predict outcomes. Therefore, we can directly interpret the fitted parameters in terms of both the viral dynamics and the pre-treatment state of HIV-1 populations within patients.

Predicting trial outcomes from genetically informed evolutionary models

Predicting rebound times

We expect different distributions of patient outcomes depending on whether they have been recently infected and thus have relatively low viral diversity, or whether their infection is longstanding with

a diverse viral population. To construct the distribution of initial population diversities θ_{ts} for simulating trial outcomes, we apply the θ_{ts} inference procedure (Equation 22) to pre-treatment sequence datasets available from the clinical trials under consideration. In Figure 2D this set of pre-trial θ_{ts} is compared to the longitudinal in-patient θ_{ts} from Zanini et al., 2017. We used random draws from the inferred θ_{ts} values for patients to generate θ_{ts} for simulations.

We found that there was considerably more viral escape and non-responders in our simulations than in the observed data as shown in Figure 4—figure supplement 3A. This is in addition to the fact that the patients were screened to have only susceptible variants to the antibodies used in trials (Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018). In theory, there should be zero non-responders, as such patients should have been excluded by screening. The over-prediction of both non-responders and late rebounds is a signature of undercounting the effective diversity of the viral populations.

The failure of both screening and our naive prediction in undercounting the diversity in the viral population can be explained by an effective viral reservoir. Viral variants which mediate rebound can come from compartments such as including bone marrow, lymph nodes, and organ tissues, and can be genetically distinct from those sample from the plasma T-cells during screening (Chaillon et al., 2020; Wong and Yukl, 2016; Chun et al., 2005). This reservoir of viral diversity can reappear in plasma after infusion of a bNAb and could in part contribute to treatment failure (Avettand-Fenoel et al., 2007; Shan and Siliciano, 2013; Sharkey et al., 2011; Tagarro et al., 2018).

Determining patient diversity enhancement due to latent reservoirs

We model the effect of reservoirs as a simple inflation of the diversity observed by a multiplicative factor ξ . We fit ξ directly to trial observations, using a disparity-based approach by minimizing an empirical divergence estimator Ekström, 2008 between the observed and simulated data. To do so, we characterize the Hellinger distance Lindsay, 1994; Simpson, 1987 between the true distribution of rebound times $P(t)$ and the rebound times $Q(t|\xi)$ generated by simulations with a given reservoir factor ξ ,

$$D_H(P(t)||Q(t|\xi)) = \int dt (Q(t|\xi)^{\frac{1}{2}} - P(t)^{\frac{1}{2}})^2 \approx \sum_{(i)}^{n_q} (Q_{(i)}(\xi)^{\frac{1}{2}} - P_{(i)}^{\frac{1}{2}})^2 \tag{32}$$

Algorithm 5 (Appendix 5) defines the procedure that we use to estimate the Hellinger distance $D_H(Q(t)||P(t|\xi))$. Specifically, we use n_q quantiles of the observed data $x_{(i)} \sim \mathbf{Q}$ to partition the space of observations into discrete outcomes

$$Q_{(i)}(\xi) = \int dt P(t|\xi) [t_{(i)} \leq t < t_{(i+1)}] P_{(i)} = \frac{1}{n_q} \tag{33}$$

where $P_{(i)}$ is a constant by construction, and $P_{(i)}(\xi)$ is estimated by simulations, and $[\cdot]$ is the Iverson bracket Graham et al., 1989; see Algorithm 5 (Appendix 5).

$$[B] = \begin{cases} 1 & \text{if } B = \text{true} \\ 0 & \text{if } B = \text{false} \end{cases} \tag{34}$$

To simulate data for this analysis, we generate S rebound times $(T_{1:S})$ by simulations, given the scaled diversity values $\xi\theta_{ts}$. We then find the optimal value ξ^* by minimizing the disparity with the observed rebound times $t_{1:p}$ by brute-force search,

$$R(\xi|t_{1:p}) = \sum_{\text{trials}} \text{ReboundDisparity}(t_{1:p}, T_{1:S}|\xi) \tag{35}$$

$$\xi^* = \arg \min_{\xi} R(\xi|t_{1:p}) \tag{36}$$

Here, ReboundDisparity is the function defined by Algorithm 5 (Appendix 5); see Ekström, 2008 for details. We find the optimal reservoir factor to be $\xi^* = 2.1$, which we use in subsequent therapy prediction. The disparity over various values of ξ for different trials is shown in Figure 4—figure supplement 3.

Simulating outcomes of clinal trials

Given the reservoir-corrected estimate of the diversity $\xi\theta$ and the posterior samples for selection factors $\hat{\sigma}$, we now summarize how we simulated the outcome of clinical trials.

For a given bNAb, we draw the selection factor at each of the escape mediating sites from the corresponding Bayesian posterior on $\hat{\sigma}$; the posterior distributions are shown in **Figure 4A**, and summarized in **Appendix 1—table 1**. We also use the mutational target size (forward μ and backward μ^\dagger rates) associated with each of the escape mediating sites of a given bNAb; see **Appendix 1—table 1**. The result can be summarized in a mutation / selection matrix \hat{M}_a for a given bNAb a , where each column corresponds to an escape mediating site against the bNAb,

$$\hat{M}_a = \begin{pmatrix} \hat{\mu}_1 & \dots & \hat{\mu}_i \\ \hat{\mu}_1^\dagger & \dots & \hat{\mu}_i^\dagger \\ \hat{\sigma}_1 & \dots & \hat{\sigma}_i \end{pmatrix} \quad (37)$$

The elements of the matrix \hat{M}_a are the scaled mutation and selection factors, i.e., $\hat{\mu}_i = \mu_i/\mu_{ts}$, $\hat{\mu}_i^\dagger = \mu_i^\dagger/\mu_{ts}$, and $\hat{\sigma}_i = \sigma_i/\theta_{ts}$, where the absolute value of mutation rate is set to $\mu_{ts} = 1.11 \times 10^{-5}$ /nt/day from **Zanini et al., 2017**.

For each patient in our simulated trial, we then draw diversity θ_{ts} from the patient pool, and scale it by our fitted $\xi = 2.1$, resulting in patient-specific selection and mutation factors,

$$\xi \times \theta_{ts} \times \hat{M}_a = \begin{pmatrix} \theta_1 \dots \theta_i \\ \theta_1^\dagger \dots \theta_i^\dagger \\ \sigma_1 \dots \sigma_i \end{pmatrix} \mu_{ts} \times \hat{M}_a = \begin{pmatrix} \mu_1 \dots \mu_i \\ \mu_1^\dagger \dots \mu_i^\dagger \\ \Delta_1 \dots \Delta_i \end{pmatrix} \quad (38)$$

These parameters are then used to initialize the state of an HIV population within a patient according to (Appendix 5), and to determine the absolute rates in Algorithm 3 (Appendix 5) for the population evolution according to **Equation 38**. The decay rate is set to the fitted trial average of $r = 0.31 \text{ days}^{-1}$ (**Equation 8**). The carrying capacity N_k is set according to **Equation 21**. This determines all parameters of the birth-death process simulating the intra-patient evolution of HIV, which are used in Algorithm 3 (Appendix 5).

We evolve a population through time until 56 days have elapsed since treatment, or until the escape fraction relative to the carrying capacity x_t is above 0.8; see Algorithm 3 (Appendix 5). After $x_t > .8$ the evolution is governed by the deterministic equations, and the stochastic simulation ends. The rebound time T , defined as the intersection of the exponential envelope and the carrying capacity, can then be calculated analytically as,

$$T = \frac{1}{\gamma} \log \left(1 + \exp(\gamma t) \frac{1-x_t}{x_t} \right). \quad (39)$$

The resulting distribution for rebound times are shown as model predictions in **Figure 3B–D**.

Rebound times generated in this fashion were also used to estimate the probability of late rebound to characterize the efficacy of a given bNAb in curbing viral rebound. The probability of late rebound was estimated from 10^4 simulated patients. The interdecile quantiles (0.1–0.9) of early rebound (t_{56} days) probability over 200 values of scaled selection coefficients $\hat{\sigma}$ drawn from the posteriors in **Figure 4A** are shown in **Figure 4C**.

Model robustness

Effect of genomic linkage on the inference of selection

In our inference of selection (**Equation 31**, **Figure 4A**), we assume that the escape-mediating sites are at linkage equilibrium and that the distribution of allele frequencies can be approximated by a skewed Beta distribution (**Equation 12**), reflecting the equilibrium of allele frequencies. In reality, despite recombination, the HIV genome exhibits linkage effects, especially at nearby sites **Zanini et al., 2015**, and the viral populations experience changing selective pressures by the immune system **Feder et al., 2021**; **Theys et al., 2018**; **Nourmohammad et al., 2019**, and the transient population bottlenecks during therapy **Feder et al., 2016**.

To test the limits on the validity of our inference procedure, we applied it to in silico populations generated by full-genome forward-time simulations (Appendix 5, Algorithm 3) in the presence and

absence of recombination. To do so, we considered an ensemble of ten patients with 100 genomes sampled at 10 time points, and used two diversity parameters $\theta_{ts} = 0.01$ and $\theta_{ts} = 0.1$, to cover the range reflected in patient data (**Figure 2D**, **Figure 4—figure supplement 2**).

One relevant scenario to consider is the impact of other selected sites in the genome on the distribution of alleles at the escape mediating sites against bNAbs. The sites under a strong constant selection are likely to be already fixed (or at a high a frequency) at their favorable state in the population. However, the strong selection on a large fraction of antigenic sites can be thought as time-varying, due to the changing pressure imposed by the immune system or therapy. To capture this effect, we simulated whole genome evolution in which linked sites were under strong selection ($0.1 \times$ growth rate), and where the sign of selection changed after exponentially distributed waiting times (i.e. as a Poisson process); this model of fluctuating selection has been used in the context of influenza evolution **Strelkova and Lässig, 2012**, and for somatic evolution of B-cell repertoires in HIV patients **Nourmohammad et al., 2019**. The resulted evolutionary dynamics in this case can involve strong selective sweeps and clonal interference due to the continuous rise of beneficial mutations (in the linked sites) within a population.

To test the robustness of our selection inference, we evaluated the distribution of maximum likelihood estimates (MLEs) for the selection values $\hat{\sigma} = \sigma/\theta_{ts}$ at the escape mediating sites, inferred from the ensemble of sequences obtain from simulated data with linkage. **Figure 4—figure supplement 2** shows that even for fully linked genomes (zero recombination) our MLE estimate of selection has little bias relative to the true values used in the simulations. Adding recombination into the simulations only further attenuates the effect of linkage (**Figure 4—figure supplement 2**), making the estimates more accurate.

The reason that selective sweeps of linked beneficial mutations have only minor effects on our inference of selection for the escape mediating sites is two-fold: First, the primary mechanism by which selective sweeps change the strength of selection at linked sites is via a reduction in the effective population size **Hill and Robertson, 2009; Comeron et al., 2008**. However, variations in the effective population size are already accounted for in our inference procedure: The selection likelihood in **Equation 31** is conditioned on the measured neutral-site diversity, $\theta = 2N_e\mu$. The change in the effective population size impacts the selection coefficient $\sigma = 2N_e\Delta$ and the diversity $\theta = 2N_e\mu$ in the same way, and therefore, the (scaled) selection parameter $\hat{\sigma} = \sigma/\theta$ that we infer from data remains insensitive to changes in the effective population size.

The second reason for the robustness of our selection inference to linkage is due to the fact that a beneficial allele in a linked locus can appear on a genetic background with or without a susceptible variant, leading to the rise of either variants in the population. As a results, the impact of such hitchhiking remains a secondary issue in inference of selection at the escape sites, for which an ensemble of populations from different patients with distinct evolutionary histories of HIV-1 are used.

Robustness of selection inference to intra-patient temporal correlations of HIV alleles

To infer the selection effect of mutations from the longitudinal deep sequencing data of **Zanini et al., 2015**, we use time averaging of the likelihood (**Equation 31**) to avoid conflating our results due to temporal correlations between the circulating alleles within patients (**Figure 2B**). We can view this choice as being one choice among two extremes: (i) to treat each patient as effectively one independent data point so that all patients are given the same weight or (ii) to treat each time point as independent, giving patients with more time points a higher weight. These two choices correspond to different log-likelihood functions for the (scaled) selection factor $\hat{\sigma}$:

$$\mathcal{L}(\hat{\sigma}) = \begin{cases} \sum_{p,t} \log P_p(m_t, s_t | \theta_{ts}(t), \hat{\sigma}) & (t\text{-independent}) \\ \sum_p \frac{1}{T_p} \sum_t \log P_p(m_t, s_t | \theta_{ts}(t), \hat{\sigma}) & (t\text{-averaged}) \end{cases} \quad (40)$$

where T_p is the total number of time points from patient p , and $P_p(m_t, s_t | \theta_{ts}(t), \hat{\sigma})$ is the probability to observe m escape mutants, and s susceptible variants at time t in patient p , given the neutral diversity $\theta_{ts}(t)$ and the scaled selection factor $\hat{\sigma}$.

We find that both of these approaches result in similar posteriors for selection $\hat{\sigma}$ (Model robustness **Figure 4—figure supplement 1A** and C), although the t -averaged likelihood has a higher uncertainty due to fewer independent time points. Thus, our inference of selection is insensitive to the exact choice of the likelihood function given in **Equation 40**, yet our time-averaged approach remains the more conservative choice between the two.

Statistical significance of the reservoir-corrected diversity

In **Equation 36**, we introduced the reservoir factor $\xi^* = 2.1$ to account for the diversity of HIV-1 that is not sampled from a patient's plasma prior to therapy, which resulted in a better fit of the rebound time distributions (**Figure 3**) compared to a reservoir-free model (**Figure 4—figure supplement 3A**). Here, we quantify the importance of the reservoir factor with a statistical test on the null hypothesis, $\xi_0 = 1$. Specifically, we perform a hypothesis test to test the necessity of using an inflated diversity ξ_{ts} relative to using the bare diversity observed in pre-trial sequence data θ_{ts} . To do so, we construct a disparity-based test statistic **Ekström, 2008**, which is analogous to the likelihood ratio test statistic.

Recall that the optimal reservoir factor $\xi^* = \arg \min_{\xi} R(\xi|t_{1:p})$ was obtained by minimizing the disparity function $R(\xi|t_{1:p})$ across measurements of rebound times $t_{1:p}$ from all the p patients in data (**Equation 36**). We can estimate the test statistic for the reduction in disparity between the null hypothesis, $\xi_0 = 1$ and the fitted reservoir factor ξ^* as,

$$\Delta_R(t_{1:p}) = R(\xi_0|t_{1:p}) - \min_{\xi} R(\xi|t_{1:p}). \quad (41)$$

We can then determine the p-value by estimating the quantile of the observed test statistic $\Delta_R(t_{1:p})$ relative to that inferred from the distribution of $\Delta_R(T_{1:p})$ obtained from simulations under null hypothesis $\xi_0 = 1$ (**Fisher, 1956**). Specifically,

$$\text{p-value} = \mathbb{E}_{T_{1:p}|\xi_0} (\mathbb{I}[\Delta_R(T_{1:p}|\xi_0) > \Delta_R(t_{1:p})]) \quad (42)$$

where $\mathbb{E}_{T_{1:p}|\xi_0} (\cdot)$ denotes the expectation over the rebound times $T_{1:p}$ obtained from 1000 realizations of simulated populations each with p patients, and under the null hypothesis $\xi_0 = 1$ is the Iverson bracket that takes value 1 when its argument is true and 0, otherwise (**Equation 34**). The observed Δ_R (**Equation 41**), the distribution of simulated values of $\Delta_R(T_{1:p}|\xi_0)$ under the null hypothesis, and the resulting p-value = 0.004 are shown in **Figure 4—figure supplement 3B, C**.

It should be noted that here we use the disparity measure because the corresponding likelihood function for the reservoir factor is inaccessible through forward simulations of populations. However, a general analogy exists between our approach and the more commonly used likelihood approach. Specifically, in an analogous likelihood-ratio test, the test statistic $\Delta_L = \max_{\xi} \log p(\xi|t_{1:p}) - \log p(\xi_0|t_{1:p})$ would be asymptotically χ^2 -distributed with one degree of freedom under the null hypothesis **Fisher, 1954**, and the quantile under the null-hypothesis (p-value) would be estimated by inverting the χ^2 cumulative distribution function (i.e. a χ^2 test).

Robustness of selection inference to strains from different clades of HIV

The longitudinal deep sequencing data of **Zanini et al., 2015** is collected from 11 patients, 9 of whom are infected with clade B strains of HIV-1, which is the dominant clade circulating in Europe **Spira et al., 2003**. All of the clinical trials we considered **Caskey et al., 2015; Caskey et al., 2017; Baron et al., 2018** are from patients carrying clade B strains. For the results presented in the main text, we included all the 11 patients in our analysis. Here, we test whether our inference of selection is sensitive to the choice of including or excluding non-clade B patients in our analysis. We therefore repeated our inference procedure for selection by excluding the two non-clade B patients. Model robustness **Figure 4—figure supplement 1A,B** shows a strong agreement between the Bayesian posterior for selection factors in the two cases, with a slight increase in uncertainty for the case with only clade B patients. This increased uncertainty is related to the reduction in sample size by excluding the non-clade B patients from data. Nonetheless, the richness of the intra-patient diversity makes the inference robust to the exclusion of one or two patients.

Robustness of predictions for trial efficacy to the inferred values of selection strength

How sensitive are the outcomes of our predictions for the rebound time distributions (**Figure 1**) to the exact values of inferred selection strengths we used for our simulations? We addressed this question by performing a disparity analysis similar to that for the diversity θ in **Equation 41**. Specifically, we assessed whether we might need to rescale our inferred selection strength σ/θ_{ts} by a multiplicative factor ξ_s (**Figure 4—figure supplement 4**).

In contrast to diversity, the reduction in disparity for adjustment of selection with a factor ξ_s is small (**Figure 4—figure supplement 4A**) and not statistically significant (p-value = 0.49; **Figure 4—figure supplement 4B**), and could be attributed to count noise. Still, we cannot discount the possibility that selection was slightly overestimated, possibly due to the effect of compensatory mutations in linked genome, the interplay between the reservoir and the inference of selection, or other biological factors. Nonetheless, in absolute terms, the null hypothesis (i.e. $\xi_s = 1$) cannot be rejected and we have no statistical justification for adding an adjustment factor for selection inferred from untreated patients.

Robustness of rebound time predictions to methods of identifying escape mutations

Our predictions rely on identifying escape variants against each bNAb, either based on in vivo trial data and patient surveillance, or in vitro DMS assays. To test the sensitivity of our results to these methods, we compare the predictions of rebound time distributions when identifying escape sites from the DMS data versus the trial data. In Appendix 4, we perform this comparison for the 10–1074 and the PGT121 bNAbs; we do not include the 3BNC117 bNAb in this analysis since it targets the CD4 binding site of HIV-1 and the DMS data is unreliable for identifying escape variants against it. As shown in **Appendix 4—figure 1**, both trial-inferred and DMS-inferred escape sites result in good predictions for rebound time distributions. However, it appears that using the DMS-inferred escape sites could lead to a more optimistic prediction for treatment success (i.e. a later rebound). This inconsistency may be due to the fact that DMS data is collected in vitro, and other biological factors could be influencing the in vivo escape patterns. Moreover, the differences in the genetic composition of the HIV-1 strains circulating in patients enrolled in trials and the strains used for the DMS experiments could lead to different epistatic interactions that can enhance or reduce the chances of escape. For a systematic understanding of these differences and limitations, more experiments would be necessary. Nonetheless, DMS data can provide baseline to gauge the efficacy of different bNAbs for therapy and further in vivo investigation.

Acknowledgements

We are thankful to Jesse Bloom and Adam Dingens for their input on the Deep Mutational Scanning data of HIV, and Michael Lässig and Matthijs Meijers for helpful discussions. This work has been supported by the NSF CAREER award (grant No: 2045054), DFG grant (SFB1310) for Predictability in Evolution, and the MPRG funding through the Max Planck Society.

Additional information

Competing interests

Armita Nourmohammad: Reviewing editor, eLife. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft	1310	Armita Nourmohammad
National Science Foundation	2045054	Armita Nourmohammad

Funder	Grant reference number	Author
Max Planck Institute for Dynamics and Self-organization	Open access funding	Colin LaMont Armita Nourmohammad

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Colin LaMont, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing; Jakub Otwinowski, Conceptualization, Methodology, Writing – review and editing; Kanika Vanshylla, Conceptualization, Data curation, Writing – review and editing; Henning Gruell, Florian Klein, Conceptualization, Writing – review and editing; Armita Nourmohammad, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

Author ORCIDs

Henning Gruell  <http://orcid.org/0000-0002-0725-7138>

Armita Nourmohammad  <http://orcid.org/0000-0002-6245-3553>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.76004.sa1>

Author response <https://doi.org/10.7554/eLife.76004.sa2>

Additional files

Supplementary files

- MDAR checklist

Data availability

The current manuscript is a computational study, so no data have been generated for this manuscript. Reference to the previously published data used in this manuscript is provided. Modelling code is uploaded on GitHub at <https://github.com/StatPhysBio/HIVTreatmentOptimization>, (copy archived at [swh:1:rev:0194cf6e554996a066633e99dd53cd5901da552e](https://www.swh.io/rev/0194cf6e554996a066633e99dd53cd5901da552e)) and in the Julia package <https://github.com/StatPhysBio/EscapeSimulator> (copy archived at [swh:1:rev:9a343f598820bafddfc7ea4547cefa90bf96fd6e](https://www.swh.io/rev/9a343f598820bafddfc7ea4547cefa90bf96fd6e)).

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Zanini et al	2015	Project: PRJEB9618	https://www.ebi.ac.uk/ena/browser/view/PRJEB9618?show=reads	European Nucleotide Archive, PRJEB9618
Caskey et al	2015	HIV-1 isolate 2A1_DB7_02 from USA envelope glycoprotein (env) gene, partial cds	https://www.ncbi.nlm.nih.gov/nuccore/KX016803	NCBI GenBank, KX016803
Caskey et al	2017	Antibody 10-1074 suppresses viremia in HIV-1-infected individuals	https://www.ncbi.nlm.nih.gov/nuccore/KY323724	NCBI GenBank, KY323724-KY324834
Bar-On et al	2018	Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in viremic individuals	https://www.ncbi.nlm.nih.gov/nuccore/MH632763	NCBI GenBank, MH632763

References

- Allen LJS. 2010. An Introduction to Stochastic Processes with Applications to Biology. Boca Raton: CRC press. DOI: <https://doi.org/10.1201/b12537>
- Avettand-Fenoel V, Mahlaoui N, Chaix ML, Milliancourt C, Burgard M, Cavazzana-Calvo M, Rouzioux C, Blanche S. 2007. Failure of bone marrow transplantation to eradicate HIV reservoir despite efficient HAART. *AIDS* 21:776–777. DOI: <https://doi.org/10.1097/QAD.0b013e3280b01836>, PMID: 17413704
- Bar KJ, Sneller MC, Harrison LJ, Justement JS, Overton ET, Petrone ME, Salantes DB, Seamon CA, Scheinfeld B, Kwan RW, Learn GH, Proschan MA, Kreider EF, Blazkova J, Bardsley M, Refsland EW, Messer M, Clarridge KE, Tustin NB, Madden PJ, et al. 2016. Effect of HIV antibody VRC01 on viral rebound after treatment interruption. *The New England Journal of Medicine* 375:2037–2050. DOI: <https://doi.org/10.1056/NEJMoa1608243>, PMID: 27959728
- Baron Y, Gruell H, Schoofs T, Pai JA, Nogueira L, Butler AL, Millard K, Lehmann C, Suárez I, Oliveira TY, Karagounis T, Cohen YZ, Wyen C, Scholten S, Handl L, Belblidia S, Dizon JP, Vehreschild JJ, Witmer-Pack M, Shimeliovich I, et al. 2018. Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in viremic individuals. *Nature Medicine* 24:1701–1707. DOI: <https://doi.org/10.1038/s41591-018-0186-4>, PMID: 30258217
- Barouch DH, Whitney JB, Moldt B, Klein F, Oliveira TY, Liu J, Stephenson KE, Chang H-W, Shekhar K, Gupta S, Nkolola JP, Seaman MS, Smith KM, Borducchi EN, Cabral C, Smith JY, Blackmore S, Sanisetty S, Perry JR, Beck M, et al. 2013. Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature* 503:224–228. DOI: <https://doi.org/10.1038/nature12744>, PMID: 24172905
- Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. 2004. Evidence for positive epistasis in HIV-1. *Science* 306:1547–1550. DOI: <https://doi.org/10.1126/science.1101786>, PMID: 15567861
- Burton DR, Poignard P, Stanfield RL, Wilson IA. 2012. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* 337:183–186. DOI: <https://doi.org/10.1126/science.1225416>, PMID: 22798606
- Burton DR, Hangartner L. 2016. Broadly neutralizing antibodies to HIV and their role in vaccine design. *Annual Review of Immunology* 34:635–659. DOI: <https://doi.org/10.1146/annurev-immunol-041015-055515>, PMID: 27168247
- Cardozo-Ojeda EF, Perelson AS. 2021. Modeling HIV-1 Within-Host dynamics after passive infusion of the broadly neutralizing antibody VRC01. *Frontiers in Immunology* 12:710012. DOI: <https://doi.org/10.3389/fimmu.2021.710012>, PMID: 34531859
- Caskey M, Klein F, Lorenzi JCC, Seaman MS, West AP Jr, Buckley N, Kremer G, Nogueira L, Braunschweig M, Scheid JF, Horwitz JA, Shimeliovich I, Ben-Avraham S, Witmer-Pack M, Platten M, Lehmann C, Burke LA, Hawthorne T, Gorelick RJ, Walker BD, et al. 2015. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* 522:487–491. DOI: <https://doi.org/10.1038/nature14411>, PMID: 25855300
- Caskey M, Klein F, Nussenzweig MC. 2016. Broadly neutralizing antibodies for HIV-1 prevention or immunotherapy. *The New England Journal of Medicine* 375:2019–2021. DOI: <https://doi.org/10.1056/NEJMp1613362>, PMID: 27959740
- Caskey M, Schoofs T, Gruell H, Settler A, Karagounis T, Kreider EF, Murrell B, Pfeifer N, Nogueira L, Oliveira TY, Learn GH, Cohen YZ, Lehmann C, Gillor D, Shimeliovich I, Unson-O'Brien C, Weiland D, Robles A, Kümmerle T, Wyen C, et al. 2017. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nature Medicine* 23:185–191. DOI: <https://doi.org/10.1038/nm.4268>, PMID: 28092665
- Chaillon A, Gianella S, Dellicour S, Rawlings SA, Schlub TE, De Oliveira MF, Ignacio C, Porrachia M, Vrancken B, Smith DM. 2020. HIV persists throughout deep tissues with repopulation from multiple anatomical sources. *The Journal of Clinical Investigation* 130:1699–1712. DOI: <https://doi.org/10.1172/JCI134815>, PMID: 31910162
- Chelate. 2022a. HIVTreatmentOptimization. swh:1:rev:0194cf6e554996a066633e99dd53cd5901da552e. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:776b6f1495f749cd2e687e53359fdde9593b5a47;origin=https://github.com/StatPhysBio/HIVTreatmentOptimization;visit=swh:1:snp:ac91d9b0c04a7418b4e7a98db0addd24199f914;anchor=swh:1:rev:0194cf6e554996a066633e99dd53cd5901da552e>
- Chelate. 2022b. EscapeSimulator. swh:1:rev:9a343f598820bafddfc7ea4547cefa90bf96fd6e. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:c6cfa184d349a3272b0211efac8fc1f0f3d398e3;origin=https://github.com/StatPhysBio/EscapeSimulator;visit=swh:1:snp:1727701be1d35fdeb61fdd39bc0a9bbc0c1cb90a;anchor=swh:1:rev:9a343f598820bafddfc7ea4547cefa90bf96fd6e>
- Chen L, Kwon YD, Zhou T, Wu X, O'Dell S, Cavacini L, Hessel AJ, Pancera M, Tang M, Xu L, Yang ZY, Zhang MY, Arthos J, Burton DR, Dimitrov DS, Nabel GJ, Posner MR, Sodroski J, Wyatt R, Mascola JR, et al. 2009. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science* 326:1123–1127. DOI: <https://doi.org/10.1126/science.1175868>, PMID: 19965434
- Chun T-W, Nickle DC, Justement JS, Large D, Semerjian A, Curlin ME, O'Shea MA, Hallahan CW, Daucher M, Ward DJ, Moir S, Mullins JI, Kovacs C, Fauci AS. 2005. HIV-infected individuals receiving effective antiviral therapy for extended periods of time continually replenish their viral reservoir. *The Journal of Clinical Investigation* 115:3250–3255. DOI: <https://doi.org/10.1172/JCI26197>, PMID: 16276421
- Cameron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* 100:19–31. DOI: <https://doi.org/10.1038/sj.hdy.6801059>, PMID: 17878920

- Crow JF**, Kimura M. An Introduction to Population Genetics Theory. Caldwell: The Blackburn Press; 2010.
- Dingens AS**, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive mapping of HIV-1 escape from a broadly neutralizing antibody. *Cell Host & Microbe* **21**:777–787. DOI: <https://doi.org/10.1016/j.chom.2017.05.003>, PMID: 28579254
- Dingens AS**, Arenz D, Weight H, Overbaugh J, Bloom JD. 2019. An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* **50**:520–532. DOI: <https://doi.org/10.1016/j.immuni.2018.12.017>, PMID: 30709739
- Durham ND**, Agrawal A, Waltari E, Croote D, Zanini F, Fouch M, Davidson E, Smith O, Carabajal E, Pak JE, Doranz BJ, Robinson M, Sanz AM, Albornoz LL, Rosso F, Einav S, Quake SR, McCutcheon KM, Goo L. 2019. Broadly neutralizing human antibodies against dengue virus identified by single B cell transcriptomics. *eLife* **8**:e52384. DOI: <https://doi.org/10.7554/eLife.52384>, PMID: 31820734
- Ekiert DC**, Wilson IA. 2012. Broadly neutralizing antibodies against influenza virus and prospects for universal therapies. *Current Opinion in Virology* **2**:134–141. DOI: <https://doi.org/10.1016/j.coviro.2012.02.005>, PMID: 22482710
- Ekström M**. 2008. Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence. *Journal of Statistical Planning and Inference* **138**:1778–1791. DOI: <https://doi.org/10.1016/j.jspi.2007.06.031>
- Feder AF**, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. 2016. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife* **5**:e10670. DOI: <https://doi.org/10.7554/eLife.10670>, PMID: 26882502
- Feder AF**, Kline C, Polacino P, Cottrell M, Kashuba ADM, Keele BF, Hu SL, Petrov DA, Pennings PS, Ambrose Z. 2017. A spatio-temporal assessment of simian/human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host. *PLOS Pathogens* **13**:e1006358. DOI: <https://doi.org/10.1371/journal.ppat.1006358>, PMID: 28542550
- Feder AF**, Pennings PS, Petrov DA. 2021. The clarifying role of time series data in the population genetics of HIV. *PLOS Genetics* **17**:e1009050. DOI: <https://doi.org/10.1371/journal.pgen.1009050>, PMID: 33444376
- Ferguson AL**, Mann JK, Omarjee S, Ndung'u T, Walker BD, Chakraborty AK. 2013. Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**:606–617. DOI: <https://doi.org/10.1016/j.immuni.2012.11.022>, PMID: 23521886
- Fisher RA**. Statistical Methods for Research Workers. Twelfth ed. New York: Hafner Publishing; 1954.
- Fisher RA**. 1956. Statistical Methods and Scientific Inference. London: Oliver and Boyd.
- García F**, Plana M, Vidal C, Cruceta A, O'Brien WA, Pantaleo G, Pumarola T, Gallart T, Miró JM, Gatell JM. 1999. Dynamics of viral load rebound and immunological changes after stopping effective antiretroviral therapy. *AIDS* **13**:F79–F86. DOI: <https://doi.org/10.1097/00002030-199907300-00002>, PMID: 10449278
- García F**, Plana M, Ortiz GM, Bonhoeffer S, Soriano A, Vidal C, Cruceta A, Arnedo M, Gil C, Pantaleo G, Pumarola T, Gallart T, Nixon DF, Miró JM, Gatell JM. 2001. The virological and immunological consequences of structured treatment interruptions in chronic HIV-1 infection. *AIDS* **15**:F29–F40. DOI: <https://doi.org/10.1097/00002030-200106150-00002>, PMID: 11416735
- Gaschen B**, Kuiken C, Korber B, Foley B. 2001. Retrieval and on-the-fly alignment of sequence fragments from the HIV database. *Bioinformatics* **17**:415–418. DOI: <https://doi.org/10.1093/bioinformatics/17.5.415>, PMID: 11331235
- Geman S**, Geman D. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**:721–741. DOI: <https://doi.org/10.1109/tpami.1984.4767596>, PMID: 22499653
- Gillespie DT**. 2002. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**:2340–2361. DOI: <https://doi.org/10.1021/j100540a008>
- Graham RL**, Knuth DE, Patashnik O, Liu S. 1989. Concrete mathematics: A foundation for computer science. *Computers in Physics* **3**:106. DOI: <https://doi.org/10.1063/1.4822863>
- Gruell H**, Klein F. 2018. Antibody-mediated prevention and treatment of HIV-1 infection. *Retrovirology* **15**:73. DOI: <https://doi.org/10.1186/s12977-018-0455-9>, PMID: 30445968
- Haddox HK**, Dingens AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. *eLife* **7**:e34420. DOI: <https://doi.org/10.7554/eLife.34420>, PMID: 29590010
- Hastings WK**. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**:97–109. DOI: <https://doi.org/10.1093/biomet/57.1.97>
- Hill WG**, Robertson A. 2009. The effect of linkage on limits to artificial selection. *Genetical Research* **8**:269–294. DOI: <https://doi.org/10.1017/S0016672300010156>
- Horwitz JA**, Halper-Stromberg A, Mouquet H, Gitlin AD, Tretiakova A, Eisenreich TR, Malbec M, Gravemann S, Billerbeck E, Dorner M, Büning H, Schwartz O, Knops E, Kaiser R, Seaman MS, Wilson JM, Rice CM, Ploss A, Bjorkman PJ, Klein F, et al. 2013. HIV-1 suppression and durable control by combining single broadly neutralizing antibodies and antiretroviral drugs in humanized mice. *PNAS* **110**:16538–16543. DOI: <https://doi.org/10.1073/pnas.1315295110>, PMID: 24043801
- Illingworth CJR**, Raghvani J, Serwadda D, Sewankambo NK, Robb ML, Eller MA, Redd AR, Quinn TC, Lythgoe KA. 2020. A de novo approach to inferring within-host fitness effects during untreated HIV-1 infection. *PLOS Pathogens* **16**:e1008171. DOI: <https://doi.org/10.1371/journal.ppat.1008171>, PMID: 32492061

- Ioannidis JP**, Havlir DV, Tebas P, Hirsch MS, Collier AC, Richman DD. 2000. Dynamics of HIV-1 viral load rebound among patients with previous suppression of viral replication. *AIDS* **14**:1481–1488. DOI: <https://doi.org/10.1097/00002030-200007280-00003>, PMID: 10983634
- Julg B**, Pegu A, Abbink P, Liu J, Brinkman A, Molloy K, Mojta S, Chandrashekar A, Callow K, Wang K, Chen X, Schmidt SD, Huang J, Koup RA, Seaman MS, Keele BF, Mascola JR, Connors M, Barouch DH. 2017. Virological control by the CD4-Binding site antibody n6 in simian-human immunodeficiency virus-infected rhesus monkeys. *Journal of Virology* **91**:e00498-17. DOI: <https://doi.org/10.1128/JVI.00498-17>, PMID: 28539448
- Kimura M**. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *PNAS* **78**:454–458. DOI: <https://doi.org/10.1073/pnas.78.1.454>, PMID: 6165991
- Klein F**, Halper-Stromberg A, Horwitz JA, Gruell H, Scheid JF, Bournazos S, Mouquet H, Spatz LA, Diskin R, Abadir A, Zang T, Dorner M, Billerbeck E, Labitt RN, Gaebler C, Marcovecchio P, Incesu RB, Eisenreich TR, Bieniasz PD, Seaman MS, et al. 2012. HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**:118–122. DOI: <https://doi.org/10.1038/nature11604>, PMID: 23103874
- Klein F**, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. 2013. Antibodies in HIV-1 vaccine development and therapy. *Science* **341**:1199–1204. DOI: <https://doi.org/10.1126/science.1241144>, PMID: 24031012
- Kong R**, Louder MK, Wagh K, Bailer RT, deCamp A, Greene K, Gao H, Taft JD, Gazumyan A, Liu C, Nussenzweig MC, Korber B, Montefiori DC, Mascola JR. 2015. Improving neutralization potency and breadth by combining broadly reactive HIV-1 antibodies targeting major neutralization epitopes. *Journal of Virology* **89**:2659–2671. DOI: <https://doi.org/10.1128/JVI.03136-14>, PMID: 25520506
- Kwon YD**, Georgiev IS, Ofek G, Zhang B, Asokan M, Bailer RT, Bao A, Caruso W, Chen X, Choe M, Druz A, Ko S-Y, Louder MK, McKee K, O'Dell S, Pegu A, Rudicell RS, Shi W, Wang K, Yang Y, et al. 2016. Optimization of the solubility of HIV-1-Neutralizing antibody 10E8 through somatic variation and structure-based design. *Journal of Virology* **90**:5899–5914. DOI: <https://doi.org/10.1128/JVI.03246-15>, PMID: 27053554
- Kwong PD**, Mascola JR, Nabel GJ. 2013. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nature Reviews Immunology* **13**:693–701. DOI: <https://doi.org/10.1038/nri3516>
- LaBranche CC**, McGuire AT, Gray MD, Behrens S, Kwong PD, Chen X, Zhou T, Sattentau QJ, Peacock J, Eaton A, Greene K, Gao H, Tang H, Perez LG, Chen X, Saunders KO, Kwong PD, Mascola JR, Haynes BF, Stamatatos L, et al. 2018. HIV-1 envelope glycan modifications that permit neutralization by germline-reverted VRC01-class broadly neutralizing antibodies. *PLoS Pathogens* **14**:e1007431. DOI: <https://doi.org/10.1371/journal.ppat.1007431>, PMID: 30395637
- Lemey P**, Rambaut A, Pybus OG. 2006. HIV evolutionary dynamics within and among hosts. *AIDS Reviews* **8**:125–140 PMID: 17078483.
- Liao H-X**, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, Zhu J, Shapiro L, NISC Comparative Sequencing Program, Mullikin JC, Gnanakaran S, Hraber P, Wiehe K, Kelsoe G, Yang G, Xia S-M, et al. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**:469–476. DOI: <https://doi.org/10.1038/nature12053>, PMID: 23552890
- Lienhardt C**, Raviglione M, Spigelman M, Hafner R, Jaramillo E, Hoelscher M, Zumla A, Gheuens J. 2012. New drugs for the treatment of tuberculosis: needs, challenges, promise, and prospects for the future. *The Journal of Infectious Diseases* **205**:S241–S249. DOI: <https://doi.org/10.1093/infdis/jis034>, PMID: 22448022
- Lindsay BG**. 1994. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *The Annals of Statistics* **22**:1081–1114. DOI: <https://doi.org/10.1214/aos/1176325512>
- Liu B**, Zhang W, Zhang H. 2019. Development of CAR-T cells for long-term eradication and surveillance of HIV-1 reservoir. *Current Opinion in Virology* **38**:21–30. DOI: <https://doi.org/10.1016/j.coviro.2019.04.004>, PMID: 31132749
- Louie RHY**, Kaczorowski KJ, Barton JP, Chakraborty AK, McKay MR. 2018. Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *PNAS* **115**:E564–E573. DOI: <https://doi.org/10.1073/pnas.1717765115>, PMID: 29311326
- Lu CL**, Murakowski DK, Bournazos S, Schoofs T, Sarkar D, Halper-Stromberg A, Horwitz JA, Nogueira L, Golijanin J, Gazumyan A, Ravetch JV, Caskey M, Chakraborty AK, Nussenzweig MC. 2016. Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 in vivo. *Science* **352**:1001–1004. DOI: <https://doi.org/10.1126/science.aaf1279>, PMID: 27199430
- Lynch RM**, Wong P, Tran L, O'Dell S, Nason MC, Li Y, Wu X, Mascola JR. 2015. HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *Journal of Virology* **89**:4201–4213. DOI: <https://doi.org/10.1128/JVI.03608-14>, PMID: 25631091
- Margolis DM**, Koup RA, Ferrari G. 2017. HIV antibodies for treatment of HIV infection. *Immunological Reviews* **275**:313–323. DOI: <https://doi.org/10.1111/imr.12506>, PMID: 28133794
- Mayer BT**, deCamp AC, Huang Y, Schiffer JT, Gottardo R, Gilbert PB, Reeves DB. 2022. Optimizing clinical dosing of combination broadly neutralizing antibodies for HIV prevention. *PLoS Computational Biology* **18**:e1010003. DOI: <https://doi.org/10.1371/journal.pcbi.1010003>, PMID: 35385469
- McCullagh P**, Nelder JA. 2019. Generalized Linear Models. Boca Raton: Routledge. DOI: <https://doi.org/10.1201/9780203753736>
- Meijers M**, Vanshilla K, Gruell H, Klein F, Lässig M. 2021. Predicting in vivo escape dynamics of HIV-1 from a broadly neutralizing antibody. *PNAS* **118**:e2104651118. DOI: <https://doi.org/10.1073/pnas.2104651118>, PMID: 34301904

- Meintjes PL, Rodrigo AG. 2005. Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. *Journal of Bioinformatics and Computational Biology* **3**:157–168. DOI: <https://doi.org/10.1142/s0219720005000953>, PMID: 15751118
- Mendoza P, Gruell H, Nogueira L, Pai JA, Butler AL, Millard K, Lehmann C, Suárez I, Oliveira TY, Lorenzi JCC, Cohen YZ, Wyen C, Kümmerle T, Karagounis T, Lu C-L, Handl L, Unson-O'Brien C, Patel R, Ruping C, Schlotz M, et al. 2018. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* **561**:479–484. DOI: <https://doi.org/10.1038/s41586-018-0531-2>, PMID: 30258136
- Mouquet H, Nussenzweig MC. 2013. HIV: Roadmaps to a vaccine. *Nature* **496**:441–442. DOI: <https://doi.org/10.1038/nature12091>, PMID: 23552894
- Neher RA, Leitner T. 2010. Recombination rate and selection strength in HIV intra-patient evolution. *PLOS Computational Biology* **6**:e1000660. DOI: <https://doi.org/10.1371/journal.pcbi.1000660>, PMID: 20126527
- Nielsen R. 2006. *Statistical Methods in Molecular Evolution*. New York: Springer. DOI: <https://doi.org/10.1007/0-387-27733-1>
- Nourmohammad A, Otwinowski J, Plotkin JB. 2016. Host-Pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLOS Genetics* **12**:e1006171. DOI: <https://doi.org/10.1371/journal.pgen.1006171>, PMID: 27442127
- Nourmohammad A, Otwinowski J, Łuksza M, Mora T, Walczak AM, Leitner T. 2019. Fierce selection and interference in b-cell repertoire response to chronic HIV-1. *Molecular Biology and Evolution* **36**:2184–2194. DOI: <https://doi.org/10.1093/molbev/msz143>, PMID: 31209469
- Otsuka Y, Schmitt K, Quinlan BD, Gardner MR, Alfant B, Reich A, Farzan M, Choe H. 2018. Diverse pathways of escape from all well-characterized VRC01-class broadly neutralizing HIV-1 antibodies. *PLOS Pathogens* **14**:e1007238. DOI: <https://doi.org/10.1371/journal.ppat.1007238>, PMID: 30125330
- Owen A, Zhou Y. 2000. Safe and effective importance sampling. *Journal of the American Statistical Association* **95**:135–143. DOI: <https://doi.org/10.1080/01621459.2000.10473909>
- Pancera M, Changela A, Kwong PD. 2017. How HIV-1 entry mechanism and broadly neutralizing antibodies guide structure-based vaccine design. *Current Opinion in HIV and AIDS* **12**:229–240. DOI: <https://doi.org/10.1097/COH.0000000000000360>, PMID: 28422787
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**:1582–1586. DOI: <https://doi.org/10.1126/science.271.5255.1582>, PMID: 8599114
- Perelson AS. 2002. Modelling viral and immune system dynamics. *Nature Reviews. Immunology* **2**:28–36. DOI: <https://doi.org/10.1038/nri700>, PMID: 11905835
- Reeves DB, Huang Y, Duke ER, Mayer BT, Cardozo-Ojeda EF, Boshier FA, Swan DA, Rolland M, Robb ML, Mascola JR, Cohen MS, Corey L, Gilbert PB, Schiffer JT. 2020. Mathematical modeling to reveal breakthrough mechanisms in the HIV Antibody Mediated Prevention (AMP) trials. *PLOS Computational Biology* **16**:e1007626. DOI: <https://doi.org/10.1371/journal.pcbi.1007626>, PMID: 32084132
- Ribeiro RM, Bonhoeffer S. 2000. Production of resistant HIV mutants during antiretroviral therapy. *PNAS* **97**:7681–7686. DOI: <https://doi.org/10.1073/pnas.97.14.7681>, PMID: 10884399
- Risken H. 1989. *The Fokker-Planck Equation*. Berlin, Heidelberg: Springer-Verlag. DOI: <https://doi.org/10.1007/978-3-642-61544-3>
- Rong L, Feng Z, Perelson AS. 2007. Emergence of HIV-1 drug resistance during antiretroviral treatment. *Bulletin of Mathematical Biology* **69**:2027–2060. DOI: <https://doi.org/10.1007/s11538-007-9203-3>, PMID: 17450401
- Rong L, Dahari H, Ribeiro RM, Perelson AS. 2010. Rapid emergence of protease inhibitor resistance in hepatitis C virus. *Science Translational Medicine* **2**:30ra32. DOI: <https://doi.org/10.1126/scitranslmed.3000544>, PMID: 20445200
- Saha A, Dixit NM. 2020. Pre-existing resistance in the latent reservoir can compromise VRC01 therapy during chronic HIV-1 infection. *PLOS Computational Biology* **16**:11. DOI: <https://doi.org/10.1371/journal.pcbi.1008434>, PMID: 33253162
- Scheid JF, Horwitz JA, Bar-On Y, Kreider EF, Lu C-L, Lorenzi JCC, Feldmann A, Braunschweig M, Nogueira L, Oliveira T, Shimeliovich I, Patel R, Burke L, Cohen YZ, Hadrihan S, Settler A, Witmer-Pack M, West AP Jr, Juelg B, Keler T, et al. 2016. HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. *Nature* **535**:556–560. DOI: <https://doi.org/10.1038/nature18929>, PMID: 27338952
- Schommers P, Gruell H, Abernathy ME, Tran MK, Dingens AS, Gristick HB, Barnes CO, Schoofs T, Schlotz M, Vanshylla K, Kreer C, Weiland D, Holtick U, Scheid C, Valter MM, van Gils MJ, Sanders RW, Vehreschild JJ, Cornely OA, Lehmann C, et al. 2020. Restriction of HIV-1 escape by a highly broad and potent neutralizing antibody. *Cell* **180**:471–489. DOI: <https://doi.org/10.1016/j.cell.2020.01.010>, PMID: 32004464
- Schoofs T, Klein F, Braunschweig M, Kreider EF, Feldmann A, Nogueira L, Oliveira T, Lorenzi JCC, Parrish EH, Learn GH, West AP, Bjorkman PJ, Schlesinger SJ, Seaman MS, Czartoski J, McElrath MJ, Pfeifer N, Hahn BH, Caskey M, Nussenzweig MC. 2016. HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1. *Science* **352**:997–1001. DOI: <https://doi.org/10.1126/science.aaf0972>, PMID: 27199429
- Shan L, Siliciano RF. 2013. From reactivation of latent HIV-1 to elimination of the latent reservoir: the presence of multiple barriers to viral eradication. *BioEssays* **35**:544–552. DOI: <https://doi.org/10.1002/bies.201200170>, PMID: 23613347

- Sharkey M**, Babic DZ, Greenough T, Gulick R, Kuritzkes DR, Stevenson M. 2011. Episomal viral cDNAs identify a reservoir that fuels viral rebound after treatment interruption and that contributes to treatment failure. *PLOS Pathogens* **7**:e1001303. DOI: <https://doi.org/10.1371/journal.ppat.1001303>, PMID: 21383975
- Shingai M**, Nishimura Y, Klein F, Mouquet H, Donau OK, Plishka R, Buckler-White A, Seaman M, Piatak M, Lifson JD, Dimitrov DS, Nussenzweig MC, Martin MA. 2013. Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature* **503**:277–280. DOI: <https://doi.org/10.1038/nature12746>, PMID: 24172896
- Simpson DG**. 1987. Minimum hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association* **82**:802–807. DOI: <https://doi.org/10.1080/01621459.1987.10478501>
- Sok D**, Burton DR. 2018. Recent progress in broadly neutralizing antibodies to HIV. *Nature Immunology* **19**:1179–1188. DOI: <https://doi.org/10.1038/s41590-018-0235-7>, PMID: 30333615
- Sparrow E**, Friede M, Sheikh M, Torvaldsen S, Newall AT. 2016. Passive immunization for influenza through antibody therapies, a review of the pipeline, challenges and potential applications. *Vaccine* **34**:5442–5448. DOI: <https://doi.org/10.1016/j.vaccine.2016.08.057>, PMID: 27622299
- Spira S**, Wainberg MA, Loemba H, Turner D, Brenner BG. 2003. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *The Journal of Antimicrobial Chemotherapy* **51**:229–240. DOI: <https://doi.org/10.1093/jac/dkg079>, PMID: 12562686
- Stephenson KE**, Julg B, Tan CS, Zash R, Walsh SR, Rolle C-P, Monczor AN, Lupo S, Gelderblom HC, Ansel JL, Kanjilal DG, Maxfield LF, Nkolola J, Borducchi EN, Abbink P, Liu J, Peter L, Chandrashekar A, Nityanandam R, Lin Z, et al. 2021. Safety, pharmacokinetics and antiviral activity of PGT121, A broadly neutralizing monoclonal antibody against HIV-1: A randomized, placebo-controlled, phase 1 clinical trial. *Nature Medicine* **27**:1718–1724. DOI: <https://doi.org/10.1038/s41591-021-01509-0>, PMID: 34621054
- Stoddart JA**, Taylor JF. 1988. Genotypic diversity: estimation and prediction in samples. *Genetics* **118**:705–711. DOI: <https://doi.org/10.1093/genetics/118.4.705>, PMID: 17246421
- Strelkova N**, Lässig M. 2012. Clonal interference in the evolution of influenza. *Genetics* **192**:671–682. DOI: <https://doi.org/10.1534/genetics.112.143396>, PMID: 22851649
- Tagarro A**, Chan M, Zangari P, Ferns B, Foster C, De Rossi A, Nastouli E, Muñoz-Fernández MA, Gibb D, Rossi P, Giaquinto C, Babiker A, Fortuny C, Freguja R, Cotugno N, Judd A, Noguera-Julian A, Navarro ML, Mellado MJ, Klein N, et al. 2018. Early and highly suppressive antiretroviral therapy are main factors associated with low viral reservoir in european perinatally HIV-Infected children. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **79**:269–276. DOI: <https://doi.org/10.1097/QAI.0000000000001789>
- Theys K**, Feder AF, Gelbart M, Hartl M, Stern A, Pennings PS. 2018. Within-patient mutation frequencies reveal fitness costs of CpG dinucleotides and drastic amino acid changes in HIV. *PLOS Genetics* **14**:e1007420. DOI: <https://doi.org/10.1371/journal.pgen.1007420>, PMID: 29953449
- Tomaras GD**, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, Decamp AC, Parks RJ, Ashley VC, Lucas JT, Cohen M, Eron J, Hicks CB, Liao H-X, Self SG, Landucci G, Forthal DN, Weinhold KJ, Keele BF, Hahn BH, et al. 2008. Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology* **82**:12449–12463. DOI: <https://doi.org/10.1128/JVI.01708-08>, PMID: 18842730
- Wagh K**, Bhattacharya T, Williamson C, Robles A, Bayne M, Garrity J, Rist M, Rademeyer C, Yoon H, Lapedes A, Gao H, Greene K, Louder MK, Kong R, Karim SA, Burton DR, Barouch DH, Nussenzweig MC, Mascola JR, Morris L, et al. 2016. Optimal combinations of broadly neutralizing antibodies for prevention and treatment of HIV-1 clade c infection. *PLOS Pathogens* **12**:e1005520. DOI: <https://doi.org/10.1371/journal.ppat.1005520>, PMID: 27028935
- Walker LM**, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, et al. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**:285–289. DOI: <https://doi.org/10.1126/science.1178746>, PMID: 19729618
- Walker LM**, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**:466–470. DOI: <https://doi.org/10.1038/nature10373>
- West AP**, Scharf L, Scheid JF, Klein F, Bjorkman PJ, Nussenzweig MC. 2014. Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* **156**:633–648. DOI: <https://doi.org/10.1016/j.cell.2014.01.052>, PMID: 24529371
- Wilkinson DJ**. 2019. Stochastic Modelling for Systems Biology. Third ed. Boca Raton: CRC Press.
- Wong JK**, Yukl SA. 2016. Tissue reservoirs of HIV. *Current Opinion in HIV and AIDS* **11**:362–370. DOI: <https://doi.org/10.1097/COH.0000000000000293>, PMID: 27259045
- Yu WH**, Su D, Torabi J, Fennessey CM, Shiakolas A, Lynch R, Chun TW, Doria-Rose N, Alter G, Seaman MS, Keele BF, Lauffenburger DA, Julg B. 2019. Predicting the broadly neutralizing antibody susceptibility of the HIV reservoir. *JCI Insight* **4**:130153. DOI: <https://doi.org/10.1172/jci.insight.130153>, PMID: 31484826
- Zanini F**, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA. 2015. Population genomics of inpatient HIV-1 evolution. *eLife* **4**:e11282. DOI: <https://doi.org/10.7554/eLife.11282>, PMID: 26652000
- Zanini F**, Puller V, Brodin J, Albert J, Neher RA. 2017. In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evolution* **3**:vex003. DOI: <https://doi.org/10.1093/ve/vex003>, PMID: 28458914

- Zhang T**, Dai L, Barton JP, Du Y, Tan Y, Pang W, Chakraborty AK, Lloyd-Smith JO, Sun R, Zhang J. 2020. Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLOS Genetics* **16**:e1009009. DOI: <https://doi.org/10.1371/journal.pgen.1009009>, PMID: 33085662
- Zhou T**, Georgiev I, Wu X, Yang Z-Y, Dai K, Finzi A, Do Kwon Y, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD. 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**:811–817. DOI: <https://doi.org/10.1126/science.1192819>, PMID: 20616231
- Zhou T**, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, Joyce MG, Lingwood D, Soto C, Bailer RT, Ernandes MJ, Kong R, Longo NS, Louder MK, McKee K, O'Dell S, Schmidt SD, Tran L, Yang Z, Druz A, et al. 2015. Structural repertoire of HIV-1-Neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell* **161**:1280–1292. DOI: <https://doi.org/10.1016/j.cell.2015.05.007>, PMID: 26004070
- Zwick MB**, Labrijn AF, Wang M, Spenlehauer C, Saphire EO, Binley JM, Moore JP, Stiegler G, Katinger H, Burton DR, Parren PWHI. 2001. Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus Type 1 glycoprotein gp41. *Journal of Virology* **75**:10892–10905. DOI: <https://doi.org/10.1128/JVI.75.22.10892-10905.2001>, PMID: 11602729

Appendix 1

Appendix 1—table 1. Selection and mutational target size for escape-mediating sites against each bNAb.

Shown are the sites (column 1) and the susceptible and the escape amino acids (column 2) for each bNAb. We called patterns for 10–1074 and CD4 binding site targetting antibodies VRC01 and 3BNC117 using genetic trial data and the remainder using DMS data. The inferred mutational target size of escape at each site (forward μ and backward μ^\dagger mutation rates) is shown in column 3. The major quantiles (10%, 50% (median), 90%) associated with the inferred site-specific Bayesian posterior of the scaled selection strength $\hat{\sigma} = \sigma/\theta_{ts}$ are shown in column 4. The corresponding quantiles for the strength of selection, $\sigma = \hat{\sigma}\theta_{ts}$ after convolving the posterior for the scaled selection $\hat{\sigma}$ with the reservoir-corrected intra-patient diversity of HIV-1 $\xi\theta$ are shown in column 5.

bNAb	site	susceptible AA	escape AA	μ	μ^\dagger	σ/θ quantiles			σ quantiles		
						0.1	0.5	0.9	0.1	0.5	0.9
10–1074	325	DN	EGK	0.76	0.51	560	1,100	2,900	8.8	29	93
	332	N	DHIKSTY	2.8	0.4	190	290	450	2.6	7.1	16
	334	S	AFGINRY	1.3	0.44	100	170	250	1.4	4	9.5
10E8	671	KNS	RT	0.41	0.51	38	110	220	0.68	2.5	7.3
	673	F	LV	1.4	0.47	150	290	510	2.4	7	18
3BNC117	279	DN	AHK	0.33	0.22	190	470	930	3.4	11	31
	281	AP	TV	1.1	1.1	62	140	250	0.97	3.2	8.6
	282	KY	ENR	1.2	0.8	120	230	400	1.9	5.6	14
	459	G	DN	0.5	1	54	140	270	0.91	3.3	9.5
PG9	160	N	KY	0.4	0.2	150	280	480	2.3	6.7	17
	162	ST	AIP	1.2	1.1	540	1600	2600	9.5	34	98
	169	GIKMRWW	ELT	0.53	0.92	240	590	1200	4.1	14	41
PGT121	171	KPR	AEGHNQST	1.4	0.62	100	190	310	1.5	4.5	11
	330	FHLRSY	Q	0.12	1.4	7.4	160	530	0.15	3.6	16
	332	AENV	DIKT	1.1	1.2	80	180	320	1.3	4.2	11
PGT145	334	DS	GN	1	2	15	80	170	0.26	1.8	5.9
	121	K	E	1	1	34	92	170	0.58	2.1	5.8
	160	N	KY	0.4	0.2	140	250	440	2.1	6.2	15
	162	ST	AIP	1.2	1.1	680	1200	2000	11	29	73
PGT151	166	KR	AEGST	0.73	0.58	58	120	200	0.9	2.8	7.2
	169	GIKLTV	EMRW	0.59	1.4	8.8	80	190	0.16	1.8	6.1
	512	A	GT	1.1	0.57	860	2,700	4100	16	59	160
	611	GN	DS	1.3	2	360	890	1300	6.5	20	50
	613	ST	CN	0.28	0.7	1200	2,800	3500	21	58	140
VRC01	637	DN	EKST	0.83	0.41	120	210	350	1.8	5.1	13
	639	T	IM	1	1	1100	1600	2600	16	42	96
	197	DN	S	0.5	1	730	1600	2300	12	35	88
	279	DN	AHK	0.33	0.22	200	420	940	3.2	10	30
	280	N	D	1	1	440	1000	1800	7.3	24	63
	281	AP	TV	1.1	1.1	52	130	250	0.9	3.1	8.6
	458	G	D	0.5	1	290	1200	2200	6	26	75

Appendix 1—table 1 Continued on next page

Appendix 1—table 1 Continued

bNAb	site	susceptible AA	escape AA	μ	μ^\dagger	σ/θ quantiles			σ quantiles		
						0.1	0.5	0.9	0.1	0.5	0.9
VRC34	512	A	GT	1.1	0.57	720	1400	2900	11	34	92
	524	GP	AERS	1.5	0.67	45	95	150	0.71	2.2	5.6
	88	N	K	0.26	0.26	570	1400	1900	9.8	30	74
	90	ST	AEK	0.48	0.8	1400	2500	3300	19	57	130

Appendix 2

Fokker-Planck description of the birth-death process

The individual-based birth-death model introduced above specifies the stochastic dynamics of a population state over time. We characterize the state of a population by vector $\mathbf{n} = (n_1, \dots, n_M)$, where n_a is the number of type a variants within the population. Using the concept of chemical reactions, suitable for Gillespie algorithm [Wilkinson, 2019](#); [Gillespie, 2002](#), we can determine the propensity $a_r(\mathbf{n})$ for a given reaction r (i.e., birth, death, or mutation) in a population of state \mathbf{n} , which in turn determines the rate at which the reactions occur ([Equation 9](#)). We denote the resulting change in the state of a population due to reaction r by ν_r . Taken together, the impact of the reactions in the birth-death model can be summarized as, Reaction Rate parameter Propensity State change

Reaction	Rate parameter	Propensity $a_r(\mathbf{n})$	State change ν_r
Birth	$\beta_a = \frac{\lambda + (f_i - \phi)}{2}$	$n_a \beta_a$	$+\hat{e}_i$
Death	$\delta_a = \frac{\lambda - (f_i - \phi)}{2}$	$n_a \delta_a$	$-\hat{e}_i$
Mutation	$\mu = \mu_{a \rightarrow b}$ ($\mu^\dagger = \mu_{b \rightarrow a}$)	$n_a \mu_{a \rightarrow b}$	$+\hat{e}_b - \hat{e}_a$

where \hat{e}_i is a vector of size M equal to size of the population state vector, in which the i^{th} element equal to one and the rest are zero. For example, a mutation reaction $\mu(a \rightarrow b)$ destroys a variant a and creates a variant b , resulting in the following change in the state vector,

$$\nu_{\mu(i \rightarrow j)} = -\hat{e}_i + \hat{e}_j \tag{44}$$

The reactions in [Equation 43](#) specify a Master equation for the change in the probability of the population state $p(\mathbf{n})$,

$$\dot{p}(\mathbf{n}) = \sum_r a_r(\mathbf{n} - \nu_r) p(\mathbf{n} - \nu_r) - a_r(\mathbf{n}) p(\mathbf{n}) \tag{45}$$

where $\mathbf{n} = \sum_i n_i \hat{e}_i$ is the state vector. Using a Kramers-Moyal expansion [Risken, 1989](#), we arrive at a Fokker-Planck approximation for the change in the probability distribution of the population state $p(\mathbf{n})$,

$$\frac{d}{dt} p(\mathbf{n}) = \left[\sum_r \left(\frac{1}{2} \sum_{i,j} \frac{\partial}{\partial n_i} \frac{\partial}{\partial n_j} \nu_r^i \nu_r^j a_r(\mathbf{n}) - \sum_i \frac{\partial}{\partial n_i} \nu_r^i a_r(\mathbf{n}) \right) \right] p(\mathbf{n}) \tag{46}$$

We can identify the drift (i.e. the deterministic force) and diffusion tensors of the Fokker-Planck operator:

$$\mathbf{b}(\mathbf{n}) = \sum_r \nu_r a_r(\mathbf{n}) \quad \Sigma(\mathbf{n}) = \sum_r \nu_r^2 a_r(\mathbf{n}) \tag{47}$$

To better demonstrate the structure of this birth-death operator, consider a bi-allelic case (e.g. susceptible and resistant) with a two-dimensional state vector, $\mathbf{n} = \begin{pmatrix} n_0 \\ n_1 \end{pmatrix}$. The drift and the diffusion tensors associated with this process follow,

$$\mathbf{b}(\mathbf{n}) = \begin{pmatrix} f_0 - \frac{n_0 f_0 + n_1 f_1}{N_k} - \mu & \mu^\dagger \\ \mu & f_1 - \frac{n_0 f_0 + n_1 f_1}{N_k} - \mu^\dagger \end{pmatrix} \cdot \mathbf{n} \tag{48}$$

$$\Sigma(\mathbf{n}) = \lambda \begin{pmatrix} n_0 & 0 \\ 0 & n_1 \end{pmatrix} + \mathcal{O}(\mu) \tag{49}$$

Note that in [Equation 49](#) we neglect the stochasticity due to mutations since the magnitude of the associated noise is much smaller than the noise due to the birth and death (i.e., genetic drift). From this Fokker-planck equation we directly recover the equilibrium distribution which links the parameters of our model to the genetic observables.

In the steady state, the population is fluctuating around carrying capacity $\sum_a n_a \approx N_k$ and we can represent the population state via allele frequencies $x_a = n_a/N_k$. In the simple case of a bi-allelic problem, the equilibrium allele frequency distribution $P_{\text{eq}}(x)$ follows the Wright-equilibrium distribution **Crow and Kimura, 2010** with modified rates,

$$P_{\text{eq}}(x) = \frac{1}{Z} \frac{e^{\frac{2N_k}{\lambda}(f_1 - f_0)x} (1-x)^{\frac{2N_k}{\lambda} \mu^\dagger} x^{\frac{2N_k}{\lambda} \mu}}{(1-x)x} \equiv \frac{1}{Z(\sigma, \theta, \theta^\dagger)} \frac{e^{-\sigma x (1-x)^{\theta^\dagger} x^\theta}}{(1-x)x} \quad (50)$$

Appendix 3

Incomplete escape of HIV-1 from bNAbs

In our model, we assume that during therapy the susceptible HIV-1 sub-population is completely neutralized by bNAbs and is removed at an exponential rate from the viral population. On the other hand, the resistant variants grow in accordance with their intrinsic fitness in the presence of bNAbs. However, this binary categorization is an approximation, and binding kinetics between viral epitopes and bNAbs implies that neutralization is a probabilistic process and could be incomplete, resulting in the removal of only a fraction of the susceptible variants. The degree of this neutralization depends on the concentration of the infused bNAbs in a patient's serum, and the binding affinity between bNAbs and the susceptible viral epitopes. In a simple model of binding kinetics, the binding affinity is characterized by the half maximal inhibitory concentration, or IC_{50} . The probability that a (susceptible) viral variant i does not bind to a neutralizing bNAb with serum concentration $[A]$ is given by,

$$u_i([A]) = \frac{1}{1 + \frac{[A]}{IC_{50}^i}}. \quad (51)$$

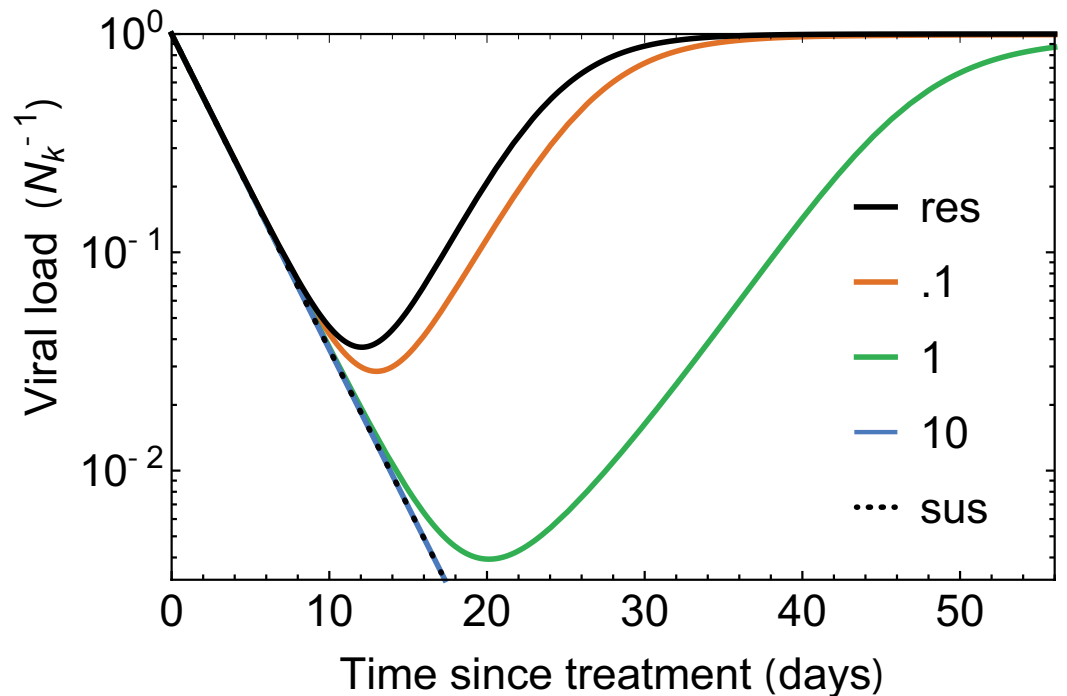
where IC_{50}^i is the half maximal inhibitory concentration of the bNAb against viral variant i . We model the growth rate $\gamma_i([A])$ of the susceptible variant i in the presence of a neutralizing bNAb with concentration $[A]$ as a weighted sum of the viral growth rate in the absence of bNAbs γ_i^0 , and the decay rate r of the virus due to neutralization; the respective weights are set according to the probability that a viral variant is bound (or unbound) to a bNAb (**Equation 51**), which results in,

$$\gamma_i([A]) = u_i([A])\gamma_i^0 + (1 - u_i([A]))r. \quad (52)$$

In the course of a trial, the concentration of an infused bNAb decays exponentially $[A](t) = [A]_0 e^{-\frac{t}{\tau_a}}$, with a characteristic time of about $\tau_a \simeq 20$ days **Stephenson et al., 2021**; here $[A]_0$ is the concentration of the bNAb upon infusion. By combining **Equation 52** with the bNAb's exponential decay over time, we find that the growth rate of variant i over time is given by

$$\gamma_i(t) = \frac{1}{1 + \frac{[A]_0}{IC_{50}^i} e^{-\frac{t}{\tau_a}}} (\gamma_i^0 - r) + r. \quad (53)$$

The growth rate of the susceptible variant is determined by the ratio $[A]_0/IC_{50}^i$ between the initial antibody concentration and the half maximal inhibitory concentration of the bNAb.



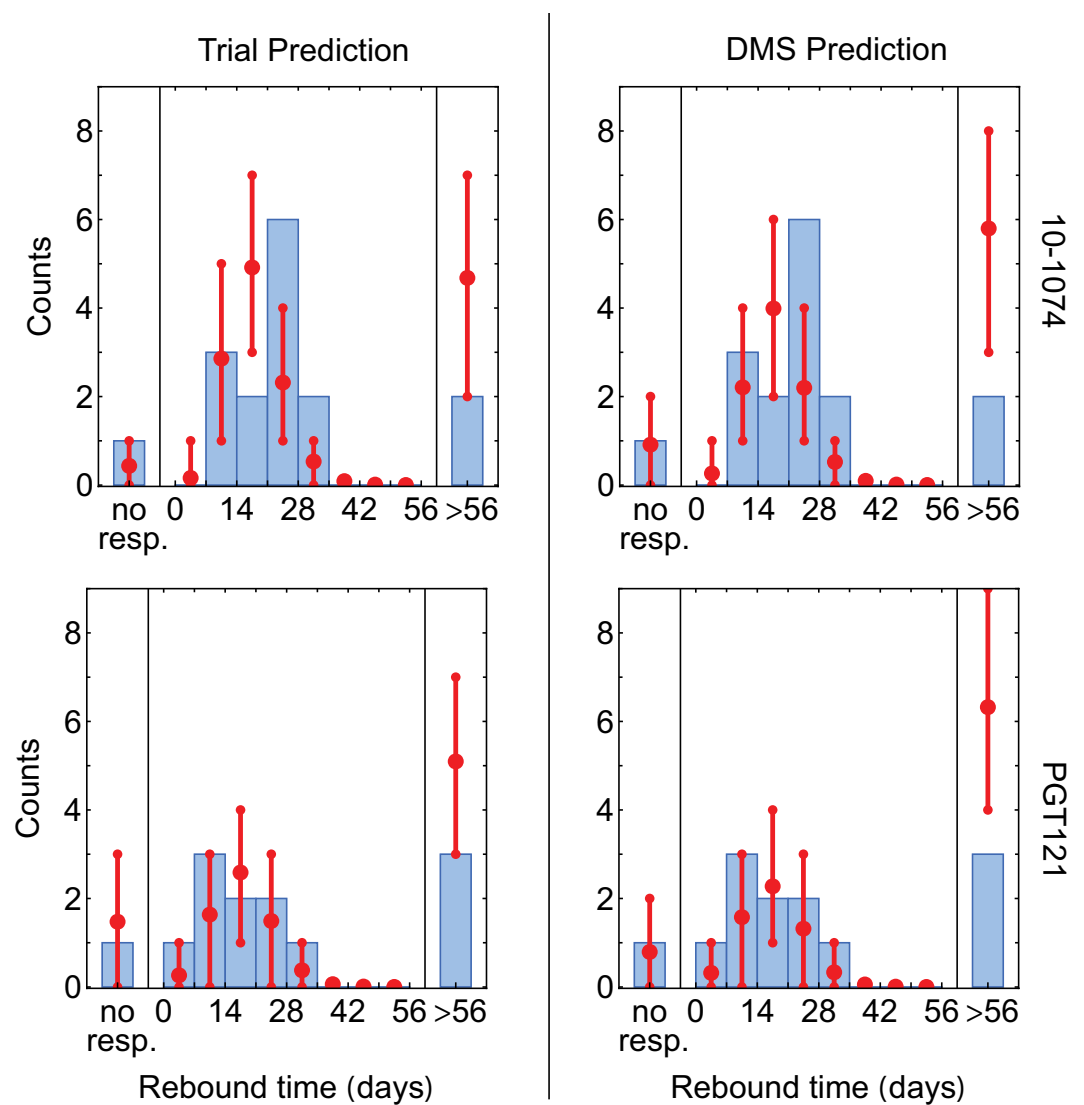
Appendix 3—figure 1. The effect of incomplete neutralization. Viral load curves, simulated under the noise-averaged model (deterministic limit) **Equation 4**, are plotted under the assumption of incomplete neutralization for different values of $[A]_0/IC_{50}$ (colors); incomplete neutralization is defined by a changing growth rate given by **Equation 53**. Small values of $[A]_0/IC_{50} \ll 1$ (red) lead to similar trajectories to the perfectly resistant variant (solid black line), while large $[A]_0/IC_{50} \gg 1$ (green line) is indistinguishable from the fully susceptible variants (dotted black line). For all curves, the escaping virus has an initial frequency $x_{res} = 10^{-3}$, with decay rate r and free growth rate γ^0 , $r = \gamma^0 = 1/3$ days.

As shown in **Appendix 3—figure 1**, if the initial antibody concentration is an order of magnitude below the IC_{50} (orange line), the dynamics of a patient's viremia closely resembles that of a fully resistant population (solid black line). On the other hand, if the initial bNAb dosage is an order of magnitude above the IC_{50} (blue line), the dynamics behaves as though the viral population is completely susceptible (dotted line), resulting in a late viral rebound (more than 8 weeks). We find that incomplete neutralization is most relevant when the initial bNAb concentration is roughly similar to IC_{50} , that is, $[A]_0/IC_{50} \approx 1$ (green line). Therefore, our binary categorization is applicable so long as the distribution of IC_{50}^i has low density around initial serum concentrations $[A]_0$. The likelihood of IC_{50}^i matching the antibody dose determines the rate at which our results will be biased by incomplete neutralization.

To compare IC_{50} with the initial bNAb concentration, we use the neutralization data from the 10–1074 bNAb trial **Caskey et al., 2017**. For this trial, TZM-bl neutralization assays against different bNAbs were obtained on 114 pseudoviruses expressing envelope proteins derived from circulating viruses in patients on day 0 (55 pseudoviruses) and week 4 (59 pseudoviruses) after 10–1074 infusion. Expectantly, a fraction of viruses from week 4 were resistant (i.e. large IC_{50}) to the 10–1074 bNAb used in the trial (orange histogram in **Appendix 3—figure 2A**). However, almost all viruses were susceptible to the (control) 3BNC117 bNAb, which the viruses had not been previously exposed to (orange histogram in **Appendix 3—figure 2B**).

For comparison, the distribution for the initial concentration $[A]_0$ (or equivalently the maximum concentration) of the infused 10–1074 bNAb in patients enrolled in this trial is shown in **Appendix 3—figure 2** (blue histograms), and is peaked around $200 \mu\text{g/ml}$. The IC_{50} values in this trial are much lower (higher) for susceptible (resistant) variants compared to the initial bNAb concentrations. Therefore, our simplified model assuming that a viral variant is either fully resistant or susceptible to a bNAb (i.e. no incomplete escape) is a reasonable approach for capturing the statistics of treatment failure at the concentrations tested in these trials. Nonetheless, developing a genotype-to-neutralization

model such as the ones developed by *Wagh et al., 2016; Yu et al., 2019*, may allow for a more nuanced approach to modeling of neutralization in future work.



Appendix 3—figure 2. The distribution of IC_{50} and $[A]_0$. The distribution of IC_{50} for envelope proteins derived from circulating viruses in patients after 10-1074 infusion is shown against (A) 10-1074, (B) 3BNC117 as a control. For comparison, the distribution of the maximum (initial) 10-1074 bNAb concentration used in the trial is shown in blue. The maximum measurable $IC_{50} = 50 \mu\text{g/ml}$ associated with the TZM-bl neutralization assays used in these experiments is indicated in each panel.

Appendix 4

Comparison of DMS-inferred and trial-inferred escape pathways

In this appendix we compare the sensitivity of our rebound time predictions to the methodology used to identify sites of escape (i.e., data from DMS experiments versus therapy trials). To do so, we focus on two bNAbs, 10–1074 and PGT121 for which we have access to both the trial [Caskey et al., 2017](#); [Stephenson et al., 2021](#) and the DMS data [Dingens et al., 2019](#); [Dingens et al., 2017](#); [Schommers et al., 2020](#). The inferred resistant and susceptible variants at the identified sites for both of these bNAbs based on their respective trials and the DMS data are reported in [Appendix 4—table 1](#).

Although there is a lot of commonalities among the two approaches, the two methods can be sensitive to different mutational pathways in certain sites. This is in part expected as the DMS procedure can generate many rare variants, but all on a common genetic background; in this case the experiments scan only single point mutations away from the HIV-1 strain BF520.W14M.C2 [Dingens et al., 2017](#). On the other hand, the clinical trial data shows escape variants on the genetic background of a diverse viral population circulating in a patient, and the fate of a mutation can be strongly determined by the epistatic interactions with the background genome that it appears on.

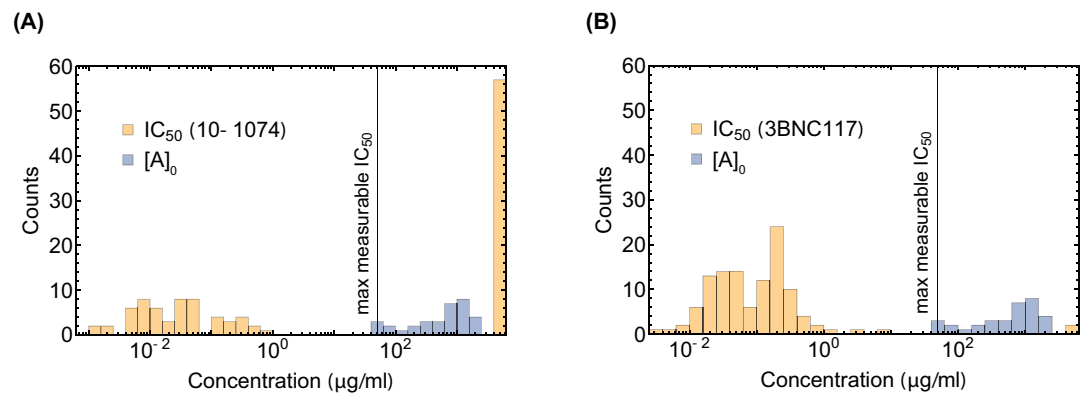
Appendix 4—table 1. Escape variants inferred from trial and DMS data.

Tables show the sites mediating escape of HIV-1 from the 10–1074 and the PGT121 bNAbs, using the respective trial data [Caskey et al., 2017](#); [Stephenson et al., 2021](#) (left) and the DMS data [Dingens et al., 2017](#) (right). The susceptible (sus) and resistant (res) variants at each site are identified according to the procedure detailed in the Methods. Sites that are identified as escape-mediating by only one of the methods are marked with (*). While substitutions at site 325 were called as escape mediating based on the PGT121 trial data [Stephenson et al., 2021](#) but are missed by the DMS data, this site was only inconsistently associated with escape, and its associated escape variants usually co-occur with other escape substitutions; see the Extended Data Figure 3 in [Stephenson et al., 2021](#). Sites are numbered according to HXB2.

10–1074 trial			10–1074 trial		
HXB2 site	sus	res	HXB2 site	sus	res
325	DN	EGK	325	DN	EGK
330*	FHLQSYR	*1,731	330	FHLQSYR	R
332	N	DHIKSTY	332	N	ADEIKTV
334	S	AFGINRY	334	S	DGN

PGT121 trial			PGT121 DMS		
HXB2 site	sus	res	HXB2 site	sus	res
325	D	NKT	325*	DNKT	*
330*	FHLQSYR	*1,733	330	FHLRSY	Q
332	NV	DSRTI	332	AENV	DIKT
334	TS	RND	334	DS	GN

To assess the sensitivity of our analyses, we make predictions for the distribution of rebound times, using the escape pathways identified based on the DMS versus the trial data in [Appendix 4—table 1](#). Overall, we see good agreements between the two approaches, but DMS-inferred escape pathways predict slightly too many patients with late rebound ($T_p > 56$ days) compared to the trial-based inference, i.e., DMS-based predictions are more optimistic ([Appendix 4—figure 1](#)). This deviation is likely to be related to the diversity of viral genetic backgrounds in clinical trials, since more escape pathways could be realized through positive epistasis, leading to a faster viral rebound.



Appendix 4—figure 1. Statistics of viral rebound with using escape pathways inferred from the DMS and the trial data. We compare the distribution of rebound times in patients from the clinical trials with 10–1074 **Caskey et al., 2017** (top) and PGT121 **Stephenson et al., 2021** (bottom) bNAbs, using escape pathways inferred from the respective trial data (left) and the DMS data **Dingens et al., 2017** (right). The error bars show the inter decile range (0.1–0.9 quantiles) generated by the simulations for the corresponding trial. The PGT121 trial predictions relied on the neutral diversity estimates θ_{ts} from the other three trials **Caskey et al., 2015**; **Caskey et al., 2017**; **Baron et al., 2018** due to the relatively limited genetic data available from this study. The fitted reservoir value of $r_{resv.} = 2.07$ is used for all predictions.

Appendix 5

Simulation algorithms

Algorithm 1. Population initialization

procedure Population Initialization ($N_k, \sigma_{1:L}, \theta_{1:L}, \theta_{1:L}^\dagger$)

▷ Generates a population of size of N_k from equilibrium. θ , and σ are the parameters defining the equilibrium values of the population state. Returns the initial vector of genotypes

```

for  $i \in 1 : L$  do
   $x_i \sim P_{\text{eq}}(x|\sigma_i, \theta_i, \theta_i^\dagger)$ 
end for
for  $v \in 1 : N_k$  do  $\rho_i^v \sim \text{Bernoulli}(x_i)$ 
end for
return  $\rho^{1:N_k}$ 
end procedure
```

Algorithm 2. Gibbs Sampler for Allele Frequencies

procedure Equilibrium Sampler ($\sigma, \theta, \theta^\dagger$ | Samples)

▷ Generates a stream of non-independent but rapidly mixing samples $X \sim p(x|\sigma, \theta, \theta^\dagger)$. Default BurnIn is 10.

```

 $N \leftarrow \text{Samples} + \text{BurnIn}$ 
 $K_0 \leftarrow \text{Round}(\sigma)$ 
for  $n \in 1 : N$  do
   $X_{n+1} \sim \text{Beta}(\theta, \theta^\dagger + K_n)$ 
   $K_{n+1} \sim \text{Poisson}(\sigma(1X_{n+1}))$ 
end for
return  $X_{\text{BurnIn}:N}$ 
end procedure
```

▷ Sample the mutant fraction
▷ Sample the auxiliary parameter
▷ Return only the mutant frequency X part of the chain (marginalize over k)

Algorithm 3. Population time step

procedure EvolvePopulation ($t, \rho^{1:N}|\lambda, \gamma, r_s$)

▷ Acts on a time t and a list of N genotypes $\rho^{1:N}$. Inherits dependency on other parameters from the fitness function $F(G)$ and the Mutate(G) operator which depend on $\Delta_{1:L}, \mu_{1:L}, \mu_{1:L}^\dagger$ and γ and the population diversity measure θ_{IS} .

```

 $\phi \leftarrow \frac{1}{N} \sum_i F(g_i)$ 
 $t' \leftarrow t + \frac{\text{RandExp}()}{\lambda N}$ 
 $i \sim \text{Rand}(1 : N)$ 
 $G \leftarrow \rho^i$ 
if IsEscaped( $G$ ) then
   $D \sim \text{Bernoulli}(\frac{\lambda - F(G) + \phi}{2\lambda})$ 
else
   $D = \text{false}$ .
  if  $D$  then
     $N' \leftarrow N - 1$ 
     $\rho^{1:N'} \leftarrow \rho^{1:\bar{i}:N}$ 
  else
     $N' \leftarrow N + 1$ 
     $\rho^{1:N'} \leftarrow \text{Append}(\rho^{1:N}, G)$ 
  end if
  if  $D$  then
     $D \sim \text{Bernoulli}(\frac{r}{\lambda})$ 
    if  $D$  then
       $N' \leftarrow N - 1$ 
       $\rho^{1:N'} \leftarrow \rho^{1:\bar{i}:N}$ 
    end if
  end if
   $j \sim \text{Rand}(1 : N')$ 
   $\rho^j \leftarrow \text{Mutate}(\rho^j)$ 
end procedure
```

▷ If the virus is escaped
▷ Determine if the virus dies ($D = \text{true}$) or lives
▷ Delete genotype at position
▷ Duplicate genotype at position
▷ If the virus is neutralized
▷ remove it at the appropriate rate
▷ Delete genotype at position
▷ Choose a random virus to mutate
▷ Apply mutation operator with intensity μ/λ
▷ Return the new time and the new population.

Inference algorithms

Algorithm 4. Importance sampled log-likelihood given a single datapoint

```

procedure SigmaLikelihood ( $\hat{\mu}, \hat{\mu}^\dagger, m, s, \theta_{ts}, N$ )
  ▶ Takes mutant  $m$  and susceptible  $s$  counts for a
  particular site at a single timepoint. Returns an approximate log-likelihood function  $l(\hat{\sigma}) = \log P(m, s | \theta_{ts}, \hat{\sigma}) + c$ ,
  up to an additive constant. SigmaLikelihood is a closure that returns a one-parameter function. We used  $N$ 
  =103 samples.
   $\theta \leftarrow \hat{\mu}\theta_{ts}$ 
   $\theta^\dagger \leftarrow \hat{\mu}^\dagger\theta_{ts}$ 
  for  $i \in 1 : N$  do
     $x_i \sim \text{Beta}(\theta, \theta^\dagger)$  ▶ Sample from the neutral distribution
     $w_i \leftarrow \frac{\mathcal{B}(\theta+m, \theta^\dagger+s)}{\mathcal{B}(\theta, \theta^\dagger)} \frac{1}{x_i^m(1-x_i)^s}$  ▶ Importance weight ratio
     $y_i \sim \text{Beta}(\theta + m, \theta^\dagger + s)$  ▶ sample from the neutral distribution,
    conditioned on the observations
     $v_i \leftarrow \frac{\mathcal{B}(\theta+m, \theta^\dagger+s)}{\mathcal{B}(\theta, \theta^\dagger)} \frac{1}{y_i^m(1-y_i)^s}$  ▶ Importance weight ratio
  end for
   $Z_0(\hat{\sigma}) := \frac{1}{N} \sum_i e^{\hat{\sigma}\theta_{ts}x_i} \frac{1}{1+w_i} + \frac{1}{N} \sum_i e^{-\hat{\sigma}\theta_{ts}y_i} \frac{1}{1+v_i}$  ▶ importance sampling mean
   $Z_1(\hat{\sigma}) := \frac{1}{N} \sum_i e^{\hat{\sigma}\theta_{ts}x_i} \frac{1}{1+w_i^{-1}} + \frac{1}{N} \sum_i e^{-\hat{\sigma}\theta_{ts}y_i} \frac{1}{1+v_i^{-1}}$  ▶ Note the inversion in the weighting
  factor compared to  $Z_0$ .
  return  $l(\hat{\sigma}) := \log Z_1(\hat{\sigma}) - \log Z_0(\hat{\sigma})$  ▶ Return a log-likelihood function. The same random variable realizations
   $x_{1:N}$  and  $y_{1:N}$  are cached in memory and used for each function evaluation, making  $l(\hat{\sigma})$  continuous and
  differentiable.
end procedure

```

Algorithm 5. Rebound-time Disparity

```

procedure ReboundDisparity ( $t_{1:P}, T_{1:S}$ )
  ▶ Takes the observed rebound times  $t_{1:P}$  and simulated late rebound times
   $T_{1:S} \sim P$  and returns a disparity estimator.
  ▶ First estimate the probabilities in the truncated observation categories
   $P_{(NR)} \leftarrow \frac{1}{P} \sum_p [t_p < 1\text{day}]$  ▶ Count the fraction of non-responders in trial
   $Q_{(NR)} \leftarrow \frac{1}{S} \sum_s [T_s < 1\text{day}]$  ▶ ... and in simulation
   $P_{(LR)} \leftarrow \frac{1}{P} \sum_p [t_p \geq 56\text{days}]$  ▶ Count the fraction of late rebounds in trial
   $Q_{(LR)} \leftarrow \frac{1}{S} \sum_s [T_s \geq 56\text{days}]$  ▶ ... and in simulation
  ▶ Then construct a histogram over the continuous data-points (i.e.  $t \in [0, 56]$ ) and estimate the probability in
  each bin
   $t_{1:P'} = \text{SortAscending}(\text{Filter}[1 \leq t < 56](t_{1:p}))$  ▶ Select only the observed rebound times
   $t_0 \leftarrow -\infty$  and  $t_{P'+1} \leftarrow \infty$ 
  for  $p \in 1 : P'$  do
     $Q_{(p)} \leftarrow \frac{1}{P}$  ▶ By design, each histogram bin contains one observed data point, and gets  $1/P$  mass
     $P_{(p)} \leftarrow \frac{1}{S} \sum_s [\max(1, \frac{1}{2}(t_{p-1} + t_p)) \leq T_s < \min(56, \frac{1}{2}(t_{p+1} + t_p)) \text{days}]$  ▶ Use the midpoints of
    adjacent points to construct the boundaries of histogram bins, and determine probability mass in each bin.
  end for
  return  $P \sum_{i \in NR, LR, 1:P'} (Q_{(i)}^{1/2} - P_{(i)}^{1/2})^2$  ▶ Return the discretized estimate of the Hellinger distance, scaled by
  the number of patients
end procedure

```