

Mild HIV-specific selective forces overlaying natural CD4+ T cell dynamics explain the clonality and decay dynamics of HIV reservoir cells

Daniel B. Reeves^{1,2,*}, Danielle N. Rigau³, Arianna Romero³, Hao Zhang⁴, Francesco R. Simonetti³, Joseph Varriale³, Rebecca Hoh⁵, Li Zhang⁶, Kellie N. Smith⁶, Luis J. Montaner⁷, Leah H. Rubin³, Stephen J. Gange³, Nadia R. Roan^{5,8}, Phyllis C. Tien⁵, Joseph B. Margolick⁴, Michael J. Peluso⁵, Steven G. Deeks⁵, Joshua T. Schiffer^{1,9}, Janet D. Siliciano³, Robert F. Siliciano^{3,10}, Anukka A. R. Antar^{3,11,**}

¹Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Center, Seattle WA 98144 USA

²Department of Global Health, University of Washington, Seattle WA 98144 USA

³School of Medicine, Johns Hopkins University, Baltimore, MD 21205 USA

⁴Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205 USA

⁵School of Medicine, University of California San Francisco, San Francisco, CA 94158 USA

⁶Bloomberg Kimmel Institute for Cancer Immunotherapy, Johns Hopkins University, Baltimore, MD 21205 USA

⁷The Wistar Institute, Philadelphia, PA 19104 USA

⁸Gladstone Institutes, San Francisco, CA 94158 USA

⁹Department of Medicine, University of Washington, Seattle WA 98144 USA

¹⁰Howard Hughes Medical Institute, Baltimore, Maryland, USA

¹¹Lead contact

*Correspondence: dreeves@fredhutch.org

**Correspondence: antar1@jhmi.edu

Summary. The latent reservoir of HIV persists for decades in people living with HIV (PWH) on antiretroviral therapy (ART). To determine if persistence arises from the natural dynamics of memory CD4+ T cells harboring HIV, we compared the clonal dynamics of HIV proviruses to that of memory CD4+ T cell receptors (TCR β) from the same PWH and from HIV-seronegative people. We show that clonal dominance of HIV proviruses and antigen-specific CD4+ T cells are similar but that the field's understanding of the persistence of the less clonally dominant reservoir is significantly limited by undersampling. We demonstrate that increasing reservoir clonality over time and differential decay of intact and defective proviruses cannot be explained by mCD4+ T cell kinetics alone. Finally, we develop a stochastic model of TCR β and proviruses that recapitulates experimental observations and suggests that HIV-specific negative selection mediates approximately 6% of intact and 2% of defective proviral clearance. Thus, HIV persistence is mostly, but not entirely, driven by natural mCD4+ T cell kinetics.

Keywords: HIV cure, HIV reservoir, latent reservoir, mathematical modeling, ecology, TCR, clonal, antigen-specific, CD4+ T cells, reservoir decay

Introduction.

HIV persists in people living with HIV (PWH) on antiretroviral therapy (ART) as proviruses integrated primarily in the genomes of CD4+ T cells¹⁻³. Most proviruses are genome defective, i.e., are not replication competent or intact^{4,5}. HIV proviruses that are replication competent comprise the HIV latent reservoir^{6,7}, which persists for decades⁸⁻¹⁴ and can seed rebound viremia if ART is stopped for any reason¹⁵⁻¹⁸. Thus, the reservoir is the major barrier to cure of HIV^{19,20}.

There is strong evidence that reservoir persistence in blood and tissues during ART is due to cellular proliferation of infected CD4+ T cells. Heavy-water (deuterium) labeling experiments in PWH on ART have demonstrated that median CD4+ T cell half-lives are generally much shorter than the half-life of intact proviruses *in vivo*^{5,23-25}, suggesting that proliferation of CD4+ T cells repopulates the reservoir. Providing direct evidence of this, replication competent proviruses with identical genomes and identical integration sites have been isolated from sister cells of clonal populations²⁶⁻³³. Model-based extrapolations of HIV proviral sequence data suggest that reservoirs are extremely clonal – after one year on ART, over 99% of CD4+ T cells harboring proviruses have divided at least once and thus are members of a clonal population³⁴. These clones are not all the same size, however. Observational and modeling work demonstrate that the reservoir contains a small number of highly expanded or dominant clones and a massive number of smaller clones^{34,35}.

The frequency of cells harboring replication competent^{8,9} proviruses among all CD4+ T cells declines over the first decade after ART initiation¹⁰⁻¹⁴. Cells harboring defective proviruses also decline, but at a slower rate^{5,10,13,14,36}. Therefore, some cells disappear and what remains in the reservoir has been observed to become increasingly concentrated in highly expanded clones over time on ART^{13,14,27,37,38}. Modeling of these observational data suggests that after a decade on ART, over half of an individual's remaining proviruses are members of the top 100-1000 most-highly expanded infected clones¹³.

It remains unknown whether the natural drivers of CD4+ T cell proliferation and death in humans (e.g., antigen stimulation and homeostatic proliferation among others) entirely account for the extraordinary clonality of HIV proviruses and their decay dynamics. Antigen-driven clonal proliferation of HIV-infected cells could certainly account for many or most of the highly expanded proviral clones³⁹⁻⁴¹. The role of homeostatic proliferation in the maintenance of the reservoir is less clear. Factors related to provirus sequence^{42,43} or integration site^{26,27,37,44,45} may confer selective advantages or disadvantages to certain proviral clones and therefore contribute to the reservoir's extraordinary clonality. While provirus integration into oncogenes is now thought to play only a minor role in reservoir clonal expansion⁴⁶, there is evidence that integration into transcriptionally inactive heterochromatin may confer a survival advantage^{44,47-50}. The largest infected clones in PWH on ART with low-level viremia are more consistently present in short-term sampling compared to uninfected memory CD4s of the same size, perhaps reflecting different proliferative drivers or HIV-specific selective advantages³⁵.

To test whether the natural physiology of memory CD4+ T cells (e.g., HIV as inert “passenger” hypothesis) is sufficient to explain reservoir clonality and dynamics or whether selective factors related to HIV infection itself are required to explain these observations, we generated longitudinal clonality data from memory CD4+ T cells in HIV seronegative individuals and from HIV proviruses and CD4s simultaneously from PWH on ART. Specifically, we investigated whether the passenger hypothesis could explain three key observations: (1) the high degree of HIV proviral clonality in PWH on ART, (2) the increase in proviral clonality over time on ART, and (3) the differential decay of cells harboring intact and defective proviruses over time on ART. We compared the longitudinal (~10 year) clonality of HIV proviruses (intact and defective) from resting CD4+ T cells with the longitudinal clonality of their putative carrier cells, resting memory CD4+ (rmCD4) T cells from the same blood samples from PWH on ART. We also compared proviral clonality of the replication competent reservoir at single time points from PWH on long-term ART with the clonality of rmCD4s from the same blood samples. To determine whether suppressed HIV infection itself is associated with significantly different mCD4 clonality, we compared the longitudinal cohort of PWH on ART to HIV-seronegative people matched on age, race, and sex. We identified antigen specific (CMV and HIV) mCD4 cells in a subset of people and compared their clonality and decay rates with that of HIV-infected cells and all mCD4s. To make these comparisons, we modeled clonotype abundance

distributions and calculated clonality and diversity metrics. Finally, we developed a mechanistic mathematical model of CD4+ clonotypes in which some harbor HIV proviruses (intact and defective). We matched this model to our comprehensive experimental data and arrived at the conclusion that mild HIV-dependent forces such as negative selection on mostly intact HIV proviruses are required to account for the increasingly oligoclonal composition of the reservoir and its decay over time on ART.

Results

Participants and sampling

We sampled people enrolled in four cohorts: two longitudinal cohorts (4 PWH on ART and 4 HIV-seronegative people) and two cross-sectional cohorts (7 PWH on ART and 33 HIV-seronegative people) (**Table 1**, **Figure 1A**, detailed information in **Table S1 & Figure S1**).

1) *Longitudinal PWH on ART (PWH): HIV provirus vs rmCD4 clonality.* To investigate the clonality of HIV-infected cells, we examined near-full length (nfl) HIV-1 provirus sequences (intact and defective) from peripheral resting CD4+ T cells (mean 57 sequences per time point, IQR 46-73) from four PWH¹¹ who remained on ART for over ten years with no documented lapses in therapy after ART initiation and with undetectable viral loads (aside from 3 blips) checked generally every six months or more frequently (**Figure S1**, **Methods**). Each participant had 3 samples collected at an average of 10 months (t0), 2 years (t1) and 11 years (t2) after ART initiation (**Figure 1A-B**)¹³. Proviruses (nfl) were sequenced at t1 and t2, and rmCD4+ TCRβ repertoires were sequenced at t0, t1, and t2 (sorting schema in **Figure S2**).

2) *Longitudinal HIV-seronegative people (Control, Ctl): rmCD4 clonality.* As controls for our longitudinal PWH cohort, we obtained longitudinal samples from four confirmed HIV-seronegative people from the MACS/WIHS combined cohort study who were originally enrolled based on having similar risk behaviors as enrolled PWH⁵¹. These four were matched one-to-one to the four longitudinally sampled PWH on ART described above on age, race, sex, time between samples, and CMV seropositive status (with one exception, **Table S1**). rmCD4s were sorted and TCRβ repertoires were obtained.

3) *Cross-sectional PWH on ART (DELSPC, DEL): Reservoir clonality vs rmCD4 clonality.* To investigate the clonality of cells comprising the replication competent reservoir, clonality was approximated using *env* sequences from outgrowth viruses from quantitative viral outgrowth assays performed at limiting dilution (QVOAs) from one sample an average of 12 years (IQR 9-15 yr) after ART initiation from a separate cohort of 7 PWH on ART⁵². rmCD4s were sorted and TCRβ repertoires were obtained from the same samples.

4) *Cross-sectional HIV-seronegative people (De Neuter, DN): mCD4 clonality.* As controls for our cross-sectional PWH cohort, we obtained single-timepoint memory CD4+ TCRβ repertoires from a cohort of 33 HIV-seronegative individuals⁵³.

Table 1. Summary of longitudinal and cross-sectional cohorts. Detailed information for all participants is available in Supplementary Table 1.

Name	Type	HIV status	Time points	N	Sex	Age range at t0	t0	t1, t2	TCRβ cell type	HIV cell type (sequence)
PWH	Longitudinal	Pos	3	4	3M 1F	35-40	TCR	TCR/ HIV	rmCD4	rCD4 (nfl seq, total proviruses)
Control (Ctl)	Longitudinal	Neg	3	4	3M 1F	36-39	TCR	TCR	rmCD4	N/A

DELPS (DEL)	Cross-sectional	Pos	1	7	7M 0F	41-60	TCR/ HIV	N/A	rmCD4	rCD4 (<i>env</i> seq from QVOA+ wells)
De Neuter (DN)	Cross-sectional	Neg	1	33	11M 22F	21-50	TCR	N/A	mCD4	N/A

Abbreviations: PWH, people living with HIV; TCR β , T cell receptor β chain; Pos, positive; Neg, negative; N, number of participants; M, male; F, female; N/A not applicable; rmCD4, resting memory CD4+ T cells; mCD4, memory CD4+ T cells; rCD4, resting CD4+ T cells; nfl, near-full length HIV provirus; seq, sequences; QVOA+, quantitative viral outgrowth assay positive, meaning virions were produced; env, HIV envelope

Clonality of HIV-infected vs mCD4+ T cells

CD4+ T cell clonality (m or rmCD4) was approximated using TCR β DNA sequencing repertoires⁵⁴ (average 1.3×10^5 TCRs per sample, **Table S1**, sorting schema in **Figure S2, Methods**).

HIV-infected cell clonality in the longitudinal cohort was approximated by identifying cells harboring proviruses with identical nucleotide sequences and, if deleted, with identical deletion junctions, as has been done previously^{13,55,56}. The clonality of HIV reservoir cells in the cross-sectional PWH cohort was approximated by identifying outgrowth viruses harboring identical *env* sequences, as has been done previously⁵⁷. HIV sequence sample sizes are limited due to the rarity of HIV-infected cells in PWH on ART and the labor-intensive nature of near-full length provirus sequencing and viral outgrowth assays⁴. Because the vast majority of rmCD4s are not infected with HIV, these experiments are in effect comparing the clonality of HIV-uninfected CD4s to that of HIV-infected cells, as has been done previously³⁵.

rmCD4 and HIV clonality both follow a characteristic log-log linear rank-abundance distribution

We first used rank-abundance distributions to enable visual and semi-quantitative comparisons of clonality (**Figure 1C-D**). In these plots, each clonotype is “ranked” or ordered (x-axis) by its observed abundance (y-axis) from most highly expanded ($x=1$) to least. Clonotypes that have the same observed abundance are randomly ranked. A log-log scale emphasizes the large numbers of unique clonotypes on the x-axis (richness) as well as the orders of magnitude differences in abundances on the y-axis. rmCD4 proportional rank abundance curves generally followed a characteristic log-log linear shape across all participants living with and without HIV and across all time points, with varying slopes (**Figure 1C**). Similar to our previous observations^{13,34}, rank abundances for HIV sequences, including for replication competent sequences, also followed a similar overall shape, though curves are substantially less smooth because of lower sample sizes (**Figure 1D**).

Richness of circulating memory CD4+ cell TCR β clonotypes saturates with increasing sample size

Richness, a term for the number of unique clonotypes found in a sample, is one simple measure of clonality/diversity. Although mCD4 TCR β sample sizes were heterogeneous across individuals (**Table S1**), we observed that sequence richness did not increase indefinitely with sample sizes and instead approximates 10^4 across all PWH and HIV-seronegative person-time points (**Figure 1C inset**). This implies a general richness, or a count of unique clonotypes, of mCD4s on the order of 10^4 in sample sizes of less than 1 million total circulating mCD4s, in line with prior observations from HIV-seronegative people⁵⁸. In contrast, the richness of HIV sequences generally increased with increasing sample sizes (**Figure 1D inset**), implying that sampling depths of up to 100 HIV-infected cells per time point do not come close to detecting most HIV-infected clonotypes in the peripheral circulation.

Quantitative comparisons of clonality require resampled diversity metrics

Ecological diversity metrics have been used to summarize the clonality of HIV-infected cells in the HIV literature³⁴. Common metrics include richness, the Shannon index⁵⁹, a measure of clonal unevenness, and

the Simpson index⁶⁰, a measure of concentration into dominant clonal groups. Here, we use normalized Hill numbers, which are transformations of these 3 metrics into a mathematically unified and more interpretable family of indices (**Supplementary Methods**)⁶¹. Hence, we use the total number of unique clonotypes (Richness, D^0), the exponential of the Shannon index which approximates the relative evenness of all clonotypes (Evenness, D^1), and the inverse of the Simpson index which approximates the relative dominance of the largest clonotypes (Dominance, D^2).

These Hill numbers are also sensitive to sample sizes⁶². By assuming an underlying distribution that was more or less clonal, we showed normalized Hill numbers decrease when the same distribution is sampled with higher sample sizes. The sample size dependence also grew as the underlying true distribution became more clonal (**Figure S3**). Hence, to make all comparisons we computationally resample data to the same size (**Figure S3A**), and only then compute normalized Hill numbers on equally sampled data. In conclusion, we report Hill numbers after resampling equally, and then divide by the sample size such that each metric D^x/N is comparable across participant time points and ranges from 0 (highly clonal) to 1 (highly diverse).

Minimal differences in mCD4 cell TCR β clonality across cohorts

There were no significant differences in rmCD4 cell TCR β clonality between the PWH on ART and HIV-seronegative longitudinal cohorts that were matched for age, race, sex, and CMV status (**Figure 1E**). The cross-sectional PWH (DEL) cohort had generally more clonal rmCD4 cells than the cross-sectional HIV-seronegative (DN) cohort for all three diversity metrics, despite the fact that DN includes actively proliferating (non-resting) CD4+ T cells. This discrepancy could be due to differences in ages and health status between the cross-sectional cohorts. For instance, DEL participants are significantly older than DN participants (mean age 51 vs 37, $p=0.001$, 1 sided Mann-Whitney U test) (**Table S1**).

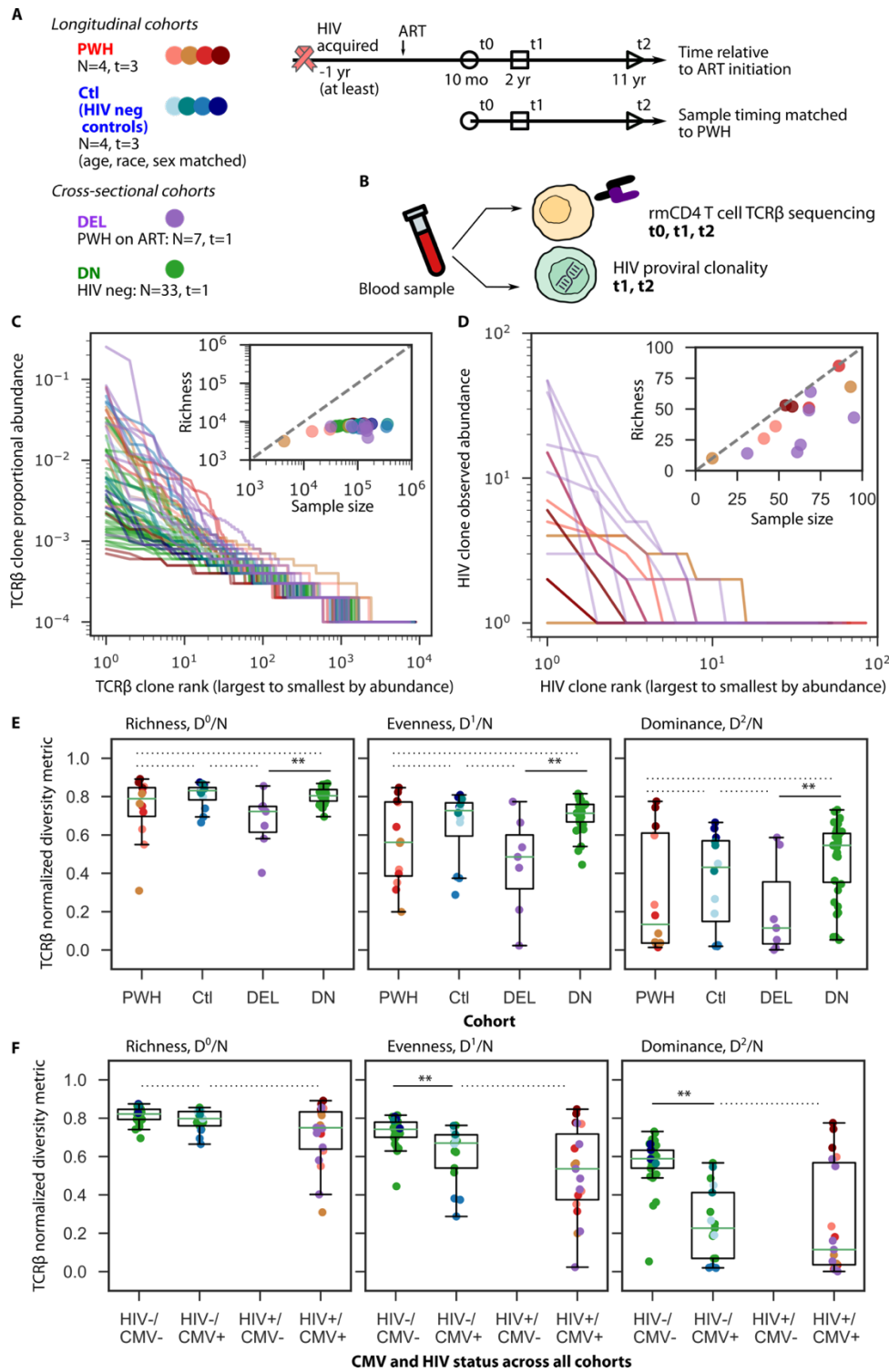


Figure 1. HIV-infected cell and resting memory CD4+ T cell clonality in PWH and HIV-seronegative individuals. A) Four PWH (Table 1) were sampled longitudinally at 3 time points on average 10 months, 2 years, and 11 years after ART initiation. Four age-, race-, and sex- matched HIV-seronegative individuals were sampled at the same intervals. Seven PWH on ART and 33 HIV-seronegative individuals were sampled at one timepoint. B) TCR β repertoires were used to assess clonality of mCD4 cells in all participants. HIV provirus sequences and env sequences from replication competent outgrowth viruses were used to assess

clonality of HIV in PWH. C) mCD4 clonality is shown as rank abundance curves, in which clonotypes are arranged from most to least abundant along the x axis (colors from **A** denote cohorts). Inset: Observed richness, or number of unique clonotypes, deviates from the line $y=x$, saturating at higher sample sizes. D) HIV clone rank abundance at absolute abundance observed and colored by cohort. Inset: Observed richness roughly follows the line $y=x$, increasing with increasing sample sizes. E) Ecological diversity metrics (Hill numbers D^x/N calculated from data resampled to the same size and normalized by dividing by the common sample size $N=10^4$) calculated for mCD4+ TCR β repertoires by cohort. F) The same ecological metrics stratified for all cohorts by HIV and CMV serostatus. Horizontal lines in **E** and **F** indicate comparisons tested with 1 sided Mann-Whitney test, ** above solid lines indicates $p<0.005$ and dashed lines indicate $p>0.05$.

Cytomegalovirus (CMV) infection but not suppressed HIV infection is associated with the clonality of mCD4+ T cell compartment

While it is well known that CMV infection in humans is associated with increased T cell clonality⁶³⁻⁶⁶, it is not clear whether HIV infection is similarly associated with perturbed clonality in the memory CD4+ T cell compartment. When data was grouped by HIV and CMV serostatus, we observed that, regardless of the presence of suppressed HIV infection, cytomegalovirus (CMV) serostatus was highly associated with decreased diversity metrics for mCD4s. No significant differences were observed for richness, but mCD4 repertoires from HIV-/CMV+ individuals were less even ($p=0.0007$) and displayed more clone dominance ($p=2.4e-7$) than those from HIV-/CMV- individuals (**Figure 1F**). Importantly, among people with positive CMV serology, there were no significant differences in mCD4 cell diversity metrics between PWH on ART and HIV-seronegative individuals. These results suggest that CMV infection in humans is associated with significantly increased mCD4 clonality while suppressed HIV infection in CMV+ individuals is not.

HIV-infected cells are more clonal than general memory CD4+ T cells even after controlling for lack of new HIV diversity during suppressive ART

Two steps were required to fairly compare the clonality of HIV-infected rCD4s, harboring either intact or defective proviruses, and rmCD4 cells from the same longitudinal peripheral blood samples. First, unlike the memory CD4+ TCR β pool which continually gains diversity as naïve T cells encounter antigen and enter the memory pool, ART reliably blocks viral replication^{34,67-69}, meaning few if any novel clonotypes enter the HIV-infected cell pool in PWH with suppressed viral loads on ART. HIV-infected CD4s have been observed to be more clonal than total memory CD4+ T cells from the same sample without controlling for this factor^{40,71}. Therefore, we compared the clonality of HIV proviruses to “persistent TCR β ” clonotypes, which are TCR β s found at t1 or t2 that were also confirmed present at t0, a time point close to the initiation of ART (schematic in **Figure 2A**, **Figure 2B**).

Second, we computationally downsample persistent TCR β clonotype distributions to the same sample size as HIV sequence sample sizes at the matching participant time point⁴⁶. Rank abundance curves immediately suggest that HIV-infected cells are more clonal than persistent rmCD4s (**Figure 2C**).

Further, our ecological diversity metrics (Hill numbers, normalized by sample size) indicate that HIV proviruses are generally still more clonal than downsampled persistent TCR β ($p<0.05$ for normalized richness and normalized exponential Shannon (D^0/N and D^1/N) metrics, **Figure 2D**, **Figure S4**). The normalized inverse Simpson metric (D^2/N) was not significantly different, suggesting that despite differences in overall clonality, measures of clone dominance in rmCD4 and HIV proviruses are similar. Together, these analyses imply that even within the same individuals and even corrected for novel clonotypes (diversity) entering the rmCD4+ T cell compartment after ART initiation, HIV-infected CD4s in PWH on ART are generally more clonal and less diverse than the underlying population of persisting rmCD4+ T cells, though both contain dominant, highly expanded clones. The dominant, highly expanded clones may be responsive to common antigens across cohorts, such as CMV. To investigate the clonality of dominant clones further, we use a novel mathematical approach.

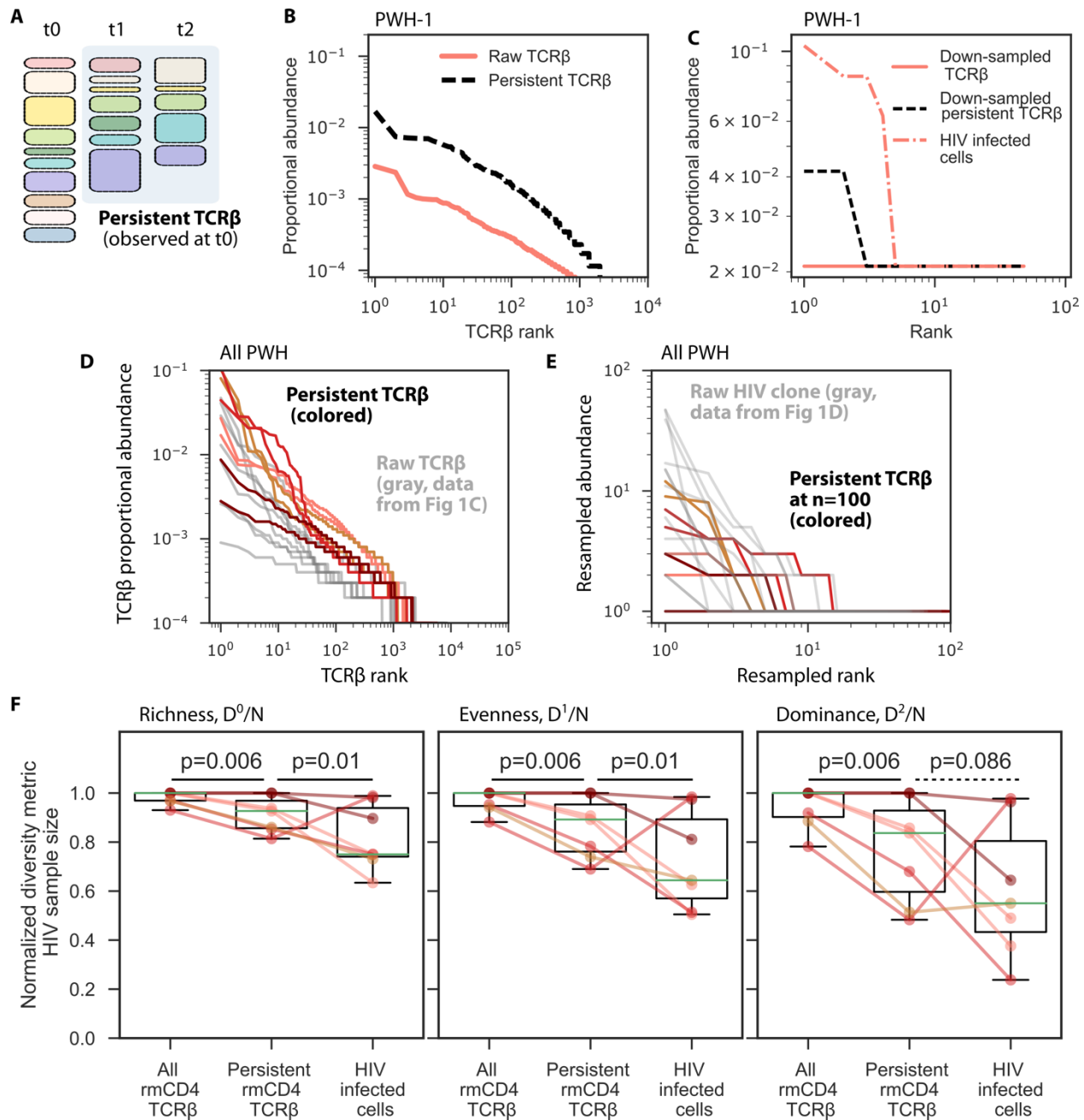


Figure 2. HIV proviruses are more clonal than persistent resting memory CD4+ T cells from the same peripheral samples. A) To fairly compare the clonality of HIV proviruses in PWH on ART to rmCD4 cells, which naturally gain new clonotypes over time as naïve T cells encounter antigen and enter the memory compartment, we define “persistent TCR” as TCRβs present at t1 and t2 that were also present in a sample close to the initiation of ART, t0. B) Proportional abundance of persistent TCRβ vs all raw rmCD4 clones for PWH-1(t1). C) Proportional abundances of downsampled TCRβ vs persistent TCRβ vs HIV proviruses for PWH-1(t1). Downsample size was set to the sample size of HIV proviruses from that participant time point ($n=n_{HIV}$). D) All PWH persistent TCRβ (colored) vs raw TCRβ (gray) proportional rank abundances. E) All PWH persistent TCRβ downsampled to HIV sample size vs raw HIV-infected cell rank abundances. Rank abundances are statistically different, see **Figure S4**. F) Ecological diversity metrics for downsampled total

rmCD4 cells, downsampled persistent TCR β , and HIV proviruses from the seven (of eight) t1 and t2 PWH timepoints with $n_{HIV}>10$. All downsample sizes were set to the sample size of HIV proviruses from that participant time point and metrics were normalized to this value. P-values indicate paired differences using a one-sided Mann-Whitney test.

Higher resolution rank abundance modeling reveals two potentially mechanistically different categories of resting memory CD4+ TCR β clones

The incongruity between HIV-infected rCD4s and general rmCD4+ TCR β clonality is interesting, given that mCD4+ T cells are the putative carrier cells for HIV proviruses, and that HIV clonality is generally understood to be driven by CD4+ T cell proliferation. Therefore, we refined our comparison using mathematical modeling of clonality distributions.

We previously modeled HIV sequence distributions using a power law model³⁴ where abundance (a) follows rank (r) with $a(r) \sim r^{-\alpha}$. This distribution suggests clonal abundances arise from stochastic dynamics where clonotypes have heterogeneous proliferation and death rates⁷⁰. Meanwhile, the clonal structure of CD4+ T cells has also been modeled using a power-law distribution^{71,72}. Therefore, we first tested our prior power law model on all mCD4+ TCR β data (N=64 time points from 48 individuals, **Table S1**). Upon close inspection, we observed that rank abundances were not always precisely log-log linear. Instead, in some samples, there appeared to be two slopes, with a steeper slope for higher abundance clonotypes (examples in **Figure 3A**). To quantitatively test this hypothesis, we fit a double power law, or two-slope, model: $a(r) \sim r^{-\alpha_1} + \phi r^{-\alpha_2}$. Across all participants, the double power law model had lower root-mean-square (RMS) error, $p=1e-5$ (**Figure 3B**) and was a better fit to many distributions for both PWH on ART and HIV-seronegative individuals.

In addition, the two slopes in this model identify different regions of the rank abundance distribution. We calculated the “change point” or the rank distinguishing at which the slope changes (**Methods**). Among individuals for whom the two-slope model was preferred (2-fold improvement in fit error vs 1 slope model), the first slope pertained generally to the top 2-80 clonotypes, with median change point $r=9$ (**Figure 3C**). In other words, on average, the abundance of the top nine most highly expanded clones in a rmCD4 distribution were best modeled with a steeper power law slope than the rest of the clones. We hypothesized these larger clones might be driven by different dynamical mechanisms, such as recent antigenic response or inflationary memory, than the tens of thousands of smaller TCR β clonotypes.

CMV seropositivity associates with a two slope rmCD4+ T cell rank-abundance model

Some support for the hypothesis that clones in the larger of the two power law slope rank abundances are driven by antigenic expansions is generated by stratifying the individuals by HIV and CMV seropositivity. rmCD4+ TCR β rank abundance distributions from HIV and CMV seropositive individuals were substantially more likely to require a two slope model than seronegative participants ($p=0.003$ and $p=2e-8$ for HIV and CMV, respectively, **Figure 3D**). The ratio of the error of the models showed the two slope model had 2-20x better fit for seropositive individuals whereas errors between one slope and two slope models were roughly equivalent for CMV seronegative, HIV seronegative participants. This finding implies that a steeper section of the rmCD4 clonotype rank abundance curve is sometimes found in HIV seropositive individuals and even more commonly found in CMV seropositive individuals, suggesting that the most highly expanded rmCD4 clonotypes could be responding to these viral antigens⁷³.

We explored clinical variables including participant sex, age, months on ART, current CD4+ T cell count, nadir CD4+ T cell count, and estimated months of HIV infection before initiation of ART for associations with power law slope model parameters (**Figure S5**). After Bonferroni correction, significant associations were observed between male sex, positive CMV status, and low nadir CD4 count for the larger clone slope (α_1), and between positive CMV status and low nadir CD4 count for the change point of the large and small clone slopes.

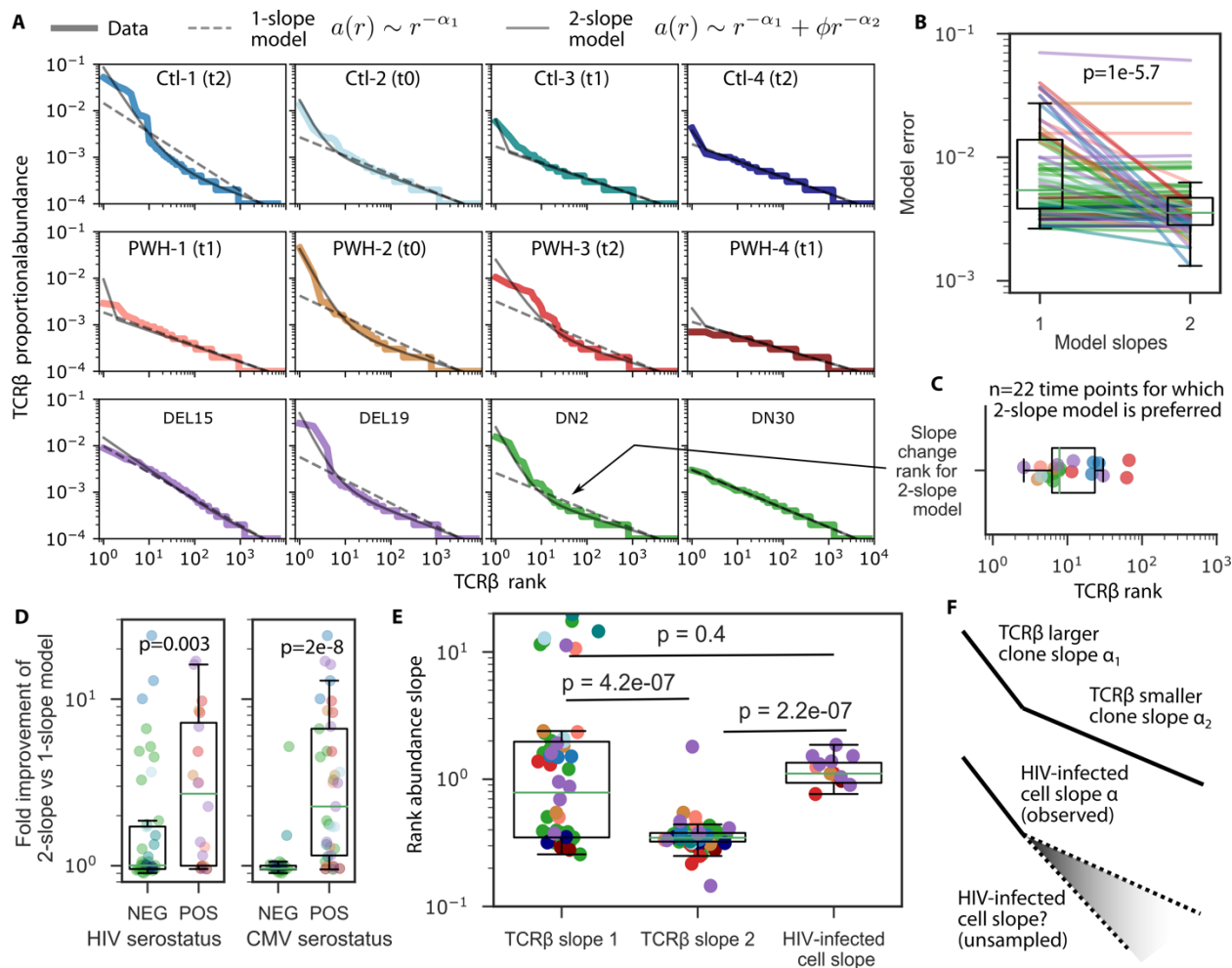


Figure 3. HIV infected cell clonality resembles that of the most highly expanded resting memory CD4+ T cell clones. A) Examples of fitting one slope and two slope power law models to rank abundances from PWH and HIV-negative individuals. B) Comparison of one and two slope model root-mean-square errors. C) The slope change rank, or the rank distinguishing the first and second slopes of the rank abundance curve. D) Fold improvement of model fit for HIV and CMV seropositive and seronegative individuals shows two-slope models are more likely (2-20 fold) to match data in seropositive individuals. E) Comparison of large and small TCR β clone slopes to HIV-infected cell slopes. F) Cartoon illustration of the distribution shapes: HIV proviral clones and large TCR β clones have similar slopes, but since sample size of HIV sequences is lower, we are not able to predict potentially smaller clone slopes in HIV clones. In all panels, p -values indicate 1 sided Mann-Whitney U test.

rmCD4+ T cell large clone slope resembles HIV-infected CD4+ T cell clone slope

The first, or larger, clone slope from rmCD4s was significantly steeper than the smaller clone slope ($p=1.3e-8$, **Figure 3E**). More intriguingly, it was not significantly different from the HIV sequence rank abundance slope ($p=0.2$, **Figure 3E**), suggesting that the clonality of the most highly-expanded, ‘observed’ HIV proviruses may be driven by the same biological factors that drive the clonality of the most highly expanded rmCD4 clonotypes, e.g. antigen-specificity, effector ratio, lack of exhaustion^{21,23,24,74}. The magnitude of the first slope indicates there are still large relative differences in clonal abundances; for example, the larger TCR β clone slope of $\alpha_1 \sim 0.8$ implies the 10th largest clone is less than 1/6 the size of

the largest clone. In contrast, the smaller TCR β clone slope of $\alpha_2 \sim 0.4$ (**Figure 3**) means the smaller clones are more similar in size; the 100th clone may not be much smaller than the 10th, and the 1000th is likely not much more than 1/10 the size of the 10th. Put another way, for clones modeled by the first slope, every 10-fold reduction in rank is associated with a 6-fold reduction in clone size; for clones modeled by the second slope, every 10-fold reduction in rank is associated with a 2.5-fold reduction in clone size.

Although we can compare the clonality distributions of the largest rmCD4 clones and the largest HIV-infected clones, we cannot comment on similarities of smaller, less-expanded, HIV provirus clones to project whether they are best modeled by two slope power law models like rmCD4+ TCR β clones, simply because the sampling depth of HIV-infected cells is not sufficient due to technical limitations of the assay (**Figure 3F**). Our observations imply that the mechanisms underlying the clonality of the largest observed HIV clones (e.g., antigen-driven stimulation) could be preferentially seen to maintain reservoirs by the HIV cure field, despite the possibility that other mechanisms (e.g. homeostatic proliferation) may maintain a larger in number, but less clonally dominant, population of reservoir cells.

Finally, we wanted to compare the clonality of the replication competent reservoir to all HIV-infected cells – which contain mostly defective proviruses. Indeed, we found DEL cohort participants had statistically higher HIV reservoir slopes compared to all proviruses from longitudinal PWH participants (**Figure S6**). This could be due to differences in cohort makeup, adherence patterns, time on ART, that DEL sequence clones are determined by HIV *env* sequencing vs near-full length sequencing, or because the replication competent reservoir truly is more clonal than all HIV-infected cells, but we could not distinguish these several factors.

CMV-responding and HIV-responding rmCD4+ T cells have similar clonality to HIV infected cells

To address whether the most highly expanded rmCD4 clonotypes were specifically responding to two relevant and potentially common antigens in these participants, we identified CMV- and HIV-responding TCR β sequences using the viral functional expansion of specific T cells (ViraFEST) assay from PWH-1 and -4, who were available to donate samples specifically for this assay at an additional time point t3 (16 and 19 years after ART initiation, **Figure 4A, Methods, Table S1**)⁷⁵. The ViraFEST assay identifies TCR β sequences of CD4+ T cells that proliferate *ex vivo* in response to antigen stimulation. We chose ViraFEST over other methods that identify cells expressing activation markers in response to antigen⁷⁶ since we were interested in a clone's ability to proliferate. We identified 335 CMV-specific and 69 HIV-specific TCR β sequences from PWH-1 and 323 CMV-specific and 25 HIV-specific TCR β sequences from PWH-4, which distributed throughout rank abundance distributions at all time points (**Figure 4B**). Across all participant time points, two CMV-specific clonotypes were found in the top 10 most highly expanded rmCD4 clonotypes and 24 CMV-specific clonotypes and 3 HIV-specific clonotypes were in the top 100 rmCD4 clonotypes.

Across participant time points, CMV-specific rmCD4 clonotypes were significantly more abundant than the same number of randomly chosen rmCD4 clonotypes ($p=0.03$, **Figure 4C**). HIV-specific clonotypes were slightly more abundant than the same number of randomly chosen clonotypes but not significantly so. PWH-1 had rmCD4 distributions that required a two slope model, and the slope change ranks for this participant's time points t0, t1, and t2 were $\phi=18, 3,$ and 10. One CMV-specific clonotype was observed in the first slope (at t0, **Figure 4B**) and proliferated *ex vivo* in response to antigen at t3, indicating that this CMV-specific rmCD4 clone has remained at high frequency in the peripheral circulation and retained its ability to proliferate in response to antigen over at least a 15 year timespan, in agreement with other work reporting on memory-inflated CMV-specific mCD4 clones over shorter time intervals in PWH who were not on ART⁷⁷. The antigen specificity of the remaining clones in the first slope at each timepoint of PWH-1, , remains unknown.

Using normalized diversity metrics, we compared the clonality of antigen specific rmCD4 cells to that of HIV-infected cells. Metrics were calculated by downsampling and normalizing to the lower sample

size of the two types (antigen specific vs HIV-infected cells), and comparisons were made pairwise for each person-timepoint. CMV-specific rmCD4s were no more or less clonal than HIV proviruses (**Figure 4D**). HIV-specific rmCD4s were somewhat more clonal ($p \sim 0.05$) than HIV proviruses, but differences could be driven by the very low numbers of HIV-specific clones observed in 3 of 4 participant time points (N=198, 28, 14, and 29). Therefore, we examined power law exponents from rank-abundance distributions (**Figure 4E**) and observed they were not significantly different between antigen-specific cells and HIV proviruses ($p > 0.18$, **Figure 4F**).

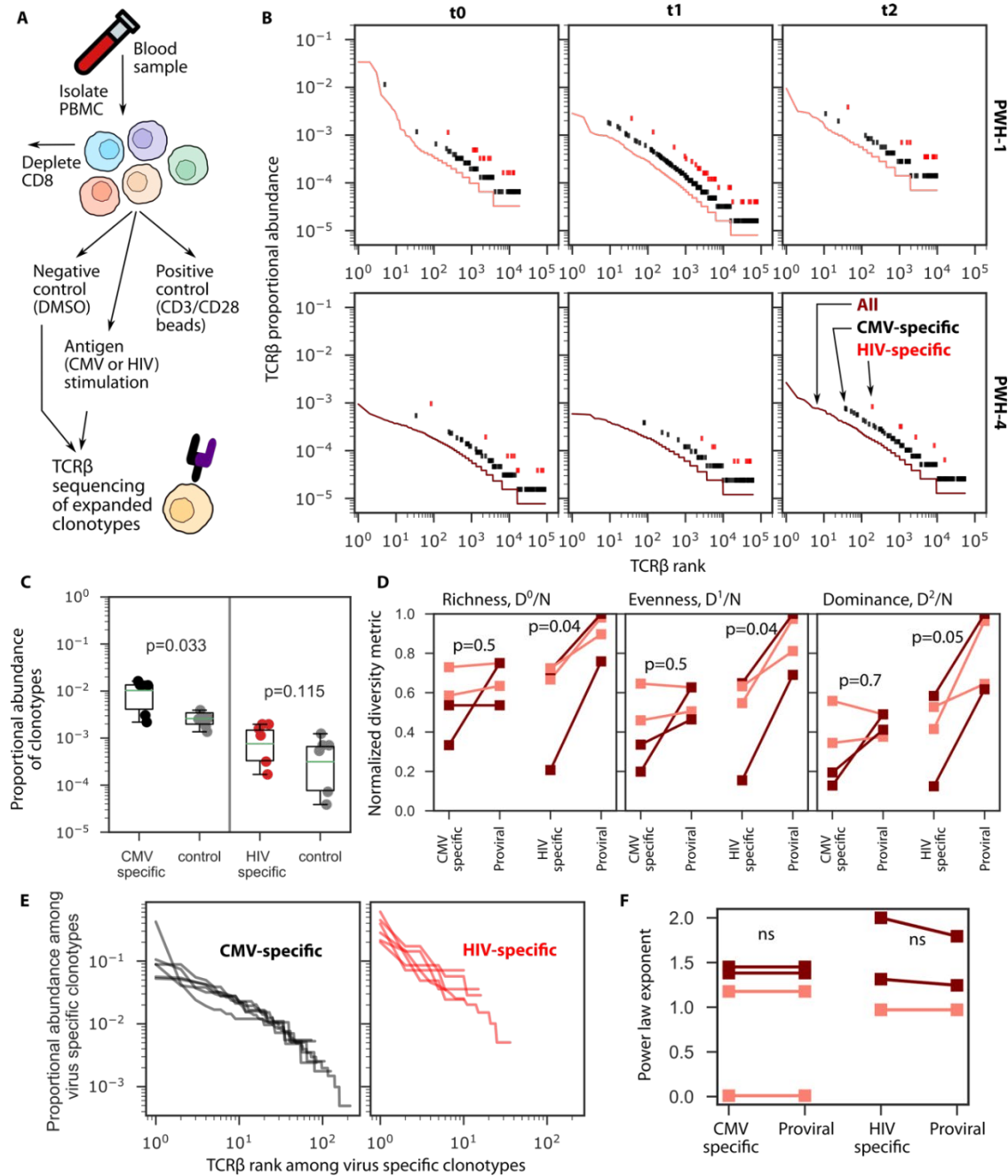


Figure 4. Clonality of CMV- and HIV-specific rmCD4+ T cells. A) Schematic of the ViraFEST assay, which identifies TCR β sequences from CD4+ T cells that proliferate in response to specific antigens. B) rmCD4+ TCR β rank abundance distributions for PWH-1 and PWH-4 (curves) with markers indicating the rank of

*clonotypes that proliferated in response to CMV (black) or HIV (red) antigen by ViraFEST at t3. C) Summed proportional abundances of antigen-specific clonotypes at each participant time point (red and black dots) vs proportional abundance of the same number of randomly selected rmCD4+ T cell clonotypes from the same person-time point. D) Normalized ecological diversity metrics comparing clonality of CMV-specific, HIV-specific, and HIV proviruses compared pairwise where each value in a pair is calculated by downsampling data to the lower sample size at each participant time point and normalizing metrics to that sample size. E) Rank abundance curves of only the CMV- and HIV-specific clonotypes, one curve per person-timepoint. F) Power law exponents of a one slope power law model of rank abundance curves of CMV-specific, HIV-specific, and HIV proviruses at each person-time point. In **D** and **F**, colors indicate PWH-1 (darker) and PWH-4 (lighter), p-values indicate paired t-test; ns, not significant $p>0.05$.*

The clonality of persistent and antigen specific rmCD4+ clones remains stable over time in people with and without HIV, whereas the clonality of HIV-infected CD4s increases over time within individuals

The characteristic shape of rmCD4 clonotype rank abundance curves for each longitudinal PWH on ART and HIV-negative participant did not have significant within-person differences across three samples spanning 10 years, although there are differences between people that are not attributable to HIV serostatus (**Figure 5A**). We previously showed that HIV-infected cells become concentrated in highly expanded clones over ~10 years on ART in these individuals¹³ (**Figure S7**). By contrast, here we observe no similar increase in clonality in persistent rmTCR β (**Figure 5B**) or all rmTCR β (**Figure S8**) over ~10 years on ART in the same individuals over the same time span. Additionally, although sample size was small, we also did not observe trends supporting changes in longitudinal clonality for CMV- and HIV-specific rmCD4+ T cells (**Figure S9**). These results agree with other demonstrations of the stability of clonally dominant CD4+ T cells over time spans of ~ 2 years⁷⁸.

rmCD4+ clones wax and wane with time

Although rank abundance shapes are relatively constant, these distributions do not indicate whether the individual TCR β clonotypes making up the distributions change in size (or in ranks). Therefore, we calculated fold-changes in clonotype sizes over time in PWH and Ctl participants (example in **Figure 5C**, all participant data in **Figure S10**). We found that most clonotypes which were initially small (observed singletons) did not change substantially at follow up time points, though occasionally they expanded. It was almost never observed that an initially large clone (observed abundance >30 in sample size $1e4$) expanded further by the second time point. Overall, across participants a strong negative correlation was observed between initial size and fold-change in size with time (median Spearman $\rho=-0.6$, **Figure 5D**). That rmCD4+ T cell clones fluctuate in size or “wax and wane,” as documented for HIV proviruses⁵⁷, despite little change in clonal structure indicates an impressive dynamic equilibrium. It appears that the identity of the clonotype inhabiting each rank in the distribution often changes even while the relative abundances between the ranks remains relatively stable. For instance, new clones may occupy the first and second ranks at a different time point, but the relative sizes of the two new clones remains very similar to the top-ranked clones from a prior or future time point.

Expansions and contractions of rmCD4+ T cells are roughly symmetric and exponentially distributed.

To further explore clonal dynamics, we quantified the frequency of expansions and contractions of rmCD4 clones, finding an approximately exponential distribution such that larger magnitude changes were less likely to occur (**Figure 5E**). rmCD4s of PWH on ART were more likely to undergo large clonal expansions when compared to HIV-seronegative participants ($p=0.01$) but contractions were similar across cohorts.

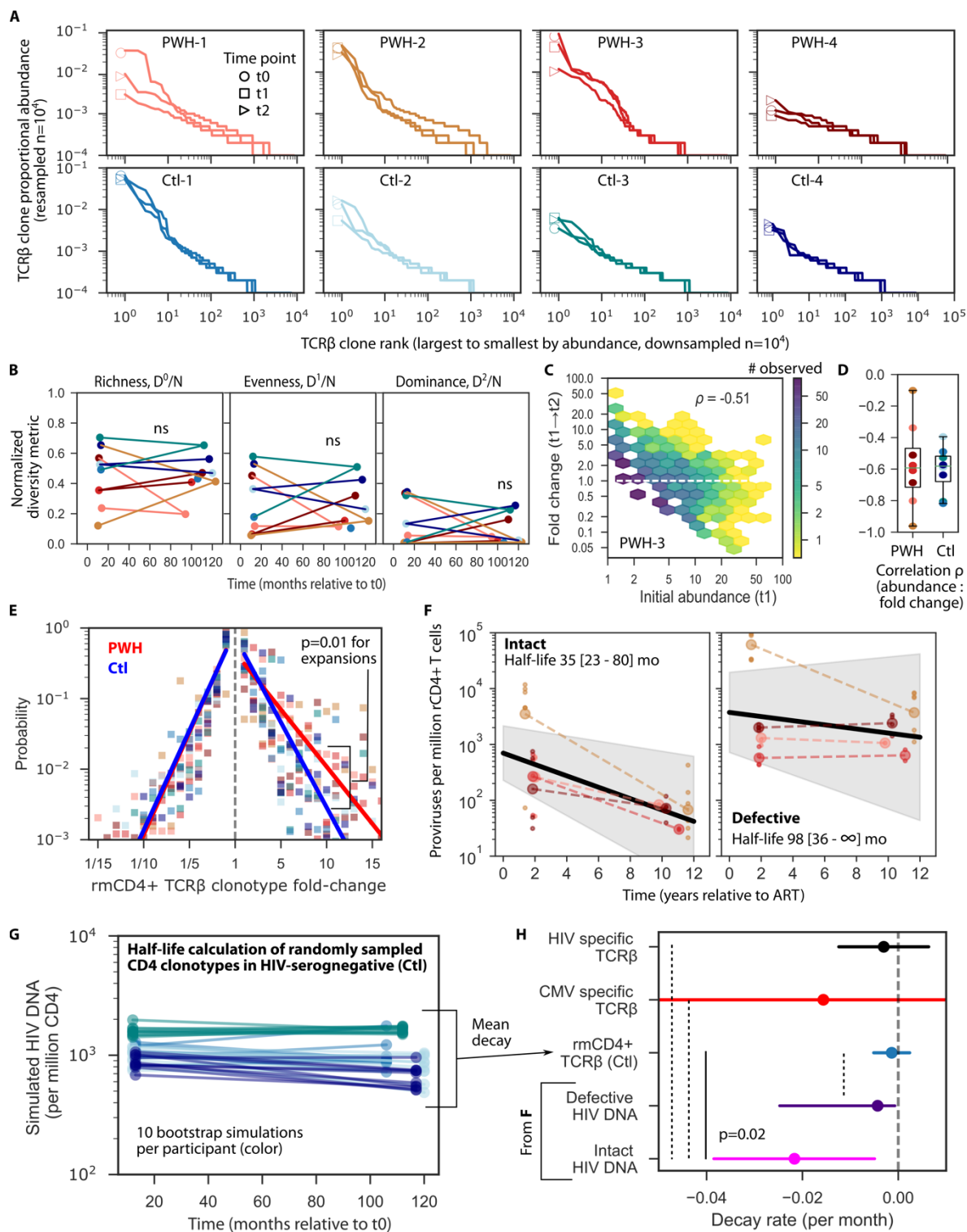


Figure 5. *rmCD4+ TCR β cell clones fluctuate in size and decay like defective but not intact HIV proviruses.*
 A) TCR β rank abundance curves from PWH and HIV-seronegative individuals sampled at 3 time points (symbol at rank=1 specifies sample time point). All data are resampled to have 10^4 total sequences. B) Longitudinal normalized ecological metrics from persistent TCR β . No decay was observed by Pearson correlation (ns). C) Example comparison of initial abundance at t1 vs. fold change to t2 of all rmCD4+ T cell clones that were observed in a single participant. Color bar shows how frequently a certain initial size

occurred (density of observations in each hex bin). The most common clonotype is initially small and remains small (purple). Initially large clones infrequently expand, and higher observed initial abundance predicts clonal contraction (Spearman $\rho=-0.5$). D) Spearman correlation coefficient (as in C) between initial abundance and fold change for all PWH and Ctl participants. E) Across all rmCD4+ TCR β that were observed at two time points, we calculated the probability of each TCR β to expand or contract in abundance at a certain level. Large expansions and contractions are less and less likely to occur, roughly following an exponential distribution with fold-change. Dots indicate values from each participant and red and blue line summarizes the slope of the probability distribution for expansions and contractions in PWH and Ctl, respectively. P value (one sided Mann-Whitney) indicates significant difference between the slope of the distribution of expansions comparing PWH and Ctl cohorts. F) Observed levels of Intact and defective HIV DNA for all 4 PWH and modeled decay (population mixed effects approach to estimate decay rates) over the first 10 years of ART. Mean population decay rates are shown as black lines and 95% CI are shown as shaded intervals. Large circles, mean; small circles, replicates. G) Simulated half-life of randomly sampled of TCR β sequences from HIV-seronegative individuals (Ctl) to mimic IPDA decay experiment (**Methods**). 10 bootstrap simulations were performed for each individual (colors). H) Observed intact and defective HIV DNA decay rates compared to the mean simulated decay rates from rmCD4+ T cells in G from all participants. Additionally, the average and 95% CI for HIV- and CMV- specific clonotypes between t1 and t2 is also presented. Dots are mean, lines are 95% CI for the mean. P-values indicate significant differences using Z-test. Dashed lines indicate non-significant comparisons.

Defective proviruses decay at a similar rate as general rmCD4+ T cells, both of which decay significantly slower than intact HIV proviruses

From our cohort of 4 PWH, we used the intact proviral DNA assay (IPDA) to measure intact and defective HIV proviral decay rates in the longitudinal PWH on ART and found them to be in line with results from other publications^{5,10,36,79,80}: the population mean intact provirus half-life was 35 months (95% CI 23-80 months) and the population mean defective provirus half-life was 98 months (95% CI 36-infinite)¹³ (**Figure 5F**). The intact provirus decay rate is faster than the defective provirus decay rate, but it remained unclear whether either of these rates simply reflected the half-life of uninfected rmCD4s. Thus, we sought to recapitulate the decay rate experiment in our matched longitudinal cohort of HIV-seronegative people. We obtained TCR β repertoires from total rCD4+ T cells at t1 and t2 in HIV-seronegative participants. Importantly, the total rCD4 repertoires included both memory and naïve cells, just as we measured HIV infected cells among total rCD4 cells in our IPDA experiments. We randomly sampled a subgroup of resting *memory* CD4+ T cells identified from rmCD4+ TCR β repertoires (having the same frequency as HIV-infected cells measured by IPDA, ~1000 per million CD4 at t1, **Figure 5F**) and then determined the change in frequency between t1 and t2 of that subgroup as a proportion of total rCD4 cells using the sum of their TCR β abundances. We repeated this random sampling and frequency tracking for each HIV-seronegative Ctl participant 10x per participant (**Figure 5G**).

We found that the decay rate of uninfected rmCD4s was on average zero. The corresponding half-life was not statistically distinguishable from that of defective proviruses but was statistically different from that of intact proviruses, which decay over time (**Figure 5H**). This suggests that rCD4s harboring defective proviruses may generally experience similar proliferative and destructive forces in humans as HIV-uninfected rmCD4s, remaining at stable levels in circulation over ten years with little appreciable selective advantage or disadvantage due to HIV-specific or other factors. In these 4 participants, HIV-infected cells harboring intact proviruses appeared to decay faster than uninfected rmCD4 cell in the first decade of ART ($p=0.02$) (**Figure 5H**). Though the means were slightly different, defective provirus decay was not significantly different from that of rmCD4 clonotypes.

Finally, we calculated the decay rate of all CMV- and HIV-specific clonotypes between t1 and t2 (**Figure 5H**). These rates were not different from zero, indicating that typical HIV- and CMV-responding

rmCD4s in the two individuals assayed with ViraFEST remained roughly stable over a decade of suppressive ART (**Figure 5E**). Because of small numbers and high heterogeneity, these rates have wide 95% confidence intervals.

Simultaneous mathematical modeling of rmCD4+ TCR β and HIV-infected cell clones

Next, we sought to design a mathematical model that was sufficient to comprehensively describe the many observations relating to both HIV-infected cell and rmCD4+ TCR β clone dynamics presented here and elsewhere. The model is a variant merging our previous work on HIV phylodynamics⁸¹ and reservoir creation⁷⁰ (**Figure 6A, Supplementary Methods**).

Briefly, rmCD4+ TCR β clones T_i are distributed into $T(0) = 10^6$ rmCD4+ T cells such that $\sum_i T_i = T(0)$. We set the fraction of cells carrying HIV DNA ($H(0) = 3000$ per million CD4s) and intact ($H^*(0) = 800$ per million CD4s) proviruses using population means from our data (**Figure 5F**). The size of each rmCD4+ T cell clonotype is initialized by drawing from a distribution. Then, TCR β clonotypes are given proliferation (b_i) and death rates (d_i), also drawn from a distribution. Thus, the size of each TCR β clone on average follows the differential equation

$$\frac{dT_i}{dt} = (b_i - d_i)T_i \quad \text{Eq 1}$$

Stability is ensured on average by setting $b_i = d_i$ for each clone. We simulate the model as a discretized stochastic system (**Methods**) such that clones expand and contract sporadically, those with higher birth and death rates do so more rapidly, and any clone can go extinct and be removed if its abundance reaches zero.

Sequential inclusion of model features to recapitulate experimental data

We sought a model that could output six key data features/observations simultaneously: 1) stability of rmCD4+ T cell population sizes over 10 years (**Supplementary Figure 1**), 2) decline of intact and slight or no decline of defective proviruses, matching prior published observations of these half-lives⁸⁻¹⁴ (**Figure 5F**), 3) fluctuations of clone sizes with waxing and waning of both rmCD4 cell and HIV-infected cell abundances as we (**Figure 5C&D**) and others^{57,82} have observed, 4) a roughly exponential and symmetric distribution of rmCD4+ TCR β expansions and contractions (**Figure 5E**), 5) a power law rank abundance (and associated ecological metrics) of rmCD4+ TCR β clonotypes that is stable over time (**Figure 5A&B**), and 6) a highly downsampled power law rank abundance of HIV clonotypes that becomes more clonal over time and matches metrics from participant data.

In **Supplementary Data 1**, we provide simulation output from all models that were tested (annotated as **Sim1** through **Sim6**), the sequential complexity that was added to each model, and any experimental observations each model could not recapitulate. For instance, we first assumed initial conditions and proliferation/death rates were drawn from uniform distributions. This model did not achieve agreement with TCR β and HIV proviral clonality (**Sim1**). Instead, we found that initializing clonotypes with a power-law distribution⁸³ $T_x(t=0) \sim x^{-z}$ where $z=0.8$ could achieve fit to observed TCR β clonality and rank abundances (**Sim1**). A key requirement for the model to achieve the simultaneously stable rmCD4 levels and rank abundance while also admitting waxing and waning was that individual clonotypes could not have the same birth and death rate for the entire simulation (10 years). We found redrawing (or “shuffling”) these rates every $s=100$ days (**Sim4**) was sufficient to achieve qualitatively reasonable behaviors, but the exact value s could not be precisely identified from the present data. In addition, an exponential distribution for proliferation and death rates $b_x = d_x \sim \alpha_T \exp(-x)$

parameterized by the average proliferation rate of central memory CD4+ T cells^{24,84}, or $\alpha_T=0.01$ per day, was preferred over a uniform distribution with the same mean to quantitatively match observed ecology and expansions and contractions of clones over time (**Sim6**).

Simulations of the optimal model show agreement with HIV and CD4+ TCR β clonality and clonal dynamics

The optimal model (**Sim6**) matches several types of observational data (**Figure 6B&C**). First, numbers of rmCD4s fall in boundaries (shaded areas, data from **Figure S1**) of general CD4+ T cell count trends from PWH on ART (**Figure 6B**). Second, rmCD4 clonotype rank abundances sampled at $n=1e4$ from the model are power-law distributed, stable over time, and fall in boundaries of participant data (shaded areas, data from **Figure 1C**). Third, intact and defective HIV DNA levels decay with half-lives matching population data (shaded areas, data from **Figure 5F**). HIV rank abundances sampled at $n=100$ at 3 time points appear like undersampled power law distributions, fall within boundaries of PWH data (shaded areas, data from **Figure 1D**) and become more clonal over time. Third, using downsampled data and calculating ecological metrics as in our analysis of experimental data, model output simultaneously matches clonality of both TCR β and HIV proviruses (shaded areas, data from **Figures 1&2**) (**Figure 6C**). Fourth, the model matches observed TCR β clone dynamics, outputting both a correlation between initial clone size and fold-change (Spearman $\rho=-0.5$), in excellent agreement with experimental observations (data from **Figure 5C**) (**Figure 6D**, top), and a symmetric distribution of expansion/contraction sizes that agrees with distributions from experimental data (shaded areas, data from **Figure 5E**) (**Figure 6D**, bottom).

Modeling reveals mild negative selection on intact and some defective proviruses

To recapitulate intact and defective proviral decay rates while maintaining all rmCD4 clonality dynamics and ecology, we required cells carrying intact (and to a lesser degree, defective) proviruses to be modulated with an additional death rate that hypothetically represents some process of viral reactivation and/or immune recognition followed by immune clearance. Each clonotype was assumed to have an additional death rate $\xi^*=6.5e-4$ (intact) and $\xi=2.3e-4$ (defective) per day depending on whether the provirus is intact vs defective^{61,62}.

However, the absolute magnitude of these additional rates in comparison to the natural death rates of CD4+ T cells is meaningful. The average proliferation and death rates in the model ($\langle b_i \rangle$ and $\langle d_i \rangle$) were set to those from central memory T cells, which turn over ~ 3.7 times per year, or 0.01 per day²⁴. This means that in our simulation, the fraction of proviral clearance that is due to HIV-specific selection is approximately $\xi^*/(\xi^* + \langle d_i \rangle)=6\%$ for intact proviruses and 2% for defective proviruses, where $\langle d_i \rangle=0.01$ per day, the average death rate in the model. Natural clearance is balanced by birth for CD4+ T cells, but over years on ART, the additional death rates on HIV-infected cells amount to a slowly decaying pool of proviruses.

One final observation on the optimal model structure is that increases in HIV clonality without commensal changes in TCR β clonality did not require individual clones that were permanently skewed toward clonal expansion -- as they might be if HIV integration site or stable pro-survival signals drove proliferation^{27,46}.

Slight survival advantages for rare intact and defective proviruses could be incorporated into the model without changing observed dynamics over 10 years of ART

It has been observed that certain infected cells may have some survival advantages compared to uninfected cells based on integration site locations or other mechanisms^{44,45,47-50,85}. As a sensitivity

analysis, we therefore added a positive “survival advantage” rate $\zeta_i=1e-5$ per day for a small fraction of HIV-infected clones $f_\zeta=1/100$. This rate was not dependent on whether the HIV provirus was intact or defective. The inclusion of this mechanism at this magnitude did not substantially change our model output comparison with experimental data. However, setting $f_\zeta=1/10$ or $\zeta_i=1e-4$ did disturb reservoir dynamics and resulted in overly clonal/stable reservoirs.

In conclusion, we iteratively generated a model (**Figure 6A**) that matches all desired qualitative and quantitative observations. The form of this ultimate model illustrates the forces on a single HIV infected clone as a balance of proliferation and death rates of the CD4+ T cell clonotype into which it is integrated – governed by Equation 1, less some additional death due to immune selection (or other mechanism) that acts differently on intact H^* vs defective H^0 HIV proviruses, and potentially plus some survival advantage related to integration site or other pro-survival mechanisms:

$$dH^{(*)}/dT = (b_i - d_i - \xi^{(*)} + \zeta)H^{(*)} \quad \text{Eq 2}$$

Model projections of individual clone dynamics

Using this final model, we could further illuminate the predicted underlying clone dynamics that generate outputs. For instance, as observed previously in PWH on ART⁴¹, the model also predicts that HIV-infected CD4 clonotypes are heterogeneous in the proportion of clonotype cells carrying given proviruses: some clonotypes have very few of their cells infected, whereas other clonotypes have a majority of their cells infected -- driven in the model by the stochastic trajectory, the size of the TCR β clonotype when it was initially infected, and the type of provirus carried (**Figure 6E**)¹⁵. HIV-infected cells carrying intact proviruses tend to decrease in fraction within individual TCR β clones (**Figure 6E**, examples in red), whereas the rare HIV-infected cells carrying proviruses with survival advantages remain stable or grow (**Figure 6E**, examples in purple). The clonal proportions of each type of cell or provirus are relatively stable over time by eye, but HIV-infected cells are slightly more clonally dominated, and the rarer intact/survival advantaged clonotypes are still more clonal (**Figure 6F**). Finally, given the rarity of proviruses with survival advantages required to match the experimental data, we project they do not make up a large proportion of HIV infected cells early on ART. However, after 10 years of simulations, levels of these proviruses begin to approximate intact provirus levels (**Figure 6F**). This projection may reconcile data from PWH on ART >15 years suggesting intact proviral decay may slow or stop entirely (potentially even transitioning to growth)^{10-12,36,80}.

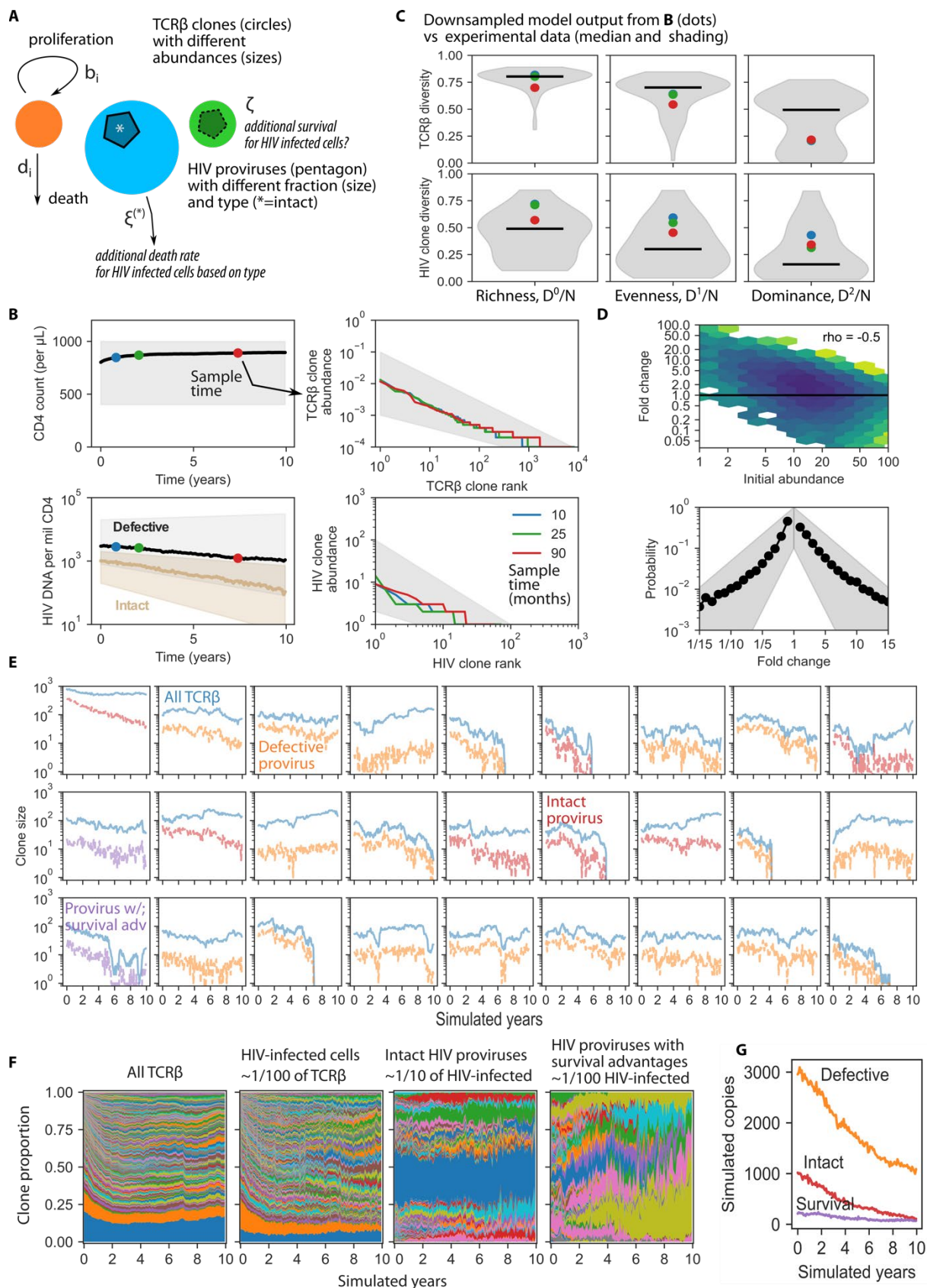


Figure 6. Simultaneous modeling of CD4+ TCR β and HIV-infected clones reveals mild negative selection against HIV proviruses and allows the possibility of HIV-specific positive selection. A) Model schematic to simulate the comprehensive system linking TCR β carriers and HIV proviral passengers during suppressive ART. TCR β clones proliferate and die, some carry HIV (pentagon), some HIV DNA is intact (*) and some is defective (dashed line). The optimal model required an additional death rate (negative selection) on HIV-infected cells, which varies for intact and defective proviruses. B-D) Model output recapitulates ranges from experimental data (shaded regions). B) mCD4 and intact and defective HIV DNA population sizes as well as associated rank abundances at 3 time points (dots in left panel indicate sampling times corresponding to rank abundance curves of same color in right panels). CD4 counts follow general trends of experimental data and are stable while HIV DNA decays and intact decays more rapidly than defective HIV DNA. TCR β rank abundances are static in time (overlying one another) whereas those of HIV become more clonal over time (red high/middle rank clones are slightly larger). C) Ecological diversity metric values calculated for TCR β and HIV provirus abundances. TCR β are sampled at $1e4$ cells and HIV provirus is sampled at 100 cells. Dots are model output from months 10, 25, and 90 as in B compared to shaded regions and medians from experimental data. D) Sampled TCR β initial clone sizes are negatively correlated with fold change with Spearman correlation coefficient matching experimental data in **Figure 5F**. Simulated expansion/contraction distributions are symmetric and fall within boundaries (shading) of participant data in **Figure 5E**. E) 27 examples of TCR β clonotypes (blue) with linked HIV provirus passenger levels (defective = orange, red = intact, purple = survival advantage) illustrate representative fluctuations in clone sizes and illustrate that TCR β clones can have a heterogeneous fraction of their associated cells that carry HIV proviruses. F) Proportional clone sizes for TCR β , total HIV infected cells, intact HIV proviral clones, and survival advantaged HIV clones. G) Total levels on non-log scale of each proviral type over 10 years of simulation time.

Discussion

In this study, we investigated whether the natural physiology of mCD4+ T cells can entirely explain the clonality and decay dynamics of the HIV reservoir in PWH on ART, or whether positive or negative selective forces acting on HIV-infected cells are also required to explain these observations. We focused on three key observations regarding HIV-infected CD4s in PWH on ART: 1) the high degree of reservoir clonality, 2) the increase in HIV-infected cell clonality over time on ART, and 3) the differential decay of CD4s harboring intact vs defective proviruses over time on ART. We found that the natural biology of mCD4+ T cells dovetails with the first observation but cannot explain the second and third observations, implying that selective advantages or disadvantages due to factors specific to HIV-infected cells contribute to the clonal kinetics of the reservoir. Moreover, we report that the shallow provirus sampling depths common in observational studies prevent accurate comparisons of the full clonality distributions between HIV-infected cells and rmCD4+ T cells. Without a full characterization of the ‘tail’ or ‘second slope’ of provirus clonality distributions, we cannot know to what degree forces like antigen stimulation and homeostatic proliferation maintain the reservoir in its totality in PWH on ART. Finally, we construct a mathematical model of HIV-infected cell and reservoir dynamics in PWH on ART through iterative testing and comparison to observational data. In this model, HIV-specific negative selective forces are mild in scope, responsible for approximately 2-6% of cell death events in mCD4s harboring defective and intact HIV proviruses. This model allows for HIV-specific pro-survival selective forces, such as integration into transcriptionally silent genomic regions or upregulation of pro-survival cellular programs, at a very low level ($\leq 1\%$ of HIV-infected cells) at ART initiation. Over the years on ART, the proportion of these survival-

advantaged HIV-infected cells among all HIV-infected cells grows, perhaps explaining why the decay of the reservoir slows over time on ART^{10–12,36,80}.

Regarding the first observation, many groups have observed the high degree of clonality in the reservoir and in HIV-infected cells (intact plus defective) in PWH on ART, reviewed previously^{86,87}. Some groups have demonstrated that HIV-infected cell clonality is greater than that of memory CD4s from the same sample^{39,85} but have not controlled for new diversity entering the mCD4 compartment after ART initiation. This is important because generally no new HIV-infected clonotypes are generated in people on suppressive ART due to absence of viral replication^{34,68,69,88}, whereas over many years, naïve CD4s encounter antigen and proliferate, with some surviving to join the long-lived mCD4 pool^{83,89}. We attempt to control for that in our investigations by examining ‘persisting’ rmCD4s – those memory clonotypes that were present close to the start of ART and persist 2 and 11 years after ART initiation. We demonstrate here that HIV-infected cell clonality in PWH on ART is greater than that of persisting rmCD4s when clonal richness (number of clones per normalized sample size) and evenness are quantified.

However, we observe that measures of clonal dominance are similar between persisting rmCD4s and HIV-infected cells. Additionally, we find that measures of clonal dominance in rmCD4s, whether quantified by normalized Hill number D^2 or a novel two-slope rank-abundance model, strongly associate with CMV serostatus, and we hypothesize that dominant clones in the reservoir and among rmCD4s become dominant due to proliferation after recent or recurrent antigen exposure. In the two longitudinal PWH available for present-day sampling, we identified TCR β sequences of mCD4s that proliferated significantly *ex vivo* after exposure to HIV antigens and after exposure to CMV antigens. We then showed that the clonality of these antigen-responsive clones was similar to that of HIV-infected cells across all measures of clonality - richness, evenness, and clonal dominance. This implies that natural forces acting on mCD4s, in this case antigen responsiveness, can account for the clonality of the observed HIV reservoir.

We document an important caveat to this conclusion. While we demonstrate that our sampling depths of rmCD4+ T cells reach clonal richness saturation, we demonstrate that HIV provirus sequencing depths of up to 100 sequences per time point do not appear close to richness saturation (**Figure 1C**). This means that provirus sequencing depths of greater than 100, and probably greater than 1000 are needed to approximate the richness of HIV-infected clones in a given sample and to assess the clonality of non-dominant HIV-infected clonotypes. However, the labor-intensive nature of provirus sequencing and the rarity of HIV-infected cells in single cell RNA sequencing approaches make it prohibitively time-consuming or expensive to reach HIV provirus sampling depths much beyond 100. One group has directly compared HIV integration site datasets – which do not distinguish intact from defective proviruses - on the order of 1000+ per sample to mCD4 TCR β repertoires and found that the largest HIV-infected clones are more stable than randomly-chosen mCD4+ T cell clones of the same size³⁵. One plausible explanation for this is that the largest HIV-infected clones are biologically more similar to the largest randomly-chosen mCD4+ T cell clones instead of the mid-sized clones with which they were compared. We show here that the largest HIV-infected cells and the largest mCD4+ T cells can be modeled with the same power-law slope, and we think it possible that both types of clones are responding to recent or recurrent antigen. Our work reveals that undersampling of reservoir datasets leads to characterization of the most highly expanded, or clonally dominant, HIV-infected cells, whose dominance is likely driven by exposure to antigen. Thus the HIV cure field may be biased towards results implying that antigen-driven proliferation provides the dominant stimulus for the persistence of the HIV latent reservoir. We show that the clonality and the longitudinal dynamics of non-dominant HIV-infected clones – the ‘tail’ or the second slope of the clonality distribution - have not yet been characterized (**Figure 3F**), although non-dominant clones probably make up the majority of HIV-infected cells in humans^{34,35,38}. Because of this gap in knowledge, it remains

unknown whether homeostatic proliferation, less frequent or intense antigen exposure, or other forces underlie the persistence of the reservoir in non-dominant clones. In order to understand the persistence of the HIV reservoir in its entirety in PWH on ART, we need more deeply sampled reservoir datasets.

Regarding the second observation, we and others have demonstrated that the clonality of HIV-infected cells increases with time on ART^{13,14,27,37,90}. Our previous work modeling this change in clonality suggested that after a decade on ART, over half of an individual's remaining proviruses are members of the top 100-1000 most-highly expanded infected clones¹³. Here we demonstrate that mostly-uninfected persisting rmCD4s from PWH on ART and uninfected rmCD4s from HIV-seronegative people do not increase in clonality with time using any of the four measures of clonality we examine. Additionally, we did not observe increases in clonality over ten years on ART of the antigen-specific clonotypes we examined, although our observations of antigen-specific cells were limited by numbers of people available for current-day sampling and should be repeated with more participants. In addition, proliferation *ex vivo* upon stimulation with high peptide/antigen density may capture only some of the clones responding to a certain pathogen *in vivo*. Taken together, HIV-specific forces seem to be required to explain observations of increasing HIV-infected cell clonality over time.

Regarding the third observation, many studies have now documented a decay in rCD4+ T cells harboring intact HIV proviruses in the first 7-10 years on ART and a relatively stability, or at least an extremely slow rate of decay, in rCD4+ T cells harboring defective HIV proviruses in PWH on ART^{8,10,13,36,91}. We demonstrate here that decay rates of defective proviruses are similar to decay rates of mostly-uninfected rmCD4s in PWH on ART, HIV-responding rmCD4s, and uninfected rmCD4s in HIV-seronegative people. Importantly, we show that clearance of intact proviruses would not be predicted by HIV-independent rmCD4+ T cell forces alone^{59,64}. Overall, this suggests that the presence of intact provirus in mCD4s is associated with enhanced death rates, at least in the first decade on ART. Our observational data still allows the possibility that defectiveness is a continuum and that certain defective clones experience more or less negative selection.

We compared normalized diversity metrics across our carefully matched longitudinal cohorts and found that rmCD4+ T cell diversity metrics did not distinguish participants with suppressed HIV infection from those who were HIV-seronegative and matched on age, sex, race, and CMV serostatus. One difference we observed is that clonal expansions of rmCD4 cells over time was slightly more common in PWH on ART than HIV-seronegative people, but we could not tell whether this was due to CMV, which is associated with clonal expansion over time, or other factors because all but one longitudinal participant had CMV. Beyond HIV, the longitudinal nature of our study admitted some novel properties of rmCD4+ T cell repertoires. First, although repertoires had distinct rank-abundances shapes for each individual, rank-abundances were remarkably stable within participants – including PWH on ART (**Figure 1 & 5**). Although aging is known to modulate T cell diversity^{92,93}, the repertoire clonality among these adults did not substantially change over a decade of observation. Second, we observed that individual clonotypes are dynamic with constant shifts in the dominant clonotypes. Large clones observed at one time point were much more likely to contract than expand at a later time point (**Figure 5C**). With respect to HIV reservoirs, dynamic TCR β expansions and contractions (waxing and waning) serve as a potential mechanistic basis for observed HIV clone waxing and waning⁵⁷. Third, we observed in HIV-seronegative people that, in contrast to intact HIV DNA levels, the frequency of a randomly selected group of rmCD4 clones among total resting CD4+ T cells (memory plus naïve) does not change in abundance over a decade of observation, a powerful example of the maintenance of immune memory in humans.

Our introduction of the double power law model of rmCD4+ T cell clonality illuminated that a model that separated TCRβs into two distinct clonality categories was associated strongly with CMV serostatus, hinting that (as observed previously using other approaches⁷³) large TCRβ clones could be responding to CMV. When we directly tested which rmCD4 clones retained the capacity to proliferate to CMV antigen, we were able to identify only one of the largest clones from years ago as currently CMV-responsive, while many smaller clones were identified as currently CMV-responsive -- similar to other groups' findings⁸⁵. This may be due to assay insensitivity (wherein most large clones respond to CMV but our assay missed them), or because large clones are responding to other antigens (including bacteria⁹⁴), or because the antigen-specific large clones identified at years 2 or 11 after ART initiation are now functionally exhausted at year 16 or 19.

Our mechanistic model grows from prior work on T cell dynamics and distributions^{83,95-99} and proliferative HIV clone dynamics (modeled similarly to our approach¹⁰⁰ and in more classical viral dynamics models^{101,102}) into what we believe is the most rigorously data-validated model to describe the linked system of uninfected and infected CD4+ T cells in PWH on ART⁷⁴. It achieves our goal of matching many observed features of HIV and CD4+ TCRβ clonal and population size dynamics simultaneously. It infers a slight but essential force of negative selection on HIV-infected cells that we estimate accounts for ~2% of defective and ~6% of intact HIV provirus clearance during ART⁴⁸. Additionally, it supports the possibility of rare (<1/100) proviruses that attain survival advantages based on integration site, viral genetic features, or upregulated pro-survival cellular programs, phenomena now observed by many groups^{26,27,37,42-45,47-50,85} that could explain reservoir stability or expansion after decades of ART. Still more or different data would be needed to estimate this population's size and dynamics precisely in our framework. Of course, the model naturally oversimplifies the system. Among many approximations, clone proliferation is stochastic but does follow programmed expansions and contractions¹⁰³, and we also institute an artificial reshuffling of clone proliferation rates. For now, we do not directly connect cellular proliferation to viral reactivation, as it is not yet clear these are consistently linked¹⁰⁴⁻¹⁰⁶.

There are important limitations in this study. We sequenced near-full length (nfl) proviruses, QVOA-outgrowth HIV *env*, and TCRβ, and each method has different biases in approximating clonality, so their comparison is not perfect. Nfl clonality is strong evidence of true clonality¹⁰⁸, especially for defective proviruses, while HIV *env* sequencing is likely biased towards higher clonality despite being the most genetically diverse region of the HIV genome. TCRαβ repertoires would be ideal for T cell clonality determination, though its expense and depth limit its use presently. A particular caveat is that we sequenced resting CD4+ T cells here, as activated CD4+ T cells may be actively dividing and short-lived^{21,24,109}. We might have missed important clonal dynamics in lymphoid associated and tissue-resident CD4, where central memory cells may be marginated until they are activated to become effector memory cells and enter the circulation. Another consideration is that more than one HIV provirus may be integrated into the same rmCD4+ T cell or clonal population. However, multiple proviruses are not often found in the same cell¹¹⁰, and if multiple proviruses were harbored by the same TCR clone this would only add ballast to the argument that HIV dynamics are driven by the largest clonotypes. Although the size of our cohort was limited by the time and expense of the assays, strengths of our study include the integration of large, published data sets and our novel matched longitudinal cohorts, in which PWH with documented ART adherence over 10+ years were age/race/sex matched with HIV-seronegative people.

In the end, our data and modeling promote a picture of powerful rmCD4+ T cell homeostasis between clonal birth and death, with dynamic clonal reshuffling underlying a remarkably stable overall clonality structure within each individual over the course of a decade in mid-life¹⁰⁷. HIV proviruses are similarly clonal to the most clonally-dominant mCD4+ T cell clones, which are likely antigen specific, but deeper

sampling is needed to understand the clonality and dynamics of the less clonally-dominant reservoir. Additional forces of negative selection on intact and some defective proviruses are required; other competing pro-survival mechanisms for proviruses may exist and addition of this modeling term reconciles HIV reservoir stability or even expansion after a decade of ART. In conclusion, the dynamics and clonality of HIV proviruses appear to be predominantly, but not completely, driven by mCD4 T cell kinetics.

Acknowledgements

We thank all the participants who donated their time and samples. We are grateful for discussions of this work with Alison Hill, Andrew Timmons, and Nicole Walch.

Samples and data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) and the Women's Interagency HIV Study (WIHS), now the MACS/WIHS Combined Cohort Study (MWCCS). MWCCS (Principal Investigators): Baltimore CRS (Todd Brown and Joseph Margolick), U01-HL146201; Data Analysis and Coordination Center (Gypsyamber D'Souza, Stephen Gange and Elizabeth Topper), U01-HL146193; Northern California CRS (Bradley Aouizerat, Jennifer Price, and Phyllis Tien), U01-HL146242. The MWCCS is funded primarily by the National Heart, Lung, and Blood Institute (NHLBI), with additional co-funding from the *Eunice Kennedy Shriver* National Institute Of Child Health & Human Development (NICHD), National Institute On Aging (NIA), National Institute Of Dental & Craniofacial Research (NIDCR), National Institute Of Allergy And Infectious Diseases (NIAID), National Institute Of Neurological Disorders And Stroke (NINDS), National Institute Of Mental Health (NIMH), National Institute On Drug Abuse (NIDA), National Institute Of Nursing Research (NINR), National Cancer Institute (NCI), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute on Minority Health and Health Disparities (NIMHD), and in coordination and alignment with the research priorities of the National Institutes of Health, Office of AIDS Research (OAR). MWCCS data collection is also supported by UL1-TR000004 (UCSF CTSA) and UL1-TR003098 (JHU ICTR). The authors gratefully acknowledge the contributions of the study participants and dedication of the staff at the MWCCS sites. Additional data and samples were collected from the Hospital of the University of Pennsylvania and from Philadelphia FIGHT.

The following reagents were obtained through the NIH HIV Reagent Program, Division of AIDS, NIAID, NIH: Peptide Pool, Human Cytomegalovirus (HCMV) pp65, ARP-11549, contributed by DAIDS/NIAID, Peptide Pool, Human Immunodeficiency Virus Type 1 Subtype B (Consensus) *gag* Region, ARP-12425, contributed by DAIDS, NIAID, Peptide Pool, Human Immunodeficiency Virus Type 1 Subtype B (Consensus) *env* Region, ARP-12540, contributed by DAIDS, NIAID, Peptide Pool, Human Immunodeficiency Virus Type 1 Subtype B (Consensus) *pol* Region, ARP-12438, contributed by DAIDS/NIAID, Peptide Pool, Human Immunodeficiency Virus Type 1 Subtype B (Consensus) *nef* Region, ARP-12545, contributed by DAIDS/NIAID, T-20 (N-acetylated derivative; Enfuvirtide acetate salt), ARP-12732, contributed by DAIDS/NIAID (produced by RayBiotech, Inc.).

This research was funded in part by awards (AA, KS, DBR) from the Johns Hopkins University Center for AIDS Research and the University of Washington/Fred Hutch Center for AIDS Research, NIH funded programs (P30AI094189, P30AI02775), which are supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIA, NIGMS, NIDDK, NIMHD. This research was also funded in part by the NIH (grants K25AI155224 - DBR, R01AI150500 - JTS, K08AI143391 - AA, K23AI157875 - MJP, UM1AI126620 - LJM, P30AI094189 - FRS, UM1 AI164566 - FRS, U01 AI177211 - FRS, DP5OD031834 - FRS), the W.W. Smith Charitable Trust (AA), the Pearl M. Stetler Foundation (AA), and the Robert I. Jacobs Fund of The Philadelphia Foundation (LJM). LJM is supported by the Herbert Kean,

M.D., Family Professorship. RS is a Howard Hughes Medical Institute (HHMI) Fellow. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

Conceptualization, AARA and DBR; formal analysis, DBR and AARA; investigation, AARA, DBR, DNR, AR, HZ, LZ, KNS; resources, AARA, DBR, JV, JDS, RFS, LJM, MJP, SGD, RH, LHR, SJG, NRR, PCT, JBM; writing-original draft, AARA and DBR; writing-review&editing, DBR, AARA, FRS, NRR, PCT, SGD, JTS, RFS; funding acquisition, AARA and DBR.

Declaration of interests

MJP reports participation on a data safety and monitoring board for American Gene Technologies.

Code availability

All code to generate all figures and perform simulations will be freely available from <https://github.com/dbrvs/TCR-HIV> at the time of peer-reviewed publication. Files are jupyter Python notebooks and make use of the *numpy*, *scipy*, and *seaborn* packages for modeling and data analysis.

Study participants and clonality data

Participant data are collected in **Table 1** and **Table S1**. For this manuscript, we use the naming convention: PWH22 = PWH-1, PWH548 = PWH-2, PWH746 = PWH-3, and PWH583 = PWH-4. The linkage between participant ID and any identifying information is not known to anyone outside the research groups.

Longitudinally sampled PWH on ART. Four research participants were selected for exceptional viral suppression over 10+ years and longitudinal samples were obtained. PWH-1 had donated samples to the Johns Hopkins HIV Clinical Cohort and research labs at Johns Hopkins, and three participants (PWH-2, -3, and -4) to the SCOPE and OPTIONS cohorts at UCSF. Sample timing, viral load, and absolute CD4+ T cell count data for these four participants are depicted in **Figure S1**. Time point 0 (t0) was a mean of 9.6 months after ART initiation (range 6.3 – 11.0 months). Time point 1 (t1) was a mean of 21.3 months after ART initiation (range 16.6 – 24.0 months). Time point 2 (t2) was a mean of 128.4 months after ART initiation (range 117.6 – 140.2 months). Time point 3 (t3) was obtained for the ViraFEST assay in the two participants willing and available to return for a dedicated research blood draw and was obtained at 189 months after ART initiation in PWH-1 and at 230 months after ART initiation in PWH-4.

Each participant maintained undetectable viral loads — checked at least every 6 months but typically more often — beginning shortly after starting ART through t2, with very few exceptions as follows (**Figure S1**): PWH-2, between t1 and t2 (10.3 years), 2 viral loads were 151 and 93 copies/mL; PWH-3, between t1 and t2 (8.4 years), 1 viral load was 42 copies/mL. Also, there were 6 instances within the >40 participant-years of follow-up in which viral loads were measured more than 6 months apart: PWH-1, once at 9 months apart; PWH-2, once at 7 months apart, once at 12 months apart, and 3 times at 9 months apart. PWH-1, -2 and -4 had negative Hepatitis C virus (HCV) IgG tests after t2, and participant PWH-3 had a negative HCV IgG test after t1 and 3.5 years before t2, with no further HCV testing available.

Longitudinally sampled HIV-seronegative people. Cryopreserved longitudinal samples from one female participant of the WIHS and three male participants from the MACS were obtained⁵¹. These participants were confirmed to be HIV-seronegative at all sampled timepoints. Participants and samples were matched

one-to-one on age, race, sex, and time between samples to the longitudinally sampled PWH on ART. These participants were all HCV seronegative.

Cross-sectionally sampled PWH on ART. *Env* sequences from limiting dilution quantitative viral outgrowth assays from seven participants in a previously published study were obtained⁵². Only *env* sequences from wells that were not treated with autologous IgG were used in this analysis. The HCV serostatus of this cohort is unknown.

Cross-sectionally sampled HIV-seronegative people. Single-timepoint memory CD4+ T cell receptor β chain repertoires from a generally healthy cohort of 33 Belgian individuals were obtained from a public database (De Neuter⁵³). These individuals were all HIV negative at the time of sampling (Dr Benson Ogunjimi, personal communication). It was not possible to match these individuals on demographic characteristics to the cross-sectionally sampled PWH on ART. The HCV serostatus of this cohort is unknown.

Study approval

The studies from which samples were obtained were approved by the UCSF Committee on Human Research, the Johns Hopkins University School of Medicine Institutional Review Board (IRB), and the University of Pennsylvania IRB. All participants provided written informed consent before enrollment.

HIV provirus sequencing

Near-full length HIV provirus sequences from resting CD4+ T cells from the four longitudinally sampled PWH on ART were obtained from a previously published study¹³. *env* sequences from limiting dilution quantitative viral outgrowth assays of resting CD4+ T cells from seven participants in a previously published study were obtained⁵². Only *env* sequences from wells that were not treated with autologous IgG were used in this analysis.

Flow sorting of resting memory CD4+ T cells

Viable frozen PBMCs were quickly thawed at 37°C, rinsed, centrifuged (5 minutes, 500 g), resuspended, and rested overnight in an incubator. Cells were incubated with Human BD Fc Block (BD Pharmingen) at room temperature for 10 minutes. Cells were stained, followed by a 30-minute incubation on ice, with an APC-labeled antibody against CD3 (BioLegend; clone UCHT1), PE-Cy7-labeled antibody against CD4 (BioLegend; clone RPA-T4), BV421-labeled antibody against CD45RA (BioLegend; clone HI100), PE-labeled antibody against CCR7 (BioLegend; clone G043H7), FITC-labeled antibodies against CD69 (BioLegend; clone FN50) and HLA-DR (BioLegend; clone L243), and PE-Cy5-labeled antibodies against CD14 (Thermo Fisher Scientific; clone 61D3), CD16 (BioLegend; clone 3G8), CD20 (BioLegend; clone 2H7), and CD8a (BioLegend; clone RPA-T8). Dead cells were excluded using propidium iodide. Cells stained with single fluorophore-labeled antibodies were used to set sorting gates. A representative gating strategy and sorting logic is provided in **Figure S2**. CD45RA expression was used to distinguish naïve-like (Na) and terminally differentiated cells from memory cells. Central memory (Tcm) cells were distinguished from effector memory (Tem) cells by the expression of CCR7. Cells were sorted using the Beckman Coulter MoFlo Legacy cell sorter.

Isolation of total resting CD4+ T cells

To obtain total (memory and naïve) resting CD4+ T cell TCR β repertoires for the denominator of frequency measurements of random groupings of rmCD4 cells in HIV-seronegative participants, viably frozen PBMCs were quick-thawed at 37°C, rinsed, centrifuged (5 minutes, 500 g), resuspended, and rested overnight in an incubator. Total CD4+ T cells were purified using an immunomagnetic negative selection kit according to the manufacturer's instructions (EasySep™ Human CD4+ T Cell Isolation Kit, STEMCELL Technologies). Immediately afterwards, resting CD4+ T cells (CD4+CD69-CD25-HLA-DR-) were enriched by a second negative depletion (biotin-conjugated anti-CD25, anti-biotin MicroBeads, CD69 MicroBead Kit II, and anti-HLA-DR MicroBeads, Miltenyi Biotec).

TCR β repertoire sequencing

Genomic DNA was isolated (QIAamp DNA Mini or Micro kit, Qiagen, or Quick-DNA Miniprep, Zymo Research) from sorted cells, quantified (Qubit 3.0 or 4 fluorometer, dsDNA quantitation, high sensitivity or broad range assays, ThermoFisher Scientific, or by NanoDrop 2000 spectrophotometer, ThermoFisher Scientific) and diluted in Tris-acetate-EDTA to a concentration of 10-50 ng/ μ L (for a total of up to 1-3 μ g per sample). TCR β sequencing data were generated using the immunoSEQ hsTCR β assay, version 4, deep or ultradeep mode (Adaptive Biotechnologies). TCR β sequences for the ViraFEST assay were generated using the AmpliSeq for Illumina TCR beta-SR panel (Illumina).

Identification of HIV-specific and CMV-specific CD4+ T cells: ViraFEST assay

The viral functional expansion of specific T cells (ViraFEST) assay identifies TCR β clonotypes that proliferate in response to viral antigen^{75,111,112}. Once antigen-specific TCR β DNA sequences were identified by ViraFEST, their corresponding amino acid sequences were identified as antigen-specific within the same participant. When quantifying abundances of antigen-specific TCR β s for a participant, any TCR β DNA sequence that yielded an antigen-specific amino acid sequence for that participant was considered antigen-specific. As outlined in **Figure 4A**, CD8-depleted PBMCs from a fresh blood draw are stimulated with either HIV or CMV viral antigen in the presence of antiretroviral drug to induce antigen-specific expansion followed by TCR β repertoire sequencing of wells to identify clonally expanded, antigen-responsive TCR β sequences. At time point 3 (t3), over 80 mL of peripheral blood was obtained from PWH-1 at 189 months and PWH-4 at 230 months after ART initiation using an acid-citrate-dextrose or ethylenediaminetetraacetic acid (EDTA) anticoagulant and PBMCs were isolated using a Ficoll density gradient as previously described¹¹³. CD8 T cells were depleted from PBMCs (CD8 Dynabeads, ThermoFisher Scientific). 2 million CD8-depleted PBMCs were plated in 2mL culture medium (IMDM, 5% human serum, 50 μ g/mL gentamicin) containing 10 IU/mL IL-2 and 10 μ M enfuvirtide. To each well one of the following was added: (1) either 5 μ g for PWH-4 or 10 μ g for PWH-1 peptide/mL HIV-1 consensus B *gag*, *pol*, *env*, and *nef* overlapping peptide pools (NIH HIV Reagent Program), (2) 10 μ g peptide/mL CMV pp65 protein overlapping peptide pool (NIH HIV Reagent Program) and 3 μ g/mL gradient-purified inactivated CMV (Virusys), (3) 0.5% sterile DMSO by volume as a negative control, or (4) anti-CD3/CD28 antibody-bound beads in a 1:1 bead:cell ratio as a positive control (Human T-Activator CD3/CD28 Dynabeads, ThermoFisher Scientific) and incubated at 37° C. Each assay condition was performed in triplicate. On days 3 and 7, half the media was replaced with fresh culture media containing IL-2 and enfuvirtide. Wells were split on day 7 if needed. On day 10, cells were harvested and CD4+ T cells were isolated using the EasySep CD4+ T cell isolation kit (negative selection, STEMCELL) or the CD4+ T cell isolation kit (negative selection, Miltenyi Biotec). DNA was extracted from the isolated CD4+ T cells using the QIAamp DNA Mini or Micro kit according to the manufacturer's instructions (Qiagen). At the time of plating with antigen, a baseline condition of ~5M CD8-depleted PBMCs was set aside for CD4+ T cell isolation and DNA extraction as above, to assess baseline frequencies of clonotypes.

TCR β repertoire sequencing of CD4+ T cell DNA from each well and data pre-processing were performed by the Johns Hopkins FEST and TCRB Immunogenomics Core Facility as previously described¹¹⁴. Processed data files were uploaded to the publicly available MANAFEST analysis web app (www.stat-apps.onc.jhmi.edu/FEST) to bioinformatically identify antigen-specific T cell clonotypes. The program uses only productive clones and for each clone, the program applies the Fisher exact test to compare the number of templates in a culture of interest and a reference culture. All clonotypes were subject to a 5-template lower threshold across all wells for consideration in the statistical analysis. The following analysis was performed separately to identify HIV-responding clonotypes and CMV-responding clonotypes. Clonotypes were deemed expanded if their abundance was significantly higher in the relevant well relative to a reference well using Fisher exact test with Benjamini-Hochberg correction for false discovery rate (FDR) $p < 0.05$. Antigen-responsive clonotypes met the following criteria: (1) significantly expanded in the antigen-containing well (in at least two of three replicate wells) compared to the reference well (negative control or baseline) by Fisher's exact test with Benjamini-Hochberg FDR, $p < 0.05$, (2) was not significantly expanded in any of the negative wells performed in tandem compared to the reference culture by the same criteria, (3) significantly expanded in at least 2 of 3 antigen-containing wells compared to the reference well (negative control or baseline) with odds ratio > 5 , (4) if there was significant expansion in a negative well compared to reference, then the total abundance of the clonotype must be greater than the maximum of 9 or the baseline well template abundance in all three antigen-containing wells. Antigen-responsive clonotypes that met the following criteria were also included: (1) significantly expanded in at least one antigen-containing well compared to the reference well (negative control or baseline) by Fisher's exact test with Benjamini-Hochberg FDR, $p < 0.05$, (2) was not significantly expanded in any of the negative or positive wells performed in tandem compared to the reference culture by the same criteria, (3) significantly expanded in at least one antigen-containing well compared to the reference well (negative control or baseline) with odds ratio > 5 , (4) was not significantly expanded in any of the negative or positive wells performed in tandem compared to the reference culture with odds ratio > 5 , and (5) its maximum frequency in an antigen-containing well divided by the sum of its maximum frequency in any other non-antigen well plus 0.001 to account for negative frequencies was > 200 .

CMV IgG testing

The presence of CMV IgG was determined qualitatively using an indirect chemiluminescence immunoassay (LIAISON[®] CMV IgG, DiaSorin Inc) in the Johns Hopkins Medical Laboratories on cryopreserved serum samples from t0 or a sample prior to t0 in all longitudinally sampled participants and from the cross-sectionally sampled PWH on ART. The CMV serostatus of the cross-sectionally sampled HIV-negative people was previously published⁵³.

Downsampling rank abundances and computing ecological metrics

To produce rank-abundance distributions (curves), we arrange all clonotypes by their rank ($r \in [1, R]$) from largest to smallest abundance $a(r)$ where R is the richness or total number of unique species in the sample. We then assume that each clones' proportional abundance is given by its observed abundance $p(r) = a(r)/N$ where N is the original experimental sample size (**Figure S3**). Note this means that many sequences are observed with the same abundance, such that their rank must be randomly assigned. Using a multinomial distribution, we then find the downsampled abundance of each clone as $M(p(r), s)$ where s is the resampled size.

After downsampling to the same sizes, we chose to compare distributions summarized with Hill numbers, ecological metrics that can be derived from a mathematically unified equation (**Supplementary Methods 1**) where the x -th Hill number is $D^x = [\sum_i p_i^x]^{1/(1-x)}$ given the clone proportional abundances

p_i where i indicates individual clonotypes. Further, we normalized these metrics to D^x/N^* where N^* is the common re-sample size such that all metrics ranged from zero to one.

Fitting to rank abundances

In the main text, we present the results of fitting with two models: the single and double power law models. Other models were tested including negative binomial, log-series, and exponential power law (see **Supplementary Methods 1**) but none worked as well as these two options and were deemed unnecessarily complex. To fit these models, we sample a model to the same N as the experimental data and estimated the model parameters that minimized (using *scipy.optimize* package) the root-mean-square error between the cumulative proportional abundances of the data and the model. We also used a version of model fitting that minimized the Kolmogorov-Smirnov statistic and found similar results.

Simulating reservoir population decline from TCR β clone data

We began by assuming that HIV DNA is at the population average level across the IPDA data, $H^*(0)=800$ intact and $H(0)=3000$ defective proviruses per million CD4+ T cells, respectively (**Figure 5F**). Then, beginning at t_0 for each participant data set we randomly choose CD4+ T cell clones, add their abundance, and continue until reaching $H(0)$ cells. The number of rmCD4+ T cells in a sample is the sum of all clone abundances $T(t) = \sum_i T_i(t)$ and then the number of HIV proviruses (inclusive of intact, but likely mostly defective) is accordingly $H(t) = \sum_j T_j(t)$ where j are the indices of TCR β clones that are infected -- note for now we assume that all members of a selected TCR β clone harbor proviruses. Thus, we calculate the proportion of total HIV DNA per million CD4+ T cells at the first time point $f_0 = H(t_0)/T(t_0) \times 10^6$.

Then, we examine the abundance of each of the j clones carrying HIV DNA at following time points. Importantly, we use TCR β sequencing inclusive of naïve cells to properly estimate the denominator $T(t_{1,2})$ size, as fractions of naïve CD4+ T cells can change in PWH on ART. At each of the following time points, we calculate the proportion of HIV DNA from clones that were observed in t_0 related to the total CD4+ T cells from all clones (f_1, f_2). This procedure numerically approximates how the proportion of HIV DNA per million CD4+ T cells would change over time if it was purely governed by TCR β clonal dynamics. Finally, from these proportions, we estimated the rate of change as in the modeling of IPDA data, assuming an exponential decay model and fitting using log-linear mixed effects. We replicated this procedure 10 times for each of the 4 Ctl participants who had naïve and memory TCR β data available. Then we calculated a mean rate of change for each participant and then the population mean and 95% CI, which could then be compared to our IPDA estimates.

Simultaneous CD4 carrier and HIV passenger mechanistic model

A complete mathematical description of the mechanistic model is provided in the **Supplementary Methods 1**. We use a tau-leap algorithm¹¹⁵ to stochastically simulate a set of ordinary differential equations (time step $\Delta t=30$ days, simulations over 10 years) where each TCR β clone has an equation with a randomly selected but balanced birth and death rate. Initial clone abundances were drawn from a distribution, which was varied (constant, uniform, exponential, power-law), as were assumptions about the distribution of balanced proliferation and death rates (constant, uniform, exponential). We compared model output to comprehensive experimental data from **Fig 1-4** in this manuscript, determining the optimal model that could match all data by accepting models that fell within experimental ranges for CD4+ T cell and HIV DNA population sizes and longitudinal dynamics and clonality of intact and defective HIV DNA and CD4+ TCR β sequences as well correlations between TCR β clone sizes and fold-changes and the

distribution of TCR β clone expansions/contractions. The metrics that each of 7 models succeeded and failed to match are documented in the **Supplementary Methods 2** file.

References

1. Cohn, L.B., Chomont, N., and Deeks, S.G. (2020). The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. *Cell Host and Microbe* 27, 519–530. 10.1016/j.chom.2020.03.014.
2. Dufour, C., Gantner, P., Fromentin, R., and Chomont, N. (2020). The multifaceted nature of HIV latency. *J Clin Invest* 130, 3381–3390. 10.1172/jci136227.
3. Siliciano, J.D., and Siliciano, R.F. (2021). In Vivo Dynamics of the Latent Reservoir for HIV-1: New Insights and Implications for Cure. *Annu. Rev. Pathol.: Mech. Dis.* 17, 1–24. 10.1146/annurev-pathol-050520-112001.
4. Ho, Y.-C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I.S., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell* 155, 540–551. 10.1016/j.cell.2013.09.020.
5. Bruner, K.M., Wang, Z., Simonetti, F.R., Bender, A.M., Kwon, K.J., Sengupta, S., Fray, E.J., Beg, S.A., Antar, A.A.R., Jenike, K.M., et al. (2019). A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* 566, 1–19. 10.1038/s41586-019-0898-8.
6. Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science* 278, 1295–1300. 10.1126/science.278.5341.1295.
7. Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C., et al. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387, 183–188. 10.1038/387183a0.
8. Siliciano, J.D., Kajdas, J., Finzi, D., Quinn, T.C., Chadwick, K., Margolick, J.B., Kovacs, C., Gange, S.J., and Siliciano, R.F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature Medicine* 9, 727–728. 10.1038/nm880.
9. Crooks, A.M., Bateson, R., Cope, A.B., Dahl, N.P., Griggs, M.K., Kuruc, J.D., Gay, C.L., Eron, J.J., Margolis, D.M., Bosch, R.J., et al. (2015). Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *The Journal of Infectious Diseases* 212, 1361–1365. 10.1093/infdis/jiv218.

10. Peluso, M.J., Bacchetti, P., Ritter, K.D., Beg, S., Lai, J., Martin, J.N., Hunt, P.W., Henrich, T.J., Siliciano, J.D., Siliciano, R.F., et al. (2020). Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy. *JCI Insight* 5, 23–14. 10.1172/jci.insight.132997.
11. Gandhi, R.T., Bosch, R.J., Mar, H., Laird, G.M., Halvas, E.K., Hovind, L., Collier, A.C., Riddler, S.A., Martin, A., Ritter, K., et al. (2023). Varied Patterns of Decay of Intact HIV-1 Proviruses Over two Decades of Art. *J Infect Dis*. 10.1093/infdis/jiad039.
12. McMyn, N.F., Varriale, J., Fray, E.J., Zitzmann, C., MacLeod, H.J., Lai, J., Singhal, A., Moskovljevic, M., Garcia, M.A., Lopez, B.M., et al. (2023). The latent reservoir of inducible, infectious HIV-1 does not decrease despite decades of antiretroviral therapy. *J. Clin. Investig.* 10.1172/jci171554.
13. Antar, A.A.R., Jenike, K.M., Jang, S., Rigau, D.N., Reeves, D.B., Hoh, R., Krone, M.R., Keruly, J.C., Moore, R.D., Schiffer, J.T., et al. (2020). Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy. *The Journal of clinical investigation* 2, 4629. 10.1172/jci135953.
14. Cho, A., Gaebler, C., Oliveira, T., Ramos, V., Saad, M., Lorenzi, J.C.C., Gazumyan, A., Moir, S., Caskey, M., Chun, T.-W., et al. (2022). Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. *P Natl Acad Sci Usa* 119, e2117630119. 10.1073/pnas.2117630119.
15. Joos, B., Fischer, M., Kuster, H., Pillai, S.K., Wong, J.K., Böni, J., Hirschel, B., Weber, R., Trkola, A., Günthard, H.F., et al. (2008). HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proceedings of the National Academy of Sciences of the United States of America* 105, 16725–16730. 10.1073/pnas.0804192105.
16. Hill, A.L., Rosenbloom, D.I.S., Fu, F., Nowak, M.A., and Siliciano, R.F. (2014). Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. *Proceedings of the National Academy of Sciences of the United States of America* 111, 13475–13480. 10.1073/pnas.1406663111.
17. Colby, D.J., Trautmann, L., Pinyakorn, S., Leyre, L., Pagliuzza, A., Kroon, E., Rolland, M., Takata, H., Buranapraditkun, S., Intasan, J., et al. (2018). Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nature Medicine* 24, 1–10. 10.1038/s41591-018-0026-6.
18. Li, J.Z., Aga, E., Bosch, R.J., Pilkinton, M., Kroon, E., MacLaren, L., Keefer, M., Fox, L., Barr, L., Acosta, E., et al. (2021). Time to Viral Rebound After Interruption of Modern Antiretroviral Therapies. *Clin Infect Dis* 74, 865–870. 10.1093/cid/ciab541.

19. Ndung'u, T., McCune, J.M., and Deeks, S.G. (2019). Why and where an HIV cure is needed and how it might be achieved. *Nature* 576, 1–9. 10.1038/s41586-019-1841-8.
20. Margolis, D.M., Archin, N.M., Cohen, M.S., Eron, J.J., Ferrari, G., Garcia, J.V., Gay, C.L., Goonetilleke, N., Joseph, S.B., Swanstrom, R., et al. (2020). Curing HIV: Seeking to Target and Clear Persistent Infection. *Cell* 181, 1–18. 10.1016/j.cell.2020.03.005.
21. Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.-R., Ghattas, G., Brenchley, J.M., et al. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine* 15, 893–900. 10.1038/nm.1972.
22. Kim, M., and Siliciano, R.F. (2016). Reservoir expansion by T-cell proliferation may be another barrier to curing HIV infection. *Proceedings of the National Academy of Sciences* 113, 1692–1694. 10.1073/pnas.1600097113.
23. Bacchus-Souffan, C., Fitch, M., Symons, J., Abdel-Mohsen, M., Reeves, D.B., Hoh, R., Stone, M., Hiatt, J., Kim, P., Chopra, A., et al. (2021). Relationship between CD4 T cell turnover, cellular differentiation and HIV persistence during ART. *PLoS Pathogens* 17, e1009214-24. 10.1371/journal.ppat.1009214.
24. Reeves, D.B., Bacchus-Souffan, C., Fitch, M., Abdel-Mohsen, M., Hoh, R., Ahn, H., Stone, M., Hecht, F., Martin, J., Deeks, S.G., et al. (2023). Estimating the contribution of CD4 T cell subset proliferation and differentiation to HIV persistence. *Nat. Commun.* 14, 6145. 10.1038/s41467-023-41521-1.
25. White, J.A., Simonetti, F.R., Beg, S., McMyn, N.F., Dai, W., Bachmann, N., Lai, J., Ford, W.C., Bunch, C., Jones, J.L., et al. (2022). Complex decay dynamics of HIV virions, intact and defective proviruses, and 2LTR circles following initiation of antiretroviral therapy. *Proc National Acad Sci* 119, e2120326119. 10.1073/pnas.2120326119.
26. Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 345, 179–183. 10.1126/science.1254194.
27. Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y.K., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 345, 570–573. 10.1126/science.1256304.
28. Stockenstrom, S. von, Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Shao, W., Hoh, R., Ho, T., et al. (2015). Longitudinal Genetic Characterization Reveals That Cell

Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *The Journal of Infectious Diseases* 212, 596–607. 10.1093/infdis/jiv092.

29. Bui, J.K., Sobolewski, M.D., Keele, B.F., Spindler, J., Musick, A., Wiegand, A., Luke, B.T., Shao, W., Hughes, S.H., Coffin, J.M., et al. (2017). Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathogens* 13, e1006283-18. 10.1371/journal.ppat.1006283.

30. Hosmane, N.N., Kwon, K.J., Bruner, K.M., Capoferri, A.A., Beg, S., Rosenbloom, D.I.S., Keele, B.F., Ho, Y.-C., Siliciano, J.D., and Siliciano, R.F. (2017). Proliferation of latently infected CD4 +T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *The Journal of Experimental Medicine* 214, 959–972. 10.1084/jem.20170193.

31. Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.-H., Hua, S., Chen, H.-R., Ouyang, Z., et al. (2017). Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *Journal of Clinical Investigation* 127, 2689–2696. 10.1172/jci93289.

32. Simonetti, F.R., Sobolewski, M.D., Fyne, E., Shao, W., Spindler, J., Hattori, J., Anderson, E.M., Watters, S.A., Hill, S., Wu, X., et al. (2016). Clonally expanded CD4 +T cells can produce infectious HIV-1 in vivo. *Proceedings of the National Academy of Sciences* 113, 1883–1888. 10.1073/pnas.1522675113.

33. Lorenzi, J.C.C., Cohen, Y.Z., Cohn, L.B., Kreider, E.F., Barton, J.P., Learn, G.H., Oliveira, T., Lavine, C.L., Horwitz, J.A., Settler, A., et al. (2016). Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proceedings of the National Academy of Sciences* 113, E7908–E7916. 10.1073/pnas.1617789113.

34. Reeves, D.B., Duke, E.R., Wagner, T.A., Palmer, S.E., Spivak, A.M., and Schiffer, J.T. (2018). A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nature Communications* 9, 49–16. 10.1038/s41467-018-06843-5.

35. Guo, S., Luke, B.T., Henry, A.R., Darko, S., Brandt, L.D., Su, L., Sun, D., Wells, D., Joseph, K.W., Demirov, D., et al. (2022). HIV infected CD4+ T cell clones are more stable than uninfected clones during long-term antiretroviral therapy. *Plos Pathog* 18, e1010726. 10.1371/journal.ppat.1010726.

36. Gandhi, R.T., Cyktor, J.C., Bosch, R.J., Mar, H., Laird, G.M., Martin, A., Collier, A.C., Riddler, S.A., Macatangay, B.J., Rinaldo, C.R., et al. (2021). Selective Decay of Intact HIV-1 Proviral DNA on Antiretroviral Therapy. *J Infect Dis* 223, 225–233. 10.1093/infdis/jiaa532.

37. Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L., McElrath, M.J., Lehmann, C., et al. (2015). HIV-1 Integration Landscape during Latent and Active Infection. *Cell* *160*, 420–432. [10.1016/j.cell.2015.01.020](https://doi.org/10.1016/j.cell.2015.01.020).
38. Pinzone, M.R., VanBelzen, D.J., Weissman, S., Bertuccio, M.P., Cannon, L., Venanzi-Rullo, E., Migueles, S., Jones, R.B., Mota, T., Joseph, S.B., et al. (2019). Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nature Communications*, 1–12. [10.1038/s41467-019-08431-7](https://doi.org/10.1038/s41467-019-08431-7).
39. Gantner, P., Pagliuzza, A., Pardons, M., Ramgopal, M., Routy, J.-P., Fromentin, R., and Chomont, N. (2020). Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. *Nature Communications* *11*, 1–9. [10.1038/s41467-020-17898-8](https://doi.org/10.1038/s41467-020-17898-8).
40. Mendoza, P., Jackson, J.R., Oliveira, T.Y., Gaebler, C., Ramos, V., Caskey, M., Jankovic, M., Nussenzweig, M.C., and Cohn, L.B. (2020). Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. *The Journal of Experimental Medicine* *217*. [10.1084/jem.20200051](https://doi.org/10.1084/jem.20200051).
41. Simonetti, F.R., Zhang, H., Soroosh, G.P., Duan, J., Rhodehouse, K., Hill, A.L., Beg, S.A., McCormick, K., Raymond, H.E., Nobles, C.L., et al. (2021). Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. *The Journal of clinical investigation* *131*. [10.1172/jci145254](https://doi.org/10.1172/jci145254).
42. Duette, G., Hiener, B., Morgan, H., Mazur, F.G., Mathivanan, V., Horsburgh, B.A., Fisher, K., Tong, O., Lee, E., Ahn, H., et al. (2022). The HIV-1 proviral landscape reveals Nef contributes to HIV-1 persistence in effector memory CD4+ T-cells. *J Clin Invest*. [10.1172/jci154422](https://doi.org/10.1172/jci154422).
43. Cannon, L. Machine Learning Bolsters Evidence That D1, Nef, and Tat Influence HIV Reservoir Dynamics. *Pathogens and Immunity*.
44. Huang, A.S., Ramos, V., Oliveira, T.Y., Gaebler, C., Jankovic, M., Nussenzweig, M.C., and Cohn, L.B. (2021). Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. *J. Exp. Med.* *218*, e20211427. [10.1084/jem.20211427](https://doi.org/10.1084/jem.20211427).
45. Dragoni, F., Kwaa, A., Traut, C.C.G., Veenhuis, R.T., Woldemeskel, B.A., Camilo-Contreras, A., Raymond, H.E., Dykema, A.G., Scully, E.P., Rosecrans, A.M., et al. (2023). Proviral location affects cognate peptide-induced virus production and immune recognition of HIV-1-infected T cell clones. *J. Clin. Investig.* *133*, e171097. [10.1172/jci171097](https://doi.org/10.1172/jci171097).
46. Coffin, J.M., Bale, M.J., Wells, D., Guo, S., Luke, B., Zerbato, J.M., Sobolewski, M.D., Sia, T., Shao, W., Wu, X., et al. (2021). Integration in oncogenes plays only a minor role in determining the in vivo distribution of HIV integration sites before or during suppressive antiretroviral therapy. *PLoS Pathogens* *17*, e1009141. [10.1371/journal.ppat.1009141](https://doi.org/10.1371/journal.ppat.1009141).

47. Einkauf, K.B., Osborn, M.R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E.M., Gladkov, G.T., Seiger, K.W., Blackmer, J.E., et al. (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell*. 10.1016/j.cell.2021.12.011.
48. Sun, W., Gao, C., Hartana, C.A., Osborn, M.R., Einkauf, K.B., Lian, X., Bone, B., Bonheur, N., Chun, T.-W., Rosenberg, E.S., et al. (2023). Phenotypic signatures of immune selection in HIV-1 reservoir cells. *Nature*, 1–9. 10.1038/s41586-022-05538-8.
49. Lian, X., Gao, C., Sun, X., Jiang, C., Einkauf, K.B., Seiger, K.W., Chevalier, J.M., Yuki, Y., Martin, M., Hoh, R., et al. (2021). Signatures of immune selection in intact and defective proviruses distinguish HIV-1 elite controllers. *Sci Transl Med* 13, eabl4097. 10.1126/scitranslmed.abl4097.
50. Clark, I.C., Mudvari, P., Thaploo, S., Smith, S., Abu-Laban, M., Hamouda, M., Theberge, M., Shah, S., Ko, S.H., Pérez, L., et al. (2023). HIV silencing and cell survival signatures in infected T cell reservoirs. *Nature* 614, 318–325. 10.1038/s41586-022-05556-6.
51. D’Souza, G., Bhondokhan, F., Benning, L., Margolick, J.B., Adedimeji, A.A., Adimora, A.A., Alcaide, M.L., Cohen, M.H., Detels, R., Friedman, M.R., et al. (2021). Characteristics of the MACS/WIHS Combined Cohort Study: Opportunities for Research on Aging With HIV in the Longest US Observational Study of HIV. *Am. J. Epidemiology* 190, 1457–1475. 10.1093/aje/kwab050.
52. Bertagnolli, L.N., Varriale, J., Sweet, S., Brockhurst, J., Simonetti, F.R., White, J., Beg, S., Lynn, K., Mounzer, K., Frank, I., et al. (2020). Autologous IgG antibodies block outgrowth of a substantial but variable fraction of viruses in the latent reservoir for HIV-1. *Proc National Acad Sci* 117, 32066–32077. 10.1073/pnas.2020617117.
53. Neuter, N.D., Bartholomeus, E., Elias, G., Keersmaekers, N., Suls, A., Jansens, H., Smits, E., Hens, N., Beutels, P., Damme, P.V., et al. (2019). Memory CD4+ T cell receptor repertoire data mining as a tool for identifying cytomegalovirus serostatus. *Genes & Immunity*, 1–6. 10.1038/s41435-018-0035-y.
54. Chiffelle, J., Genolet, R., Perez, M.A., Coukos, G., Zoete, V., and Harari, A. (2020). T-cell repertoire analysis and metrics of diversity and clonality. *Curr Opin Biotech* 65, 284–295. 10.1016/j.copbio.2020.07.010.
55. Bender, A.M., Simonetti, F.R., Kumar, M.R., Fray, E.J., Bruner, K.M., Timmons, A.E., Tai, K.Y., Jenike, K.M., Antar, A.A.R., Liu, P.-T., et al. (2019). The Landscape of Persistent Viral Genomes in ART- Treated SIV, SHIV, and HIV-2 Infections. *Cell Host and Microbe* 26, 73-85.e4. 10.1016/j.chom.2019.06.005.
56. Salantes, D.B., Zheng, Y., Mampe, F., Srivastava, T., Beg, S., Lai, J., Li, J.Z., Tressler, R.L., Koup, R.A., Hoxie, J., et al. (2018). HIV-1 latent reservoir size and diversity are stable following brief treatment interruption. *Journal of Clinical Investigation* 128, 3102–3115. 10.1172/jci120194.

57. Wang, Z., Gurule, E.E., Brennan, T.P., Gerold, J.M., Kwon, K.J., Hosmane, N.N., Kumar, M.R., Beg, S.A., Capoferri, A.A., Ray, S.C., et al. (2018). Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proceedings of the National Academy of Sciences* *115*, E2575–E2584. [10.1073/pnas.1720665115](https://doi.org/10.1073/pnas.1720665115).
58. Sun, X., Nguyen, T., Achour, A., Ko, A., Cifello, J., Ling, C., Sharma, J., Hiroi, T., Zhang, Y., Chia, C.W., et al. (2022). Longitudinal analysis reveals age-related changes in the T cell receptor repertoire of human T cell subsets. *J Clin Invest* *132*, e158122. [10.1172/jci158122](https://doi.org/10.1172/jci158122).
59. Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* *27*, 379–423. [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
60. SIMPSON, E.H. (1949). Measurement of Diversity. *Nature* *163*, 688–688. [10.1038/163688a0](https://doi.org/10.1038/163688a0).
61. Roswell, M., Dushoff, J., and Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos* *130*, 321–338. [10.1111/oik.07202](https://doi.org/10.1111/oik.07202).
62. Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., and Ellison, A.M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* *84*, 45–67. [10.1890/13-0133.1](https://doi.org/10.1890/13-0133.1).
63. Zhang, W., Morris, A.B., Peek, E.V., Karadkhele, G., Robertson, J.M., Kissick, H.T., and Larsen, C.P. (2021). CMV Status Drives Distinct Trajectories of CD4+ T Cell Differentiation. *Front. Immunol.* *12*, 620386. [10.3389/fimmu.2021.620386](https://doi.org/10.3389/fimmu.2021.620386).
64. Pourgheysari, B., Khan, N., Best, D., Bruton, R., Nayak, L., and Moss, P.A.H. (2007). The Cytomegalovirus-Specific CD4 + T-Cell Response Expands with Age and Markedly Alters the CD4 + T-Cell Repertoire. *J. Virol.* *81*, 7759–7765. [10.1128/jvi.01262-06](https://doi.org/10.1128/jvi.01262-06).
65. Khan, N., Shariff, N., Cobbold, M., Bruton, R., Ainsworth, J.A., Sinclair, A.J., Nayak, L., and Moss, P.A.H. (2002). Cytomegalovirus Seropositivity Drives the CD8 T Cell Repertoire Toward Greater Clonality in Healthy Elderly Individuals. *J Immunol* *169*, 1984–1992. [10.4049/jimmunol.169.4.1984](https://doi.org/10.4049/jimmunol.169.4.1984).
66. Sylwester, A.W., Mitchell, B.L., Edgar, J.B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P.R., Grabstein, K.H., Hosken, N.A., Kern, F., et al. (2005). Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *J. Exp. Med.* *202*, 673–685. [10.1084/jem.20050882](https://doi.org/10.1084/jem.20050882).
67. McManus, W.R., Bale, M.J., Spindler, J., Wiegand, A., Musick, A., Patro, S.C., Sobolewski, M.D., Musick, V.K., Anderson, E.M., Cyktor, J.C., et al. (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *The Journal of clinical investigation* *130*, 4629–4642. [10.1172/jci126714](https://doi.org/10.1172/jci126714).

68. Bozzi, G., Simonetti, F.R., Watters, S.A., Anderson, E.M., Gouzoulis, M., Kearney, M.F., Rote, P., Lange, C., Shao, W., Gorelick, R., et al. (2019). No evidence of ongoing HIV replication or compartmentalization in tissues during combination antiretroviral therapy: Implications for HIV eradication. *Science advances* 5, eaav2045. [10.1126/sciadv.aav2045](https://doi.org/10.1126/sciadv.aav2045).
69. Kearney, M.F., Wiegand, A., Shao, W., McManus, W.R., Bale, M.J., Luke, B., Maldarelli, F., Mellors, J.W., and Coffin, J.M. (2017). Ongoing HIV Replication During ART Reconsidered. *Open Forum Infectious Diseases* 4, e1004010-5. [10.1093/ofid/ofx173](https://doi.org/10.1093/ofid/ofx173).
70. Boshier, F.A.T., Reeves, D.B., Duke, E.R., Swan, D.A., Prlic, M., Cardozo-Ojeda, E.F., and Schiffer, J.T. (2022). Substantial uneven proliferation of CD4+ T cells during recovery from acute HIV infection is sufficient to explain the observed expanded clones in the HIV reservoir. *J Virus Erad* 8, 100091. [10.1016/j.jve.2022.100091](https://doi.org/10.1016/j.jve.2022.100091).
71. Naumov, Y.N., Naumova, E.N., Hogan, K.T., Selin, L.K., and Gorski, J. (2003). A Fractal Clonotype Distribution in the CD8+ Memory T Cell Repertoire Could Optimize Potential for Immune Responses. *J Immunol* 170, 3994–4001. [10.4049/jimmunol.170.8.3994](https://doi.org/10.4049/jimmunol.170.8.3994).
72. Meier, J., Roberts, C., Avent, K., Hazlett, A., Berrie, J., Payne, K., Hamm, D., Desmarais, C., Sanders, C., Hogan, K.T., et al. (2013). Fractal Organization of the Human T Cell Repertoire in Health and after Stem Cell Transplantation. *Biol Blood Marrow Tr* 19, 366–377. [10.1016/j.bbmt.2012.12.004](https://doi.org/10.1016/j.bbmt.2012.12.004).
73. Henrich, T.J., Hobbs, K.S., Hanhauser, E., Scully, E., Hogan, L.E., Robles, Y.P., Leadabrand, K.S., Marty, F.M., Palmer, C.D., Jost, S., et al. (2017). Human Immunodeficiency Virus Type 1 Persistence Following Systemic Chemotherapy for Malignancy. *The Journal of Infectious Diseases* 216, 254–262. [10.1093/infdis/jix265](https://doi.org/10.1093/infdis/jix265).
74. Cole, B., Lambrechts, L., Boyer, Z., Noppe, Y., Scheerder, M.-A.D., Eden, J.-S., Vrancken, B., Schlub, T.E., McLaughlin, S., Frenkel, L.M., et al. (2022). Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia before and during analytical treatment interruption. *Cell Reports* 39, 110739. [10.1016/j.celrep.2022.110739](https://doi.org/10.1016/j.celrep.2022.110739).
75. Chan, H.Y., Zhang, J., Garliss, C.C., Kwaa, A.K., Blankson, J.N., and Smith, K.N. (2020). A T Cell Receptor Sequencing-Based Assay Identifies Cross-Reactive Recall CD8+ T Cell Clonotypes Against Autologous HIV-1 Epitope Variants. *Front Immunol* 11, 591. [10.3389/fimmu.2020.00591](https://doi.org/10.3389/fimmu.2020.00591).
76. Poloni, C., Schonhofer, C., Ivison, S., Levings, M.K., Steiner, T.S., and Cook, L. (2023). T-cell activation–induced marker assays in health and disease. *Immunol. Cell Biol.* 101, 491–503. [10.1111/imcb.12636](https://doi.org/10.1111/imcb.12636).
77. Abana, C.O., Pilkinton, M.A., Gaudieri, S., Chopra, A., McDonnell, W.J., Wanjalla, C., Barnett, L., Gangula, R., Hager, C., Jung, D.K., et al. (2017). Cytomegalovirus (CMV) Epitope–Specific

CD4+ T Cells Are Inflated in HIV+ CMV+ Subjects. *J. Immunol.* *199*, 3187–3201. 10.4049/jimmunol.1700851.

78. Sponaugle, A., Weideman, A.M.K., Ranek, J., Atassi, G., Kuruc, J., Adimora, A.A., Archin, N.M., Gay, C., Kuritzkes, D.R., Margolis, D.M., et al. (2023). Dominant CD4+ T cell receptors remain stable throughout antiretroviral therapy-mediated immune restoration in people with HIV. *Cell Rep. Med.* *4*, 101268. 10.1016/j.xcrm.2023.101268.

79. Levy, C.N., Hughes, S.M., Roychoudhury, P., Reeves, D.B., Amstuz, C., Zhu, H., Huang, M.-L., Wei, Y., Bull, M.E., Cassidy, N.A.J., et al. (2021). A highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral reservoir. *Cell Reports Medicine*, 100243. 10.1016/j.xcrm.2021.100243.

80. Reeves, D.B., Gaebler, C., Oliveira, T.Y., Peluso, M.J., Schiffer, J.T., Cohn, L.B., Deeks, S.G., and Nussenzweig, M.C. (2023). Impact of misclassified defective proviruses on HIV reservoir measurements. *Nat. Commun.* *14*, 4186. 10.1038/s41467-023-39837-z.

81. Swan, D.A., Rolland, M., Herbeck, J.T., Schiffer, J.T., and Reeves, D.B. (2022). Evolution during primary HIV infection does not require adaptive immune selection. *Proc National Acad Sci* *119*, e2109172119. 10.1073/pnas.2109172119.

82. Towler, A.M.H., Ravishankar, S., Coffey, D.G., Puro, C.E., and Warren, E.H. (2022). Serial Analysis of the T-Cell Receptor β -Chain Repertoire in People Living With HIV Reveals Incomplete Recovery After Long-Term Antiretroviral Therapy. *Front Immunol* *13*, 879190. 10.3389/fimmu.2022.879190.

83. Gaimann, M.U., Nguyen, M., Desponds, J., and Mayer, A. (2020). Early life imprints the hierarchy of T cell clone sizes. *Elife* *9*, e61639. 10.7554/elife.61639.

84. Macallan, D.C., Busch, R., and Asquith, B. (2019). Current estimates of T cell kinetics in humans. *Curr Opin Syst Biology* *18*, 77–86. 10.1016/j.coisb.2019.10.002.

85. Collora, J.A., Liu, R., Pinto-Santini, D., Ravindra, N., Ganoza, C., Lama, J.R., Alfaro, R., Chiarella, J., Spudich, S., Mounzer, K., et al. (2022). Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity*. 10.1016/j.immuni.2022.03.004.

86. Lau, C.-Y., Adan, M.A., and Maldarelli, F. (2021). Why the HIV Reservoir Never Runs Dry: Clonal Expansion and the Characteristics of HIV-Infected Cells Challenge Strategies to Cure and Control HIV Infection. *Viruses* *13*, 2512. 10.3390/v13122512.

87. Liu, R., Simonetti, F.R., and Ho, Y.-C. (2020). The forces driving clonal expansion of the HIV-1 latent reservoir. *Virology Journal* *17*, 1–13. 10.1186/s12985-019-1276-8.

88. Dinoso, J.B., Kim, S.Y., Wiegand, A.M., Palmer, S.E., Gange, S.J., Cranmer, L., O’Shea, A., Callender, M., Spivak, A., Brennan, T., et al. (2009). Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 9403–9408. 10.1073/pnas.0903107106.
89. Mold, J.E., Réu, P., Olin, A., Bernard, S., Michaëlsson, J., Rane, S., Yates, A., Khosravi, A., Salehpour, M., Possnert, G., et al. (2019). Cell generation dynamics underlying naive T-cell homeostasis in adult humans. *PLOS Biology* *17*, e3000383-26. 10.1371/journal.pbio.3000383.
90. Wagner, T.A., McKernan, J.L., Tobin, N.H., Tapia, K.A., Mullins, J.I., and Frenkel, L.M. (2013). An Increasing Proportion of Monotypic HIV-1 DNA Sequences during Antiretroviral Treatment Suggests Proliferation of HIV-Infected Cells. *Journal of Virology* *87*, 1770–1778. 10.1128/jvi.01985-12.
91. Besson, G.J., Lalama, C.M., Bosch, R.J., Gandhi, R.T., Bedison, M.A., Aga, E., Riddler, S.A., McMahan, D.K., Hong, F., and Mellors, J.W. (2014). HIV-1 DNA Decay Dynamics in Blood During More Than a Decade of Suppressive Antiretroviral Therapy. *Clinical Infectious Diseases* *59*, 1312–1321. 10.1093/cid/ciu585.
92. Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.-Y., Olshen, R.A., Weyand, C.M., Boyd, S.D., and Goronzy, J.J. (2014). Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences of the United States of America* *111*, 13139–13144. 10.1073/pnas.1409155111.
93. Ndifon, W., and Dushoff, J. (2016). The Hayflick Limit May Determine the Effective Clonal Diversity of Naive T Cells. *J Immunol* *196*, 4999–5004. 10.4049/jimmunol.1502343.
94. Borgognone, A., Noguera-Julian, M., Oriol, B., Noël-Romas, L., Ruiz-Riol, M., Guillén, Y., Parera, M., Casadellà, M., Duran, C., Puertas, M.C., et al. (2022). Gut microbiome signatures linked to HIV-1 reservoir size and viremia control. *Microbiome* *10*, 59. 10.1186/s40168-022-01247-6.
95. Lythe, G., Callard, R.E., Hoare, R.L., and Molina-París, C. (2016). How many TCR clonotypes does a body maintain? *Journal of Theoretical Biology* *389*, 214–224. 10.1016/j.jtbi.2015.10.016.
96. Desponds, J., Mora, T., and Walczak, A.M. (2016). Fluctuating fitness shapes the clone-size distribution of immune repertoires. *Proceedings of the National Academy of Sciences* *113*, 274–279. 10.1073/pnas.1512977112.
97. Greef, P.C. de, Oakes, T., Gerritsen, B., Ismail, M., Heather, J.M., Hermsen, R., Chain, B., and Boer, R.J.D. (2020). The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes. *eLife* *9*, 958–24. 10.7554/elife.49900.

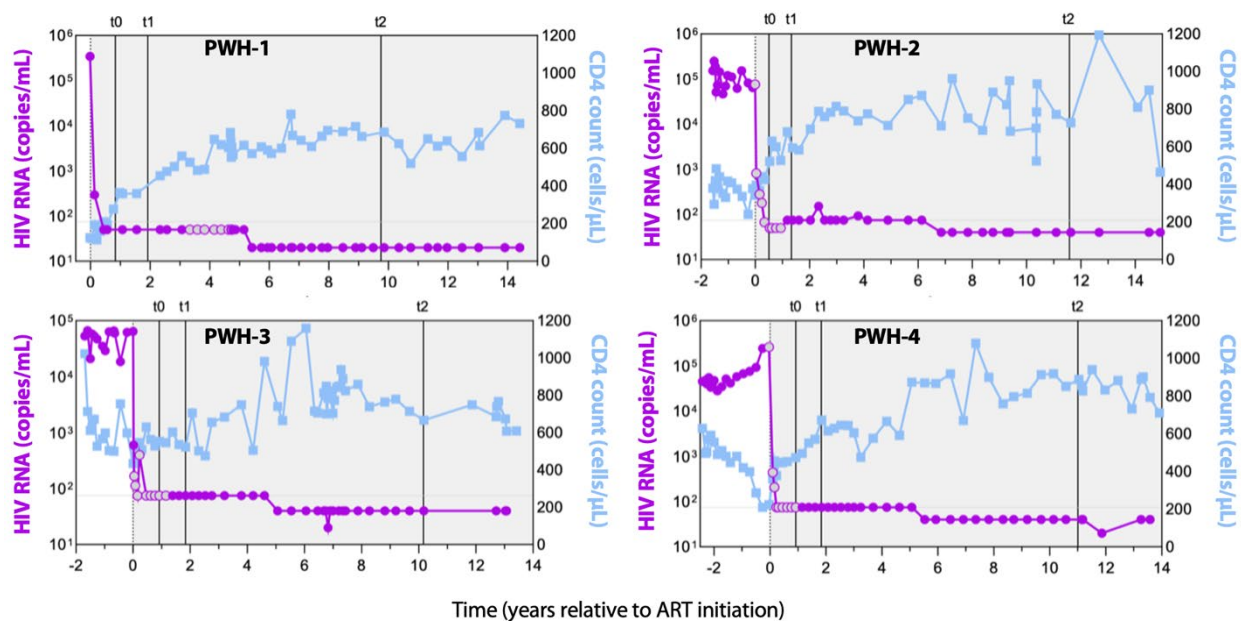
98. Dessalles, R., D'Orsogna, M., and Chou, T. (2018). Exact Steady-State Distributions of Multispecies Birth–Death–Immigration Processes: Effects of Mutations and Carrying Capacity on Diversity. *J Stat Phys* *173*, 182–221. [10.1007/s10955-018-2128-4](https://doi.org/10.1007/s10955-018-2128-4).
99. Dessalles, R., Pan, Y., Xia, M., Maestrini, D., D'Orsogna, M.R., and Chou, T. (2022). How Naive T-Cell Clone Counts Are Shaped By Heterogeneous Thymic Output and Homeostatic Proliferation. *Front Immunol* *12*, 735135. [10.3389/fimmu.2021.735135](https://doi.org/10.3389/fimmu.2021.735135).
100. Murray, J.M. (2021). Dynamics of latent HIV under clonal expansion. *Plos Pathog* *17*, e1010165. [10.1371/journal.ppat.1010165](https://doi.org/10.1371/journal.ppat.1010165).
101. Rong, L., and Perelson, A.S. (2009). Modeling Latently Infected Cell Activation: Viral and Latent Reservoir Persistence, and Viral Blips in HIV-infected Patients on Potent Therapy. *PLoS computational biology* *5*, e1000533-18. [10.1371/journal.pcbi.1000533](https://doi.org/10.1371/journal.pcbi.1000533).
102. Conway, J.M., and Perelson, A.S. (2016). Residual Viremia in Treated HIV+ Individuals. *PLoS computational biology* *12*, e1004677-19. [10.1371/journal.pcbi.1004677](https://doi.org/10.1371/journal.pcbi.1004677).
103. Mayer, A., Zhang, Y., Perelson, A.S., and Wingreen, N.S. (2019). Regulation of T cell expansion by antigen presentation dynamics. *Proceedings of the National Academy of Sciences of the United States of America* *116*, 5914–5919. [10.1073/pnas.1812800116](https://doi.org/10.1073/pnas.1812800116).
104. Bosque, A., Famiglietti, M., Weyrich, A.S., Goulston, C., and Planelles, V. (2011). Homeostatic Proliferation Fails to Efficiently Reactivate HIV-1 Latently Infected Central Memory CD4+ T Cells. *Plos Pathog* *7*, e1002288. [10.1371/journal.ppat.1002288](https://doi.org/10.1371/journal.ppat.1002288).
105. Huang, S.-H., Ren, Y., Thomas, A.S., Chan, D., Mueller, S., Ward, A.R., Patel, S., Bollard, C.M., Cruz, C.R., Karandish, S., et al. (2018). Latent HIV reservoirs exhibit inherent resistance to elimination by CD8+ T cells. *Journal of Clinical Investigation* *128*, 876–889. [10.1172/jci97555](https://doi.org/10.1172/jci97555).
106. Wonderlich, E.R., Subramanian, K., Cox, B., Wiegand, A., Lackman-Smith, C., Bale, M.J., Stone, M., Hoh, R., Kearney, M.F., Maldarelli, F., et al. (2019). Effector memory differentiation increases detection of replication-competent HIV-1 in resting CD4+ T cells from virally suppressed individuals. *Plos Pathog* *15*, e1008074. [10.1371/journal.ppat.1008074](https://doi.org/10.1371/journal.ppat.1008074).
107. Gossel, G., Hogan, T., Cownden, D., Seddon, B., and Yates, A.J. (2017). Memory CD4 T cell subsets are kinetically heterogeneous and replenished from naive T cells at high levels. *Elife* *6*, e23013. [10.7554/elife.23013](https://doi.org/10.7554/elife.23013).
108. Laskey, S.B., Pohlmeier, C.W., Bruner, K.M., and Siliciano, R.F. (2016). Evaluating Clonal Expansion of HIV-Infected Cells: Optimization of PCR Strategies to Predict Clonality. *PLoS Pathogens* *12*, e1005689-17. [10.1371/journal.ppat.1005689](https://doi.org/10.1371/journal.ppat.1005689).

109. Asquith, B., Borghans, J.A.M., Ganusov, V.V., and Macallan, D.C. (2009). Lymphocyte kinetics in health and disease. *Trends Immunol* *30*, 182–189. 10.1016/j.it.2009.01.003.
110. Josefsson, L., King, M.S., Makitalo, B., Brännström, J., Shao, W., Maldarelli, F., Kearney, M.F., Hu, W.-S., Chen, J., Gaines, H., et al. (2011). Majority of CD4+ T cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. *Proc. Natl. Acad. Sci.* *108*, 11199–11204. 10.1073/pnas.1107729108.
111. Dykema, A.G., Zhang, B., Woldemeskel, B.A., Garliss, C.C., Cheung, L.S., Choudhury, D., Zhang, J., Aparicio, L., Bom, S., Rashid, R., et al. (2021). Functional characterization of CD4+ T-cell receptors cross-reactive for SARS-CoV-2 and endemic coronaviruses. *J. Clin. Investig.* *131*. 10.1172/jci146922.
112. Danilova, L., Anagnostou, V., Caushi, J.X., Sidhom, J.-W., Guo, H., Chan, H.Y., Suri, P., Tam, A.J., Zhang, J., Asmar, M.E., et al. (2018). The Mutation-Associated Neoantigen Functional Expansion of Specific T cells (MANAFEST) assay: a sensitive platform for monitoring antitumor immunity. *Cancer Immunol. Res.* *6*, canimm.0129.2018. 10.1158/2326-6066.cir-18-0129.
113. Siliciano, J.D., and Siliciano, R.F. (2005). Human Retrovirus Protocols, *Virology and Molecular Biology*. *Methods Mol. Biol. (Clifton, NJ)* *304*, 3–15. 10.1385/1-59259-907-9:003.
114. Dykema, A.G., Zhang, B., Woldemeskel, B.A., Garliss, C.C., Rashid, R., Westlake, T., Zhang, L., Zhang, J., Cheung, L.S., Caushi, J.X., et al. (2022). SARS-CoV-2 vaccination diversifies the CD4+ spike-reactive T cell repertoire in patients with prior SARS-CoV-2 infection. *eBioMedicine* *80*, 104048. 10.1016/j.ebiom.2022.104048.
115. Gillespie, D.T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics* *115*, 1716–1733. 10.1063/1.1378322.

Supplementary Tables and Figures

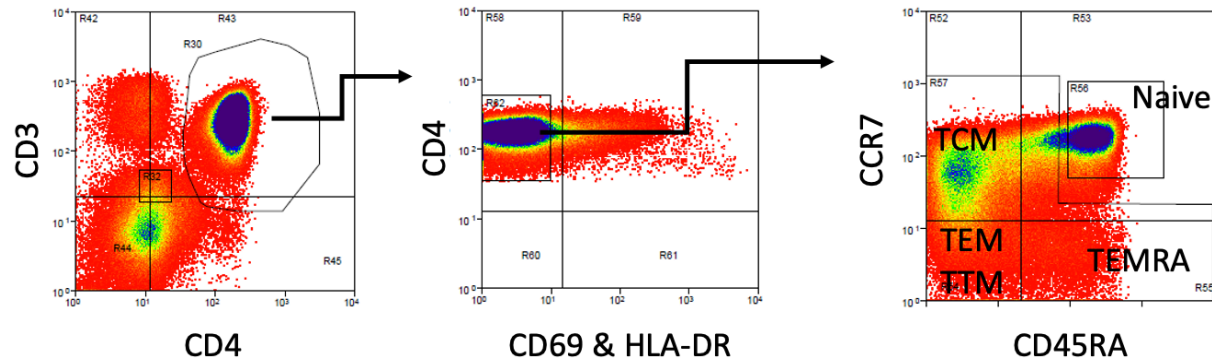
Supplementary Table 1. Detailed participant and sample characteristics. [Separate file]

Abbreviations: AA, African American; ABC, abacavir; ART, antiretroviral; CMV, cytomegalovirus; DTG, dolutegravir; EFV, efavirenz; EVG/c, elvitegravir/cobicistat; FTC, emtricitabine; JHHCC, Johns Hopkins HIV Clinical Cohort; m, months; LPV/r, lopinavir/ritonavir; ND, not done; NFL, near-full length; PWH, people living with HIV; RPV, rilpivirine; t0, t1, etc, study timepoint 1, 2; TCR β , T cell receptor beta chain; TDF, tenofovir disoproxil fumarate; Unk, unknown; W, White; ZDV, zidovudine; 3TC, lamivudine.



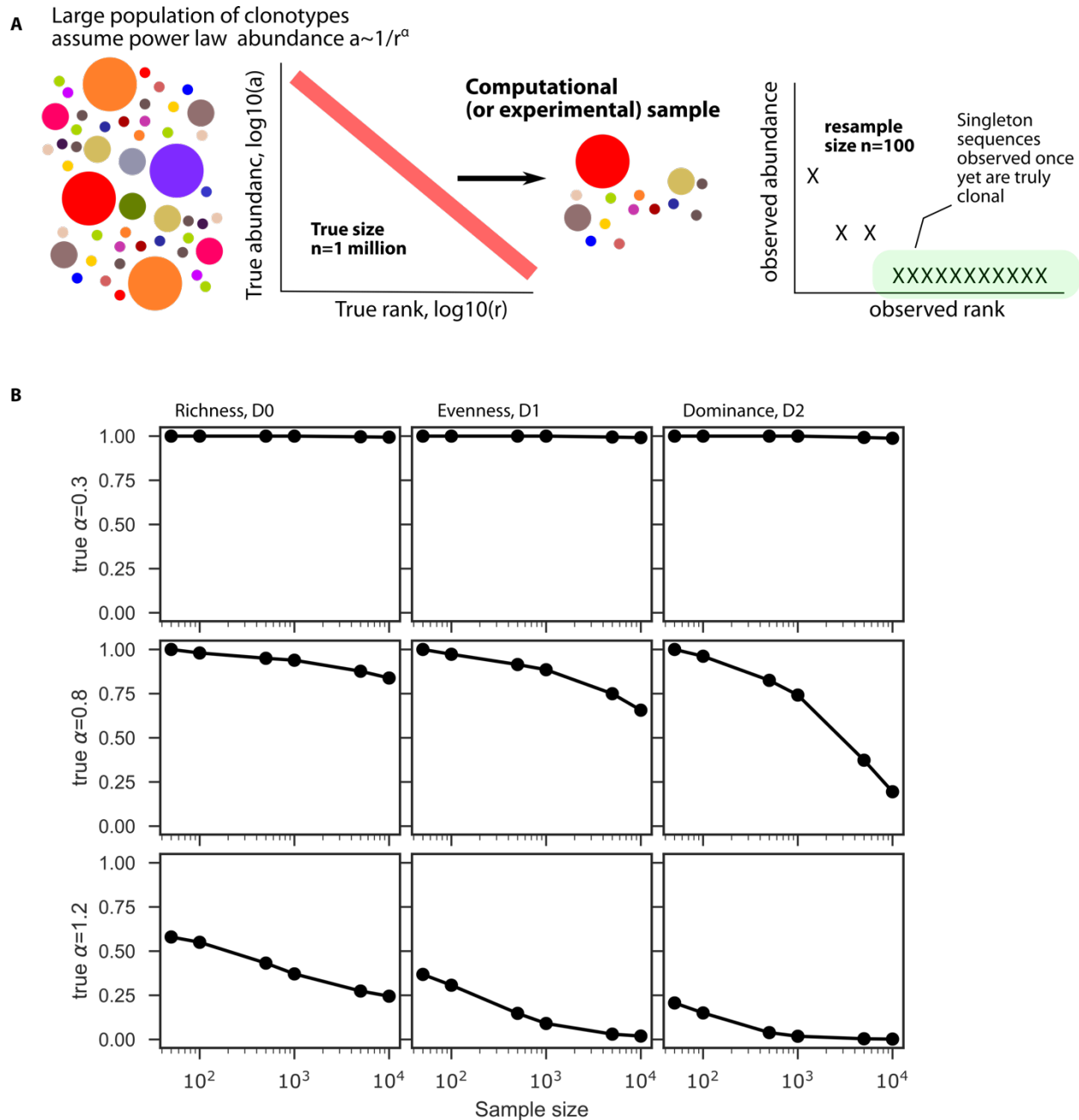
Supplementary Figure 1. Longitudinal participant viral loads and absolute CD4+ T cell counts.

HIV-1 RNA copies/mL (purple) and absolute CD4+ T cells/ μ L (blue) in the peripheral circulation of the four PWH on ART sampled longitudinally in this study. Grey shading indicates time on ART. Sample time points t0, t1, and t2 are shown for each participant. Resting memory CD4+ TCR β repertoires were sequenced at t0, t1, and t2. Near-full length HIV proviruses isolated from resting CD4+ T cells were sequenced at t1 and t2.

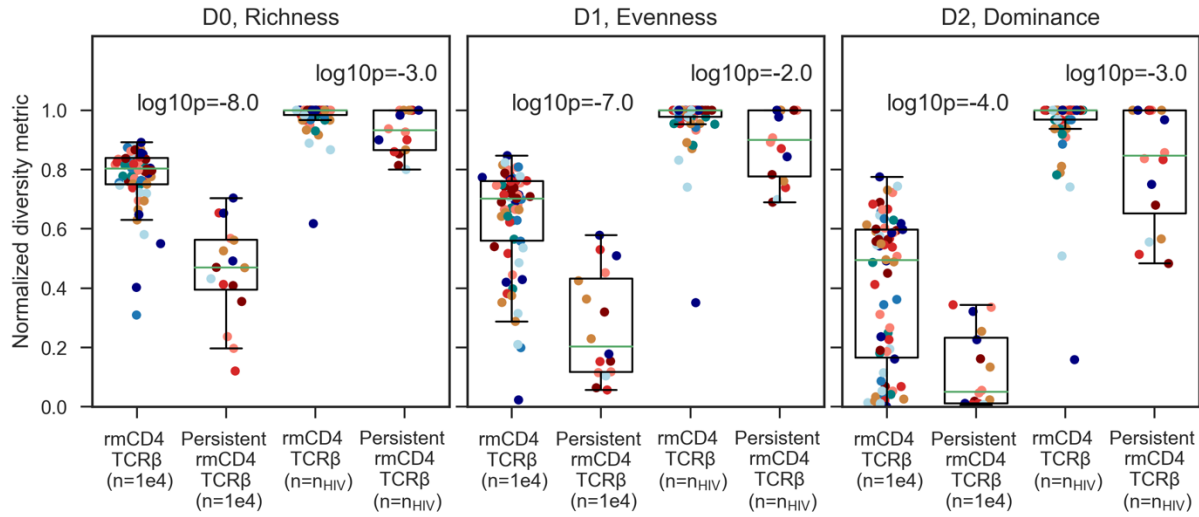


Supplementary Figure 2. Gating strategy for FACS sorting of resting memory CD4+ T cells.

A dump channel included CD8+, CD14+, CD16+, CD20+, and non-viable cells (Methods). Selected cells were CD8-CD14-CD16-CD20-CD3+CD4+CD69+HLA-DR- and either CCR7- or CCR7+CD45RA- (collected together). TCM, T central memory cells, TEM, T effector memory cells, TTM, T transitional memory cells, TEMRA, T effector memory cells re-expressing CD45RA.

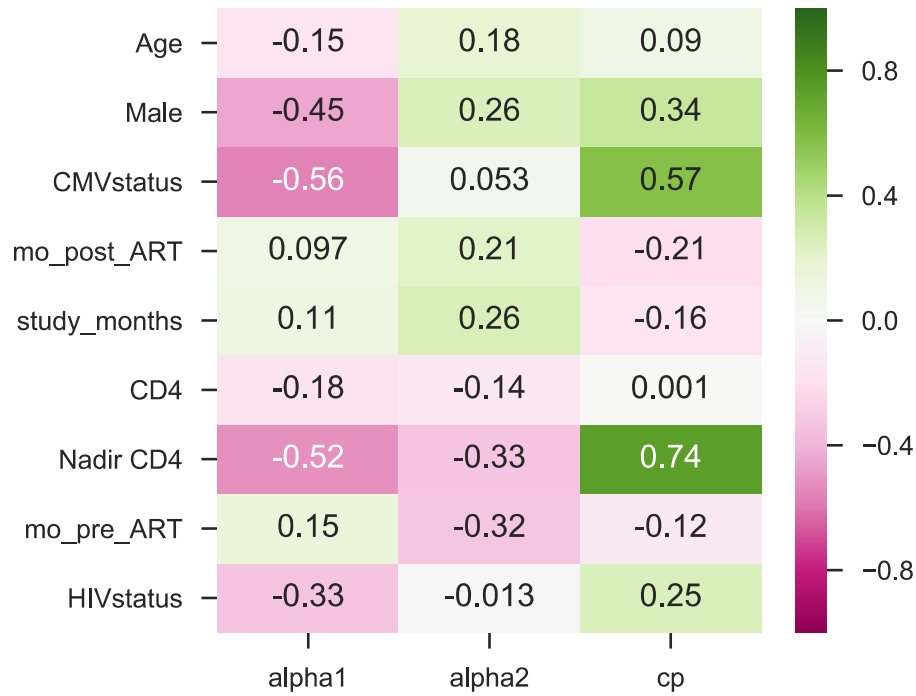


Supplementary Figure 3. Simulation study to illustrate influence of sample size on ecological metrics. A) Model schematic, we assume clones follow a strict power law rank abundance relationship with exponent alpha, $a(r) \sim r^{-\alpha}$ and true size (number of cells) $n=10^6$ and then resample these distributions to a lower sample size. Resampled rank abundance curves (here $n=100$) can contain many observed singletons that are truly clonal in the original distribution. B) Three “Hill” diversity metrics computed from different underlying power law exponents (rows) at varying sample sizes (x-axis) illustrate that observed clonality depends on sample sizes, and this dependence grows stronger with alpha values.

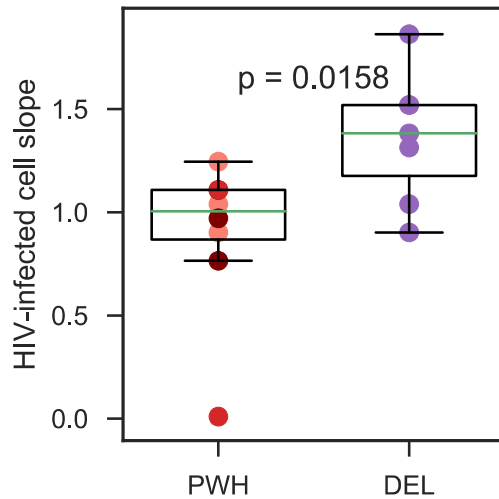


Supplementary Figure 4. Comparison of normalized diversity metrics for TCRβ and persistent TCRβ at resampled sizes for typical TCRβ and HIV proviruses.

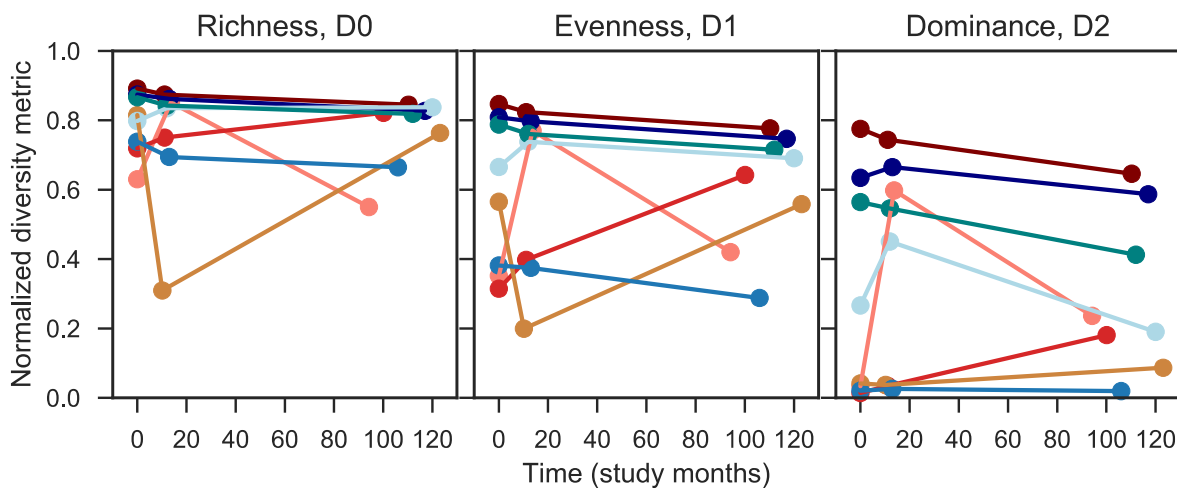
Ecological diversity metrics of rmCD4 TCRβs and persistent rmCD4 TCRβs downsampled to $n=10^4$ (left) and downsampled to the sample size of HIV proviruses from that participant time point (right). HIV proviruses were not sequenced at t₀, so we set the downsampled size for TCRβ sequences at t₀ to $n_{HIV}=60$. P-values indicate paired differences using a one-sided Mann-Whitney test.



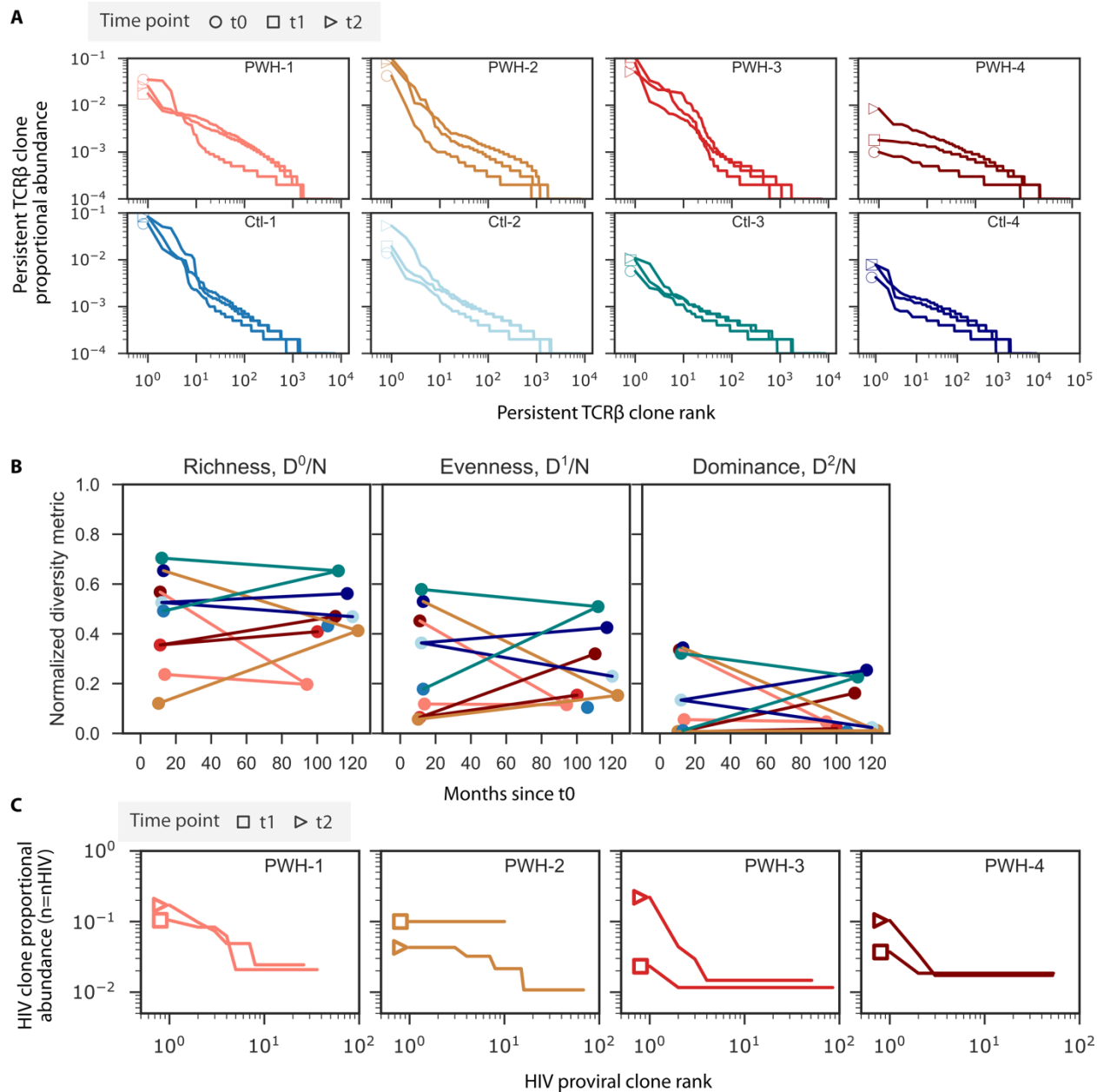
Supplementary Figure 5. Spearman correlation coefficients computed between power law slope parameters and clinical variables. We assessed associations between large clone slope (alpha1), smaller clone slope (alpha2), and change point of that slope (cp, the rank where the slope changes) and 9 clinical variables (y-axis). Associations were deemed significant with a Bonferonni correction threshold of $p < 0.05/27 = 0.002$, which occurs in this data set when correlation coefficients have absolute values above 0.4. Thus, significant associations are observed between alpha1 and Male sex, CMV status, and Nadir CD4 count; and between cp and CMV status and Nadir CD4 count.



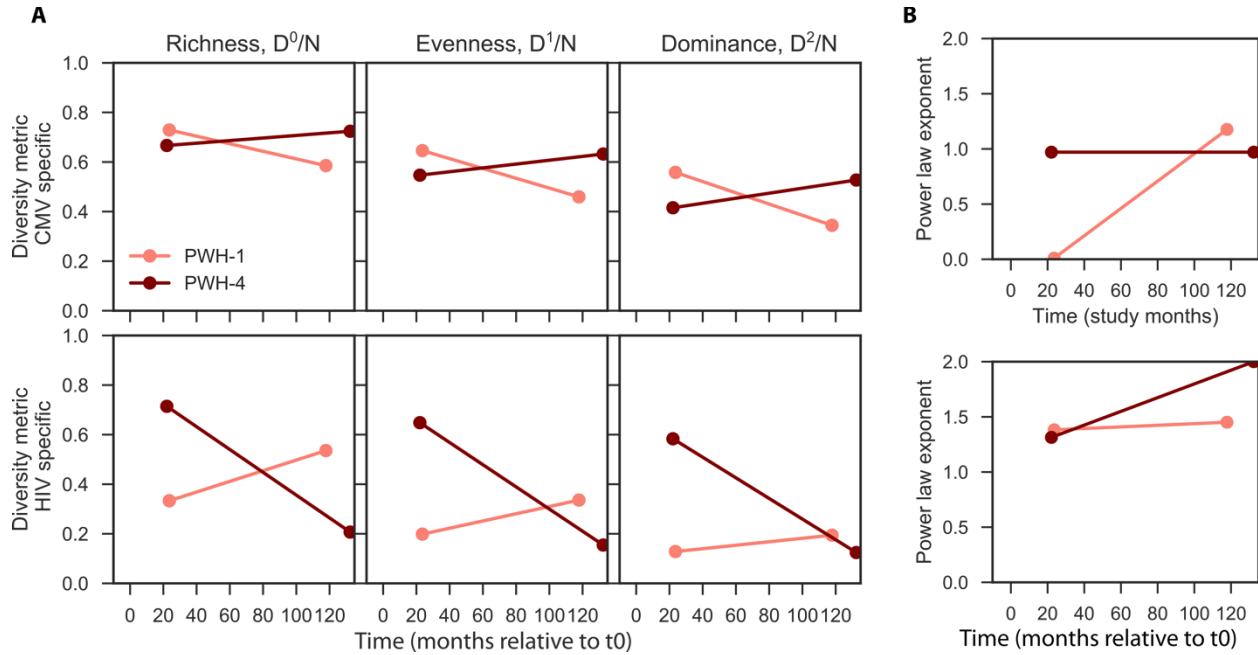
Supplementary Figure 6. Power law slope for HIV-infected cells from two cohorts. DEL cohort HIV reservoir cells (*Env* sequences from intact, inducible proviruses) had significantly higher slopes compared to the PWH cohort HIV-infected cells (near-full length sequences from intact and defective proviruses) ($p=0.02$, 1 sided Mann-Whitney U test) indicating higher clonality.



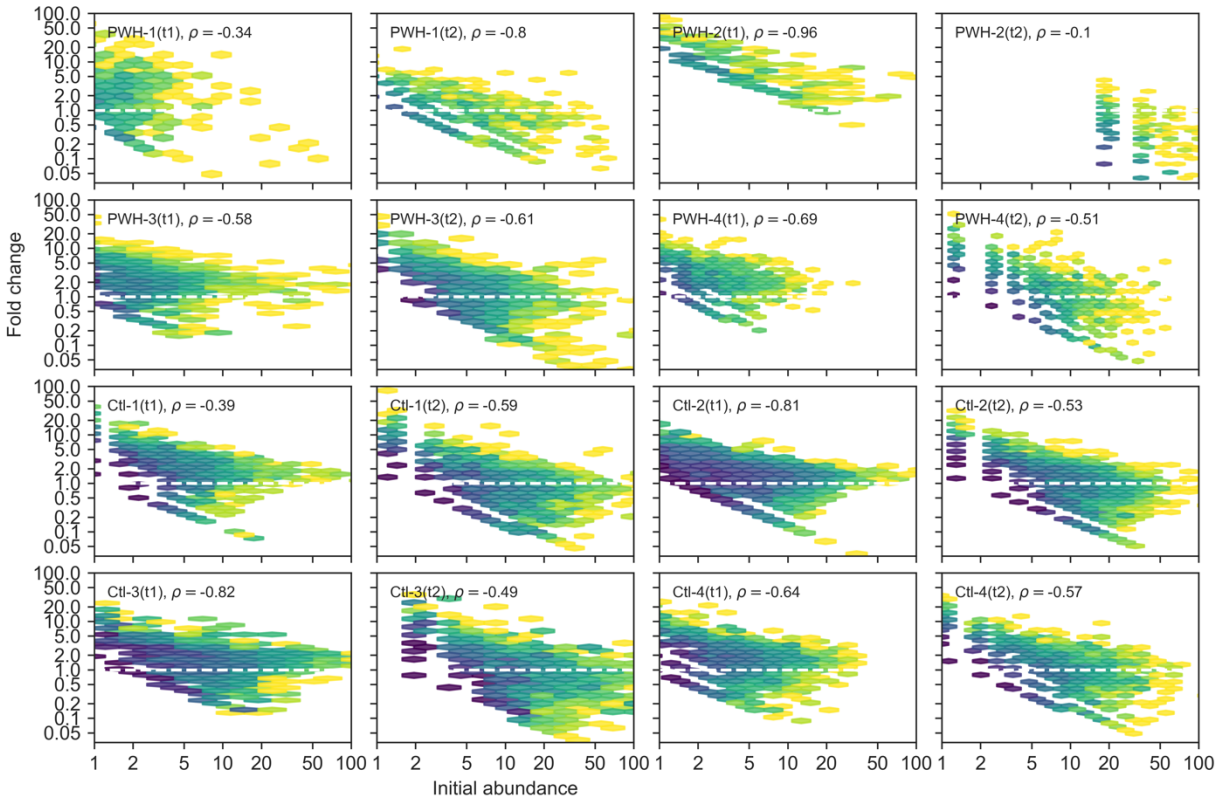
Supplementary Figure 7. Longitudinal ecological metrics for TCR β data from PWH and Ctl groups. No significant trends are detected via linear mixed effects model ($p>0.05$ on slope). Study months indicates time relative to t0. Individuals denoted by different colors, PWH in reds and oranges and HIV-seronegative people in blues and greens.



Supplementary Figure 8. Longitudinal persistent TCRβ compared with longitudinal HIV provirus rank abundance and ecology. A) Proportional rank abundance distributions of persistent TCRβ at all three time points. B) Diversity metrics from longitudinal PWH (reds) and Ctl cohorts (blues). Pearson correlation coefficients indicate no trends over time ($p > 0.05$). C) Proviral proportional rank abundance distributions of raw proviral data (not resampled) at t1 and t2 illustrate visually how clonality increases in 3/4 participants (compare triangles and squares) and in PWH 3 the change is unclear because t1 had only $n=10$ samples, all singletons.



Supplementary Figure 9. Longitudinal ecological metrics (A) and power law exponents (B) estimated from antigen specific TCR β data. Across 2 participants (PWH-1 and PWH-4, colors correspond across all panels) no significant trends are detected via pearson correlation coefficient ($p > 0.05$ on slope). Study months indicates time relative to t0.



Supplementary Figure 10. Correlation between initial mCD4 TCRβ clone size and fold-change in clone size across for all PWH and Ctl. For each participant time point (t1 and t2), the initial abundance was determined at the prior time point (t0 and t1, respectively). Then the fold change of clonotypes found at both time points is shown. All data are derived from resamples where each time point has the same $n=1e4$. Darker regions indicate higher density of points. Spearman correlation coefficients are annotated on each panel, indicating significant negative trends such that larger initial sizes predict contraction and smaller initial sizes predict expansion.