SI Methods

We applied Bayesian inference methods to estimate jointly the frequency of integrated proviruses and their reactivation probability, an approach enabled by the combined proviral sequencing and viral outgrowth assay data. We model the probability of measuring a certain number \( k_{\text{DNA}} \) of integrated proviral sequences as binomial,

\[
k_{\text{DNA}} \sim \text{Binomial}(W_{\text{DNA}}, p_{\text{DNA}}),
\]

\[
p_{\text{DNA}} = 1 - (1 - p)^{C_{\text{DNA}}} \approx 1 - e^{-C_{\text{DNA}} p}. \tag{S1}
\]

Here, \( W_{\text{DNA}} \) is the number of wells tested, \( C_{\text{DNA}} \) is the number of cells per well, and \( p \) is the frequency of CD4\(^+\) T cells with integrated provirus. Similarly, the probability of observing \( k_{\text{VOA}} \)-positive wells in the viral outgrowth assay is also binomial,

\[
k_{\text{RNA}} \sim \text{Binomial}(W_{\text{VOA}}, p_{\text{VOA}}), \quad p_{\text{VOA}} \approx 1 - e^{-C_{\text{VOA}} r}. \tag{S2}
\]

As above, \( W_{\text{VOA}} \) and \( C_{\text{VOA}} \) are the number of wells and number of cells tested per well in the viral outgrowth assay, respectively. Here, \( r \) is the probability that an integrated provirus reactivates and grows out in the VOA.

We assumed the following prior distributions:

\[
\log_{10}(p) \sim \text{Normal}(\mu, 2),
\]

\[
r \sim \text{Uniform}(0, 1), \tag{S3}
\]

where \( \mu = \max(k_{\text{DNA}}/(W_{\text{DNA}} \times C_{\text{DNA}}), 10^{-7}) \) are weakly informative (S4), intended simply to prevent pathological inferences. This approach yields conservative parameter estimates. For example, posterior distributions of reactivation probabilities for large clones are nonzero even if no reactivation events are observed. In models that include data from multiple visits, we also included an informative prior to take into account the expected decay of the latent reservoir in individuals on uninterrupted ART,

\[
\log_{10} \left( \frac{p_2}{p_1} \right) \sim \text{Normal}(\lambda t, \sigma). \tag{S4}
\]

Here, \( p_i \) and \( r_i \) are the frequency of integrated provirus and reactivation probability at visit \( i \), respectively. Their product is therefore equivalent to the number of infectious units per million cells (IUPM), divided by \( 10^6 \). We used \( \lambda = -0.007 \text{ mo}^{-1} \) as the IUPM decay rate and \( \sigma = 0.38 \) as the SD of its measurement, following previous estimates (3). The time between visits (in months) is \( t \). We implemented all models in Stan (52) using the PyStan interface.
Fig. S1. Maximum likelihood phylogenetic tree was constructed from viral env sequences from outgrowth culture supernatants as well as archived proviral DNA from all participants. Hypervariable (as defined in https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/) and other poorly aligned regions were excluded from the analysis. The tree was constructed using RAxML v. 8.0.22 (55) with a GTR+GAMMA substitution model, with 1,000 bootstrap replicates and midpoint rooted.
Figure S2. Maximum likelihood phylogenetic tree was constructed from viral env sequences from outgrowth culture supernatants for each participant. Viruses from time point 1 are red, viruses from time point 2 are orange, and bulk culture viruses are green. Asterisks indicate nodes with significant bootstrap values (bootstrap support ≥ 70%). Tables beneath each tree show concentration that inhibits response by 80% (IC<sub>80</sub>) titers for selected outgrowth culture viruses (red, IC<sub>80</sub> of 0–0.1 μg/mL; orange, IC<sub>80</sub> of 0.1–1.0 μg/mL; yellow, IC<sub>80</sub> of 1.0–10 μg/mL; green, IC<sub>80</sub> of 10–50 μg/mL). NT, not tested.
Table S1. Clinical characteristics of study subjects

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Age</th>
<th>Sex</th>
<th>Year of HIV diagnosis</th>
<th>CD4 nadir</th>
<th>Years since HIV diagnosis</th>
<th>Years on ART</th>
<th>ART regimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>B106</td>
<td>27</td>
<td>M</td>
<td>2008</td>
<td>390</td>
<td>7</td>
<td>7</td>
<td>TDF/FTC/RPV</td>
</tr>
<tr>
<td>B115</td>
<td>44</td>
<td>M</td>
<td>1993</td>
<td>200</td>
<td>22</td>
<td>22</td>
<td>DRV/r, ABC, 3TC</td>
</tr>
<tr>
<td>B155</td>
<td>59</td>
<td>M</td>
<td>1993</td>
<td>444</td>
<td>22</td>
<td>15</td>
<td>TDF/FTC/RPV</td>
</tr>
<tr>
<td>B199</td>
<td>49</td>
<td>M</td>
<td>2009</td>
<td>200</td>
<td>6</td>
<td>4</td>
<td>TDF/FTC, RAL</td>
</tr>
</tbody>
</table>

ABC, abacavir; DRV/r, darunavir/ritonavir; FTC, emtricitabine; ID, identification; M, male; RAL, raltegravir; RPV, rilpivirine; 3TC, lamivudine; TDF, tenofovir disoproxil fumarate.