HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs

Highlights
- HIV-1 sequences sampled from different reservoirs were compared to rebound viruses in 11 individuals
- Rebound viruses can originate from various cellular and anatomical compartments
- Cellular proliferation is an important driver of HIV persistence
- Cure strategies should take into account the lack of a prominent HIV rebound origin

Authors
Marie-Angélique De Scheerder, Bram Vrancken, Simon Dellicour, ..., Philippe Lemey, Sarah Palmer, Linos Vandekerckhove

Correspondence
marieangelique.descheerder@ugent.be (M.-A.D.S.), linos.vandekerckhove@ugent.be (L.V.)

In Brief
De Scheerder et al. conduct an in-depth investigation into the origins of HIV rebound. They show that viral rebound originates from multiple compartments and cell proliferation is a driver of viral persistence. Future HIV cure strategies will need to overcome the challenges associated with heterogeneous viral rebound.
HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs

Marie-Ange´ lique De Scheerder,1,* Bram Vrancken,2,15 Simon Dellicour,2,3,15 Timothy Schlub,4 Eunok Lee,5 Wei Shao,6 Sofie Rutsaert,1 Chris Verhofstede,7 Tessa Kerre,8 Thomas Malfait,9 Dimitri Hemelsoet,10 Marc Coppens,11 Annemieke Dhondt,12 Danny De Looze,13 Frank Vermassen,14 Philippe Lemey,7,16 Sarah Palmer,5,16 and Linos Vandekerckhove1,16,17,*

1HIV Cure Research Center, Department of General Internal Medicine, Ghent University Hospital, Ghent University, Ghent 9000, Belgium
2KU Leuven, Department of Microbiology and Immunology, Rega Institute, Laboratory of Evolutionary and Computational Virology, Herestraat 49, Leuven 3000 Belgium
3Spatial Epidemiology Laboratory (SpELL), Université Libre de Bruxelles, CP160/12 50, av. FD Roosevelt, 1050 Bruxelles, Belgium
4University of Sydney, Faculty of Medicine and Health, Sydney School of Public Health, Sydney 2000, NSW, Australia
5Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney 2145, NSW, Australia
6Frederick National Laboratory for Cancer Research (FNLCR), Frederick, MD, USA
7Aids Reference Laboratory, Ghent University Hospital, Ghent 9000, Belgium
8Department of Hematology, Ghent University Hospital, Ghent 9000, Belgium
9Department of Pulmonology, Ghent University Hospital, Ghent 9000, Belgium
10Department of Neurology, Ghent University Hospital, Ghent 9000, Belgium
11Department of Anesthesiology, Ghent University Hospital, Ghent 9000, Belgium
12Department of Nephrology, Ghent University Hospital, Ghent 9000, Belgium
13Department of Vascular Surgery, Ghent University Hospital, Ghent 9000, Belgium
14These authors contributed equally
15Senior author
16Lead Contact
*Correspondence: marieangelique.descheerder@ugent.be (M.-A.D.S.), linos.vandekerckhove@ugent.be (L.V.)
https://doi.org/10.1016/j.chom.2019.08.003

SUMMARY

Viral rebound upon stopping combined antiretroviral therapy poses a major barrier toward an HIV cure. Cellular and anatomical sources responsible for reinitiating viral replication remain a subject of ardent debate, despite extensive research efforts. To unravel the source of rebounding viruses, we conducted a large-scale HIV-STAR (HIV-1 sequencing before analytical treatment interruption to identify the anatomically relevant HIV reservoir) clinical trial. We collected samples from 11 participants and compared the genetic composition of (pro)viruses collected under treatment from different cellular and anatomical compartments with that of plasma viruses sampled during analytical treatment interruption. We found a remarkably heterogeneous source of viral rebound. In addition, irrespective of the compartment or cell subset, genetically identical viral expansions played a significant role in viral rebound. Our study suggests that although there does not seem to be a primary source for rebound HIV, cellular proliferation is an important driver of HIV persistence and should therefore be considered in future curative strategies.

INTRODUCTION

Although major advances have been made in identifying cells and anatomical compartments that contribute to the HIV latent reservoir, it is still unclear which of these reservoirs are responsible for viral rebound during treatment interruption in vivo (Barton et al., 2016; Chun et al., 2010). HIV is integrated into the genome of various immune cells, and considerable debate exists about which subset represents the actual latent reservoir, defined as the intact and replication-competent virus capable of viral rebound during treatment interruption (Eisele and Sili- ciano, 2012; Kandathil et al., 2016). The therapeutic implications of identifying such a specific reservoir are tremendous because targeting specific anatomical sites and/or cell subsets that harbor the large majority of replication-competent virus may increase the efficiency of HIV cure efforts in patients on long-term suppressive combined antiretroviral therapy (cART) (Leder- man et al., 2016; Margolis et al., 2016).

Although different CD4+ T cell subsets have been identified as major contributors to the HIV-1 reservoir (Chomont et al., 2009; Hiener et al., 2017; Lee et al., 2017b) and different surface markers (CXCR3, CD30, and CD32a) (Banga et al., 2018; Descours et al., 2017; Henrich et al., 2017) and immune exhaustion markers (PD-1, TIGIT, and LAG-3) (Khoury et al., 2017; McGary et al., 2017) have emerged as potential biomarkers for latently infected cells, previous studies have not converged on a clear pattern or predominant subsets responsible for HIV persistence.
In-Depth HIV-1 Reservoir Characterization Shows a Variability in Identical Sequence Expansions, Intact V1-V3 env Sequences, and Infection Rates between Subsets and Compartments

We first sought to identify whether substantial differences existed between cell subsets and anatomical compartments in the extent of identical sequence expansions, provirus intactness within the sequenced region, and infection frequency (Table 1; Figure 2).

The contribution of the various cell subsets and anatomical compartments to the persistence of HIV infection through cellular proliferation was quantified by the proportion of identical proviral env sequences within each of the different cell subsets. This proportion and its distribution between anatomical sites and the most exhaustive sampling (about >400 sequences per participant) of virus populations from different reservoirs and rebound viruses to date, allowing for a much-needed comprehensive investigation into the origins of rebound viruses.

RESULTS

In-Depth HIV Sampling before and during ATI

Twelve HIV-1 positive individuals were included in the study (STAR 1–12; Table S1, see STAR Methods for inclusion and exclusion criteria). STAR 1 was not subjected to the treatment interruption phase and therefore not included in further analyses. The study setup involved sampling of peripheral blood mononuclear cells (PBMCs), GALT, LN, plasma, and cerebrospinal fluid (CSF) on cART (time point 1, T1 in Figure 1). CD45+ cells from GALT; central (TCM), transitional (TTM), and effector (TEM) memory and naive (TN) CD4+ T cells from PBMCs; and memory CD4+ T cell subsets (TCM and TEM) were sorted from LN (STAR Methods; Figure S5 for gating strategy). In addition, plasma was sampled prior to treatment initiation (time point 0, T0), 8 to 15 days after ATI (time point 2, T2), at the first detectable viral load (time point 3, T3), and at rebound (time point 4, T4) (Figure 1). The median time to viral rebound was 21 days (range 15–36). To determine the source of viral rebound, we characterized proviral DNA derived from a broad range of cell subsets (T1) and plasma RNA (T0–T4), and we obtained V1–V3 env sequences using single-genome sequencing (Joseffson et al., 2012). In 5 out of 11 participants, we also managed to obtain RNA sequences from CSF (Table 1), and for two participants (STAR 4 and 10), we analyzed additional sequence data generated from bone marrow (BM) samples.

To ensure high-quality sequence data in subsequent statistical and phylogenetic analyses, we only included those sequences with a mean quality score of Q20 (99% accuracy) and excluded all defective sequences prior to and following the generation of the contigs (Barton et al., 2016; Ewing and Green, 1998; Joseffson et al., 2012; von Stockenstrom et al., 2015) (see STAR Methods, Table S3). For the phylogenetic analyses, recombinant lineages contributing to viral rebound were considered in separate downstream analyses (Martin et al., 2015). The quality control and filtering steps resulted in a total of 4,329 env sequences (an average of 393 sequences per participant, range 240–535) for further analyses.

In-Depth HIV Sampling before and during ATI

Twelve HIV-1 positive individuals were included in the study (STAR 1–12; Table S1, see STAR Methods for inclusion and exclusion criteria). STAR 1 was not subjected to the treatment interruption phase and therefore not included in further analyses. The study setup involved sampling of peripheral blood mononuclear cells (PBMCs), GALT, LN, plasma, and cerebrospinal fluid (CSF) on cART (time point 1, T1 in Figure 1). CD45+ cells from GALT; central (TCM), transitional (TTM), and effector (TEM) memory and naive (TN) CD4+ T cells from PBMCs; and memory CD4+ T cell subsets (TCM and TEM) were sorted from LN (STAR Methods; Figure S5 for gating strategy). In addition, plasma was sampled prior to treatment initiation (time point 0, T0), 8 to 15 days after ATI (time point 2, T2), at the first detectable viral load (time point 3, T3), and at rebound (time point 4, T4) (Figure 1). The median time to viral rebound was 21 days (range 15–36). To determine the source of viral rebound, we characterized proviral DNA derived from a broad range of cell subsets (T1) and plasma RNA (T0–T4), and we obtained V1–V3 env sequences using single-genome sequencing (Joseffson et al., 2012). In 5 out of 11 participants, we also managed to obtain RNA sequences from CSF (Table 1), and for two participants (STAR 4 and 10), we analyzed additional sequence data generated from bone marrow (BM) samples.

To ensure high-quality sequence data in subsequent statistical and phylogenetic analyses, we only included those sequences with a mean quality score of Q20 (99% accuracy) and excluded all defective sequences prior to and following the generation of the contigs (Barton et al., 2016; Ewing and Green, 1998; Joseffson et al., 2012; von Stockenstrom et al., 2015) (see STAR Methods, Table S3). For the phylogenetic analyses, recombinant lineages contributing to viral rebound were considered in separate downstream analyses (Martin et al., 2015). The quality control and filtering steps resulted in a total of 4,329 env sequences (an average of 393 sequences per participant, range 240–535) for further analyses.

In-Depth HIV-1 Reservoir Characterization Shows a Variability in Identical Sequence Expansions, Intact V1-V3 env Sequences, and Infection Rates between Subsets and Compartments

We first sought to identify whether substantial differences existed between cell subsets and anatomical compartments in the extent of identical sequence expansions, provirus intactness within the sequenced region, and infection frequency (Table 1; Figure 2).

The contribution of the various cell subsets and anatomical compartments to the persistence of HIV infection through cellular proliferation was quantified by the proportion of identical proviral env sequences within each of the different cell subsets. This proportion and its distribution between anatomical sites and
cell subsets were highly variable between participants (effect modification p < 0.001).

Across cell subsets, we found strong evidence for differences in the level of genetically identical HIV sequences (p < 0.001).

Table 1. In-Depth HIV-1 Reservoir Characterization

<table>
<thead>
<tr>
<th></th>
<th>STAR 2</th>
<th>STAR 3</th>
<th>STAR 4</th>
<th>STAR 5</th>
<th>STAR 6</th>
<th>STAR 7</th>
<th>STAR 8</th>
<th>STAR 9</th>
<th>STAR 10</th>
<th>STAR 11</th>
<th>STAR 12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TCM blood</strong></td>
<td>Total N</td>
<td>30</td>
<td>34</td>
<td>48</td>
<td>37</td>
<td>31</td>
<td>41</td>
<td>40</td>
<td>35</td>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.83</td>
<td>0.91</td>
<td>0.71</td>
<td>0.62</td>
<td>0.94</td>
<td>0.61</td>
<td>0.93</td>
<td>0.74</td>
<td>1.00</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.07</td>
<td>0.50</td>
<td>0.19</td>
<td>0.41</td>
<td>0.00</td>
<td>0.29</td>
<td>0.38</td>
<td>0.11</td>
<td>0.41</td>
<td>0.15</td>
<td>0.53</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>158</td>
<td>140</td>
<td>4658</td>
<td>970</td>
<td>397</td>
<td>219</td>
<td>190</td>
<td>2140</td>
<td>726</td>
<td>1157</td>
<td>125</td>
</tr>
<tr>
<td><strong>TEM blood</strong></td>
<td>Total N</td>
<td>35</td>
<td>26</td>
<td>43</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>27</td>
<td>40</td>
<td>41</td>
<td>38</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
<td>1.00</td>
<td>1.00</td>
<td>0.85</td>
<td>0.93</td>
<td>0.78</td>
<td>0.95</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.49</td>
<td>0.73</td>
<td>0.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.59</td>
<td>0.55</td>
<td>0.83</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>518</td>
<td>272</td>
<td>1468</td>
<td>NA</td>
<td>NA</td>
<td>146</td>
<td>4612</td>
<td>671</td>
<td>332</td>
<td>802</td>
<td>47</td>
</tr>
<tr>
<td><strong>TTM blood</strong></td>
<td>Total N</td>
<td>28</td>
<td>33</td>
<td>32</td>
<td>38</td>
<td>30</td>
<td>21</td>
<td>40</td>
<td>31</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.89</td>
<td>0.85</td>
<td>0.56</td>
<td>0.84</td>
<td>0.83</td>
<td>0.71</td>
<td>0.95</td>
<td>0.68</td>
<td>0.97</td>
<td>0.88</td>
<td>0.93</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.29</td>
<td>0.58</td>
<td>0.41</td>
<td>0.47</td>
<td>0.20</td>
<td>0.33</td>
<td>0.28</td>
<td>0.16</td>
<td>0.35</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>278</td>
<td>761</td>
<td>422</td>
<td>735</td>
<td>257</td>
<td>37</td>
<td>446</td>
<td>1209</td>
<td>173</td>
<td>220</td>
<td>43</td>
</tr>
<tr>
<td><strong>TN blood</strong></td>
<td>Total N</td>
<td>30</td>
<td>35</td>
<td>32</td>
<td>30</td>
<td>1</td>
<td>27</td>
<td>33</td>
<td>28</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.97</td>
<td>0.97</td>
<td>0.84</td>
<td>0.83</td>
<td>1.00</td>
<td>0.89</td>
<td>0.91</td>
<td>0.96</td>
<td>0.94</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.23</td>
<td>0.60</td>
<td>0.28</td>
<td>0.43</td>
<td>0.00</td>
<td>0.07</td>
<td>0.55</td>
<td>0.73</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>76</td>
<td>199</td>
<td>584</td>
<td>282</td>
<td>N/A</td>
<td>26</td>
<td>30</td>
<td>43</td>
<td>447</td>
<td>19</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>TCM LN/GALT</strong></td>
<td>Total N</td>
<td>38</td>
<td>33</td>
<td>31</td>
<td>32</td>
<td>36</td>
<td>37</td>
<td>32</td>
<td>N/A</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.95</td>
<td>0.94</td>
<td>0.71</td>
<td>0.78</td>
<td>0.97</td>
<td>0.95</td>
<td>0.84</td>
<td>N/A</td>
<td>0.94</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.16</td>
<td>0.64</td>
<td>0.00</td>
<td>0.25</td>
<td>0.17</td>
<td>0.05</td>
<td>0.13</td>
<td>N/A</td>
<td>0.38</td>
<td>0.19</td>
<td>0.67</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>292</td>
<td>190</td>
<td>911</td>
<td>555</td>
<td>200</td>
<td>92</td>
<td>154</td>
<td>N/A</td>
<td>467</td>
<td>110</td>
<td>29</td>
</tr>
<tr>
<td><strong>TEM LN/GALT</strong></td>
<td>Total N</td>
<td>33</td>
<td>28</td>
<td>23</td>
<td>35</td>
<td>38</td>
<td>14</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.76</td>
<td>0.96</td>
<td>0.87</td>
<td>0.89</td>
<td>0.92</td>
<td>0.71</td>
<td>0.89</td>
<td>0.95</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.45</td>
<td>0.61</td>
<td>0.52</td>
<td>0.69</td>
<td>0.24</td>
<td>0.00</td>
<td>0.68</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>504</td>
<td>728</td>
<td>665</td>
<td>786</td>
<td>334</td>
<td>95</td>
<td>297</td>
<td>2373</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>CD45 GALT</strong></td>
<td>Total N</td>
<td>26</td>
<td>29</td>
<td>29</td>
<td>36</td>
<td>39</td>
<td>36</td>
<td>13</td>
<td>30</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td>Prop intact</td>
<td>0.92</td>
<td>0.86</td>
<td>0.59</td>
<td>0.81</td>
<td>0.93</td>
<td>0.89</td>
<td>0.92</td>
<td>0.87</td>
<td>0.95</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>Prop identical</td>
<td>0.31</td>
<td>0.76</td>
<td>0.24</td>
<td>0.36</td>
<td>0.28</td>
<td>0.67</td>
<td>0.62</td>
<td>0.20</td>
<td>0.63</td>
<td>0.71</td>
<td>0.90</td>
</tr>
<tr>
<td>Infection frequency</td>
<td>121</td>
<td>168</td>
<td>447</td>
<td>151</td>
<td>52</td>
<td>393</td>
<td>76</td>
<td>756</td>
<td>196</td>
<td>3632</td>
<td>341</td>
</tr>
<tr>
<td><strong>Intact V1V3env RNA plasma</strong></td>
<td>Total N</td>
<td>163</td>
<td>102</td>
<td>72</td>
<td>108</td>
<td>128</td>
<td>125</td>
<td>125</td>
<td>104</td>
<td>101</td>
<td>133</td>
</tr>
<tr>
<td>T0</td>
<td>29</td>
<td>30</td>
<td>0</td>
<td>27</td>
<td>40</td>
<td>23</td>
<td>37</td>
<td>27</td>
<td>51</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>T1</td>
<td>12</td>
<td>8</td>
<td>1</td>
<td>24</td>
<td>11</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>T2</td>
<td>8</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>4</td>
<td>21</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>T3</td>
<td>25</td>
<td>24</td>
<td>28</td>
<td>28</td>
<td>41</td>
<td>29</td>
<td>29</td>
<td>40</td>
<td>28</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>T4</td>
<td>89</td>
<td>35</td>
<td>31</td>
<td>29</td>
<td>33</td>
<td>32</td>
<td>38</td>
<td>30</td>
<td>33</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td><strong>Intact V1V3env RNA CSF</strong></td>
<td>Total N</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>N/A</td>
<td>1</td>
</tr>
<tr>
<td>T1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>N/A</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

For each participant and sampling site, at time point 1 (compartment and cell subset) the total number of sequences (N), the number of sequences that were intact in the sequenced region (from V1 to V3) and the number of sequences that were identical within the subset is given as a proportion of the total N. The estimated infection frequency is based on the absolute cell N input and the N of sequences obtained. Infection frequency was presented as positive cells per 10⁶. We also provide the N of intact RNA sequences from plasma and CSF at the different sampling points.

Overall within the sample across participants and cell subsets TEM had the highest proportion of identical sequences (p < 0.001, OR = 2.47 95% CI 1.78–3.43); TEM from blood contained more genetically identical V1–V3 env sequences than TEM.
sequences from LN (p = 0.02, OR = 1.63 95% CI 1.08–2.46). At the level of anatomical compartments, we found evidence for a higher average proportion of identical V1-V3 \textit{env} sequences in GALT than the LN and blood (p < 0.001, OR = 1.94 95% CI 1.51–2.50). However, this could be explained by including only CD45+ immune cells in GALT rather than differentiating between TCM and TEM.

As rebound viruses can only emerge from intact proviruses, we also scrutinized the proportion of intact V1-V3 \textit{env} sequences across cell subsets and compartments. We consider the relative proportion of intact viruses in V1-V3 \textit{env} to reflect relative probabilities to find intact viruses in the subsets. The proportion of genetically intact proviruses (within the V1-V3 \textit{env} region) varied across the cell subsets (p < 0.001) and across participants (p = 0.04). Within the participants, TN from blood had on average the highest proportion of intact sequences (p = 0.01, OR = 1.81 95% CI 1.15–2.87) followed by TEM from blood (p = 0.02 when comparing to the other subsets excluding TN, OR = 1.70 95% CI 1.08–2.66). There was no evidence for significant variation across the anatomical compartments (p = 0.17).

Estimates of HIV infection frequency of each cell subset were also found to heavily depend on the individual participant (effect modification with cell subsets p < 0.001). On average, TEM from blood or LN and TCM from blood had the highest rate of infection (OR = 2.67 95% CI 2.45–2.92 for TEM blood or LN and TCM from blood combined against all others, p < 0.001) while TN had the lowest.

Overall, these results paint a complex reservoir picture with broad seeding and important inter-participant variability, with TEM frequently harboring high levels of genetically identical and intact proviruses.
Interestingly, in 6 out of 11 participants, we found CD45RA+/CD45RO− cells derived from blood to have intact V1-V3 env sequences identical to viral sequences found in plasma after viral rebound. This was observed both as part of a large expansion of genetically identical sequences shared over different cell subsets (Figures 3C and S2B; STAR 10) or as a unique sequence genetically identical to a rebound cluster (Figures 3A and 3B). Although these markers were primarily used to sort naı̈ve T cells, we cannot exclude a small fraction of stem memory T Cells (TSCM), which were previously reported to be enriched in HIV (Buzon et al., 2014). In 2 participants (STAR 4 and 10), we further analyzed additional sequences derived from CD4+ T cells isolated from BM (Figure S1C). These sequences are dispersed across the trees and do not show any compartmental structure, nor is their clustering indicative of a particular role in viral rebound. In addition, we analyzed CSF from 5 out of 11 participants; however, the amount of sequences that we obtained was insufficient to use them for downstream analysis (Figure S1D).

**Cellular Proliferation Is a Driver of Viral Rebound**

To investigate the importance of cellular proliferation in the rebound dynamics, we subsequently focused on proviruses from cell subsets at T1 that were genetically identical to a plasma virus after ATI (T2-T3-T4). To illustrate this, we highlighted the identical sequence expansions that match rebound sequences from T2-T3 or T4 on the phylogenetic trees (Figures 3A–3C, red and blue dashed lines).

We provide a table in the Supplemental Information (Table S4), which represents the absolute number of sequences that are 100% identical to viral rebound for all participants. In addition, we show the number of sequences with a >99% match to take into account viral and assay error rates. Overall, these data revealed that cell subsets with large expansions of identical sequences (presumably clonally expanded) contain a higher proportion of sequences that were identical to rebound plasma virus (p = 0.003); this strongly suggests a role of cellular proliferation as a driver for viral rebound. Although we could not consistently identify a specific cellular or anatomical compartment responsible for viral rebound based on phylogenetic relationships and we observe a high inter-participant variability, these data confirm that effector memory CD4+ T cells (TEM) contribute to the latent reservoir, as this cell subset showed the highest proportion of identical sequence expansions (as previously discussed and illustrated in Table 1).

**Figure 3. Heterogeneous Cellular and Anatomical Reservoir Contributions to HIV Rebound with Cellular Proliferation as a Potential Driver**

Within-host ancestor-descendant relationships between the viruses from different cell types for 3 participants and a radar plot representing the variability in viral rebound source. (A–C) Maximum likelihood phylogenetic trees from three selected participants representing the sequences from T1 cell subsets from blood (TCM, TEM, TTM, and TN), LN (TCM and TEM), and GALT (CD45+ cells) before ATI. Plasma viruses from time points T2, T3, and T4 are grouped as plasma after STOP cART. The colored strip represents sampling origin for each sequence as indicated by the legend. The trees are drawn to scale and the gray circles represent the branch length from the root expressed as the number of substitutions per site. The scale values are given in the inset (light gray numbers). The heterogeneity in potential reservoir contribution is indicated by the color mixing in the strip. Identical cellular DNA V1–V3 sequence expansions that are identical to plasma RNA sequences after cART STOP were highlighted in the trees by the bold dashed lines, with the expansions colored in red and blue alternatingly. Trees from all participants are provided in Figure S1. (D) The radar plot representing, for each of the 3 participants, the estimated number of times that a rebound virus lineage originates from the respective cell subsets (depicted as the numbers from 0 to 35). The legend indicates the color used for each participant. A radar plot across all participants is available in Figure S4.
anatomical compartments and different cell subsets at the individual sequence level (Table S4). When considering 100% matches, we find no evidence for a difference in rebound contributions across anatomical sites at the individual sequence level (p = 0.12). However, when viral and assay error rates > 99% were taken into account, we did find some evidence for a difference (p = 0.028). Here, we observe the blood, gut, and then LN to contribute to rebound, respectively, from low to high, but individual comparisons between these were not significant (p > 0.05). At the individual sequence level, we find evidence that rebound contributions are different across cell subsets both for 100% and for >99% matches (p < 0.001). Although, on average, we observe the order from lowest to highest as CD45 < TCM < TN < TTM < TEM for 100% matches, this was highly variable across participants (effect modification p < 0.001) and changed if >99% matches were considered instead. This result shows that all cell subsets, including TEM, play an important role in viral rebound. We conducted a second analysis where identical sequences were reduced to a single representative sequence to avoid counting the same expansion multiple times (Figure 4; Table S4). When taking into account sequence expansions so that identical sequences are only counted once, we no longer find evidence for a difference in rebound contribution across anatomical sites (p = 0.34) or cell subsets (p = 0.28), further indicating that all anatomical compartments and cell subsets are potential sources of viral rebound and that cellular proliferation is an important driver for these rebound events (Figures 3 and 4; Table S4).

Evidence for Kinetic Variability and Stochastic Reactivation of Rebound Viruses after Treatment Interruption

Finally, we investigate in more detail the plasma RNA sequences collected at the different time points T0 > T4). In Figure 5, we highlight the plasma sequences at the various time points of sample collection in the phylogenetic trees for 3 participants (trees from all participants are provided in Figure S1B).

First, we assessed whether residual viremia during cART can predict the source of viral rebound. In the 6 participants for whom we were able to sequence the plasma virus on cART at T1, we observed two distinct patterns. In most of these individuals, identical viral sequences dominate in the plasma at T1 (Figure 5B, green dashed lines), but in some individuals, we observe multiple unique variants (Figure 5A, red dashed lines). Remarkably when clones dominate the residual viremia, they were more likely to be identical to virus obtained in the plasma at later time points, suggesting that these viruses represent a part of the infectious reservoir and are able to fuel viral rebound after ATI (Figure 5B, green dashed lines; Figure S2A STAR 7; and Figure S2C STAR 7 right circle). Similarly, we found plasma-derived clonal sequences before start cART at T0 that matched intracellular HIV-1 DNA sequences at T1 and rebound viruses at later time points after ATI (Figure 5C, green dashed lines; Figure S2A STAR 12), again emphasizing the role of proliferation in the persistence of the viral reservoir and their contribution to viral rebound.

Second, we wondered whether we could observe evolution in the viral sequences collected at different time points after ATI. We found that plasma viruses collected after ATI (T2, T3, and T4) were usually intermixed (Figure 5C, red dashed lines). Confirming this mixed phylogenetic clustering, we did not find any pronounced phylogenetic association by time point of sampling (Table S2). The phylogenetic patterns indicate multiple independent rebound viruses or rebound lineages of highly variable size (Figure S4). Even when the rebound virus was largely represented by a single lineage there remained an extensive intermixing of sequences from the other compartments, which is in line with a heterogeneous reservoir origin (Figure 5B, red dashed lines; Figure S2A STAR 3 left circle). The nucleotide diversity profiles also show that the rebound virus diversity is generally comparable to or only slightly lower than the genetic diversity of the HIV-1 DNA sequences from the cell subsets and anatomical compartments, which is expected when the virus from multiple cell subsets and anatomical sites contribute to the plasma virus after ATI (Figures 5D and S3).

However, in some participants, we were able to detect distinct patterns of rebound kinetics. In STAR 3, we observe a clear outgrowth of plasma virus over time (Figure 5B, blue dashed lines; Figure S2A STAR 3 right circle). This is confirmed by the nucleotide diversity analysis, wherein some participants a decrease in nucleotide diversity between T3 and T4 is observed, suggesting enrichment of a particular viral strain over time (Figures 5D and S3). In some other participants, we observed that plasma RNA sequences at T4 are genetically identical to the HIV-1 DNA sequences sampled at T1, but no genetic match was found between the HIV-1 DNA sequences and the previous plasma-derived RNA sequences from time points T2 and T3 (Figure 5C, blue dashed lines; Figure S2B STAR 10), indicating these reservoirs reactivated at a later time point during the ATI. These individual findings suggest stochastic cellular reactivation and kinetic variability of viral rebound after treatment interruption. Plotting intact plasma-derived sequences over time (T0–T4) does not suggest any selection or enrichment of replication-competent virus across patients (Figure S4C).

DISCUSSION

Overall, the most important finding from this extensive phylogenetic and statistical analyses based on the largest and most in-depth sequencing data of the HIV reservoir is that the rebound virus can originate from several cellular and anatomical compartments after treatment interruption. The substantial inter-participant variability further supports that there is no prominent source of rebound viruses. We observed a clear link between the rebound viruses and their source in some participants, whereas in others, there were multiple potential contributors or the source was unclear. This explains why studies of only a few participants have pointed at various sources of rebound viruses, often in an inconsistent manner. The sample size in our study, in terms of the number of participants but importantly also in terms of the number of independent rebound lineages detected in each participant, offers strong support against HIV rebound seeded by a dominant reservoir.

In line with a broad seeding of the reservoir and the lack of clear structure in the HIV rebound dynamics, our results support a substantial degree of mixing of viruses across anatomical compartments. Moreover, our study revealed identical env V1-V3 sequences in different cellular and anatomical compartments,
indicating dynamic interchanges between compartments and little evidence for viral evolution or clustering by site. The observed expansions of genetically identical HIV sequences during cART are the result of homeostatic and/or antigen-driven cellular proliferation, rather than resulting from ongoing viral replication (Banga et al., 2018; Kearney et al., 2017; Rosenbloom et al., 2017). In participants with groups of identical proviral env V1-V3 sequences, our data indicate that these expansions play an important role in viral rebound, confirming that cellular proliferation is a crucial driver of viral persistence, irrespective of the compartment or cell subset (Josefsson et al., 2013; von Stockenstrom et al., 2015; Wang et al., 2018). In this regard, cellular proliferation may further increase TEM contribution to rebound. Interestingly these data also showed that unique cellular sources...
such as CD45RO-/RA+ T cells, consisting of a majority (> 90%) naïve T cells (cf. STAR Methods), although we cannot exclude a small fraction of TSCM, which were previously reported to be enriched in HIV (Buzon et al., 2014) and unique anatomical sources such as GALT can contribute to viral rebound. We did not find clear evidence that the presence of viral production during cART can predict the dynamics of viral rebound; however, it can reveal the presence of dominant clones responsible for viral rebound.

In our unique study setup, we assess replication competence by comparing proviral sequences on cART to the rebound sequences in plasma after ATI. Therefore we focused on the highly variable V1-V3 env region (~0.8 kb) (Kearney et al., 2009; von Stockenstrom et al., 2015), and although we can therefore only account for identical and intactness of the sequenced region, modeling and previous studies have frequently used env sequences as a surrogate and showed that multiple identical env sequences provide a strong support for clonal expansion (Hosmane et al., 2017; Laskey et al., 2016). Several studies have therefore used the envelope region to identify the source of the rebound viremia (Barton et al., 2016; Lu et al., 2018).

More recently, new strategies studying near full-length proviral genome sequences are being presented and allow for even more in-depth analysis of the HIV reservoir (Bruner et al., 2016; Hiener et al., 2017; Lee et al., 2017a); however, any comparison with rebound plasma-derived HIV RNA in these studies is based on a particular HIV subgenomic region due to the fact that full-length HIV RNA assays are not currently available and therefore do not contribute to a better understanding of rebound dynamics. One of the major problems with full-length sequencing at this stage is the relative paucity of full-length HIV sequences generated, therefore requiring a much higher cellular input. This major challenge is probably the reason why no studies on tissue reservoirs employing near full-length sequencing have been published so far. Although we recognize that implementing such a technique would in theory increase resolution and provide more direct proof of intactness and clonality of the available sequences, the current state-of-the-art cannot strengthen the answer to
the central research question (Wang and Palmer, 2018), and we are confident that the strategy we choose is a valuable surrogate for this extensive analysis of the viral reservoir and above all has little impact on our overall conclusion of heterogeneous viral rebound. However, we acknowledge that sequencing a subgenomic region overestimates the number of intact proviruses (Lee et al., 2017b). Furthermore, without integration site analysis, we are unable to determine if identical V1–V3 env sequences are derived from clonal expansions of the same cell (Cohn et al., 2015).

Our findings confirm the results of others (Cohen et al., 2018; Salantes et al., 2018; Winckelmann et al., 2017) that only a small fraction of plasma RNA env sequences sampled at ATI (T2-T3-T4) were genetically identical to viral sequences from the cells and anatomical sites sampled at T1 (approx. 20%, with a large range between participants). This suggests that although the origin of the rebound virus is very heterogeneous and diverse, only a small subset of this total pool of infected cells contributes to viremia (Kearney et al., 2016). Another potential explanation for mismatches between proviral and rebound sequences could be recombination, as described by Cohen et al. (Cohen et al., 2018). Determining the amount of viral recombination that contributes to the rebound virus within the env V1-V3 fragment is challenging to assess. Therefore, to correct for errors made during viral replication and the error rate of the assay, we determined the number of cell-derived sequences with greater than 99% identity to rebound viral sequences. In doing so, we found a significant increase in the number of genetically matching sequences (approximately 70%). However, this analysis did not change the conclusion that viral rebound is heterogeneous and cannot be attributed to one specific reservoir. By doing the >99% analysis, we furthermore took into consideration the potential error rate (calculated at 0.011% [Palmer et al., 2005]) and the potential bias associated with post-rebound diversification. However, based on the short post-rebound period and previous within-host evolutionary rate estimates this diversification can be considered as negligible. Indeed, if we consider evolutionary rate estimates for a similar env region obtained for 9 patients (Bielejec et al., 2016; Shankarappa et al., 1999), the highest average rate was estimated to be 0.001 subst/month (or 0.012 subst/year). So, for the longest post-rebound period of 36 days, we expect, on average, 0.0012 subst/site. For an alignment length of about 750 bp, this leads to an expectation of 0.9 substitutions. So, any sequence sampled 36 days post-rebound is expected to have accumulated only about one substitution since rebound. In other words, almost all diversity we observed in the post-rebound sequences has already been established. Depending on the participant and the cellular subsets, we see a mostly large amount of rebound lineages. Since immune pressure changes to the virus seem rather unlikely (cf. the analysis done above), we suggest that the main reason is the unique in-depth sequencing in this study providing several anatomical and cellular sources of viral rebound.

Although primer mismatch when sequencing a single and variable subgenomic region can occur, the env primers we used in our assay are HIV-1 subtype B specific and were chosen because they bind to the most conserved regions of the V1-V3 sequences in the Los Alamos HIV database and therefore are not selective for specific viral populations within the HIV-1 subtype B (Kearney et al., 2009).

While sampling bias can affect genetic analyses and one rebound event over time cannot capture the entire complexity of viral rebound, the substantial number of participants investigated and the depth at which they were sampled ensures that this will have little impact on our overall conclusion. As already suggested by others (Cohen et al., 2018; Wang et al., 2018), our data support stochastic reactivation, in line with our overall hypothesis of heterogenic contributions to rebound viruses.

In conclusion, our comprehensive study clearly demonstrates the complexity of latency mechanisms and the challenges this brings about for strategies aimed at purging the reservoir. Our data show that genetically identical viral expansions play a significant role in viral rebound. Focusing on mechanisms that drive antigenic and homeostatic proliferation of immune cells (Chomont et al., 2009; Hosmne et al., 2017) will be crucial to achieve progress toward an HIV cure. Although recent cure interventions (Mendoza et al., 2018) are promising, they also confirm that the response to treatment is heterogeneous among the trial participants and that an individual approach and a better insight into the mechanisms of viral persistence together with in-depth mapping of these latent reservoirs will be necessary to overcome the challenge of random reactivation.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Inclusion-and Exclusion Criteria
  - Clinical Study Design And Sample Collection
  - Sample Processing
  - Single-Genome/Proviral Sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Phylogenetic Analysis
  - Statistical Analysis
- DATA AND CODE AVAILABILITY
  - GenBank Accession Numbers
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2019.08.003.

ACKNOWLEDGMENTS

We would like to acknowledge the participants of the HIV STAR cohort; The Ghent University Hospital and all the collaborators for providing us the support and resources necessary to organize the in-depth sampling; the MDs involved at the Aids Reference Center in Ghent for recruiting of participants; and all the MDs involved in the sampling procedures, Prof. T. Kerre, Prof. D. De Loosze, Dr. Malait, Dr. D. Hemelsoet, Prof. Coppens, and all the supporting nurses and facilities, especially E. Van Caelenbergh, A. Masset, E. Caluwé, and S. Vanherreweghe.
M.-A.D.S. was granted a FWO travel grant and a Sofina Gustave-Boel fellowship. L.V. is supported by the Research Foundation Flanders (FWO 1.0.020.09.N.00). S.P. is supported by the Delaney AIDS Research Enterprise (DARE) to Find a Cure (U19AI096109 and 1UM1AI126611-01) and the Australian National Health and Medical Research Council (APP1061681 and APP1149900). P.L. is supported by the Special Research Fund, KU Leuven (“Bijzonder Onderzoeksfonds”, KU Leuven, OT/14/115), the Research Foundation Flanders FWO (GO66215N) and the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 725422-ReservoirDOC5). B.V. is supported by a postdoctoral fellowship from the Fund for Scientific Research (FWO) Flanders (Fonds voor Wetenschappelijk Onderzoek, Flanders, Belgium). S.D. is supported by the Fonds National de la Recherche Scientifique (FNRS, Belgium) and was previously supported by the Fund for Scientific Research (FWO) Flanders (Fonds voor Wetenschappelijk Onderzoek, Flanders, Belgium). S.R. received a strategic basic research fund of the Research Foundation Flanders (FWO 1S32916N). This work was supported by an MSD investigator grant (grant number: IS5 52777).

**AUTHOR CONTRIBUTIONS**


**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: February 22, 2019
Revised: May 16, 2019
Accepted: July 31, 2019
Published: August 27, 2019

**REFERENCES**


### STAR METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 neg selection kit for IMag: human CD4 T cell</td>
<td>BD Biosciences</td>
<td>Cat#557939; RRID: AB_2802162</td>
</tr>
<tr>
<td>Anti-human CD3 BB</td>
<td>BD Biosciences</td>
<td>Cat#564465; RRID: AB_2744386</td>
</tr>
<tr>
<td>Anti-human CD8 PeCy7</td>
<td>BD Biosciences</td>
<td>Cat#557746; RRID: AB_396852</td>
</tr>
<tr>
<td>Anti-human CD45 PerCPCy5.5</td>
<td>BD Biosciences</td>
<td>Cat#564105; RRID: AB_2744405</td>
</tr>
<tr>
<td>Anti-human CD45 RO PE</td>
<td>BD Biosciences</td>
<td>Cat#555493; RRID: AB_395884</td>
</tr>
<tr>
<td>Anti-human CD45 RO PerCPCy5.5</td>
<td>BD Biosciences</td>
<td>Cat#560607; RRID: AB_1277500</td>
</tr>
<tr>
<td>Anti-human CD45 RA APC</td>
<td>BD Biosciences</td>
<td>Cat#550855; RRID: AB_398468</td>
</tr>
<tr>
<td>Anti-human CD27 APC</td>
<td>BD Biosciences</td>
<td>Cat#561400; RRID: AB_10645790</td>
</tr>
<tr>
<td>Anti-human CD197 (CCR7) PE</td>
<td>BD Biosciences</td>
<td>Cat#560765; RRID: AB_2033949</td>
</tr>
<tr>
<td>Anti-human CD34 GP 105-120</td>
<td>BD Biosciences</td>
<td>Cat#560940; RRID: AB_10569308</td>
</tr>
<tr>
<td>Lineage Cocktail 1 (lin1)</td>
<td>BD Biosciences</td>
<td>Cat#340546; RRID: AB_400053</td>
</tr>
<tr>
<td>Anti-human CD206</td>
<td>BD Biosciences</td>
<td>Cat#555954; RRID: AB_396250</td>
</tr>
<tr>
<td>Viability stain 780</td>
<td>BD Biosciences</td>
<td>Cat#565388</td>
</tr>
<tr>
<td>FCR blocking reagent</td>
<td>Miltenyi Biotec</td>
<td>Cat#130-059-901</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leukosept tubes</td>
<td>Fisher Scientific</td>
<td>Cat#10349081</td>
</tr>
<tr>
<td>Lymphoprep</td>
<td>Eittech</td>
<td>AXI-1114547</td>
</tr>
<tr>
<td>RPMI</td>
<td>Thermo Fisher</td>
<td>Cat#21875034</td>
</tr>
<tr>
<td>EDTA</td>
<td>Thermo Fisher</td>
<td>Cat#15575020</td>
</tr>
<tr>
<td>PBS PH 7,2</td>
<td>Thermo Fisher</td>
<td>Cat#20012019</td>
</tr>
<tr>
<td>L-glutamine</td>
<td>Thermo Fisher</td>
<td>Cat#25030024</td>
</tr>
<tr>
<td>Penicillin/Streptomycine</td>
<td>Thermo Fisher</td>
<td>Cat#15140122</td>
</tr>
<tr>
<td>HBSS</td>
<td>Thermo Fisher</td>
<td>Cat#14175053</td>
</tr>
<tr>
<td>DTT</td>
<td>Thermo Fisher</td>
<td>Cat#20290</td>
</tr>
<tr>
<td>Costar 6 well plates</td>
<td>Sigma Aldrich Costar</td>
<td>SKU CLS3471-24EA</td>
</tr>
<tr>
<td>Lysis buffer (10mM TRisHCl, 0.5%NP-40, 0.5% Tween-20)</td>
<td>Made in the lab following protocol</td>
<td>N/A</td>
</tr>
<tr>
<td>Proteinase K 20mg/ml</td>
<td>Ambion /Lifetechnologies</td>
<td>Cat#AM2546</td>
</tr>
<tr>
<td>PCR nucleotide mix 10mM 1ml</td>
<td>Promega</td>
<td>Cat#C1145</td>
</tr>
<tr>
<td>8M guanidinium thiocyanate BioUltr</td>
<td>Sigma Aldrich</td>
<td>SKU 50980-50ml; CAS 593-84-0</td>
</tr>
<tr>
<td>Glycogen (from mussels) 20mg</td>
<td>Roche</td>
<td>SKU 10901393001</td>
</tr>
<tr>
<td>Tris buffered saline (tablets)</td>
<td>Sigma Aldrich</td>
<td>T5030-100Tab</td>
</tr>
<tr>
<td>Isopropanol (2-propanolol) for molecular biology</td>
<td>Sigma Aldrich</td>
<td>I9516; CAS 67-63-0</td>
</tr>
<tr>
<td>Ultracentrifuge tubes (optiseal, polypropylene 8.9, 16x60mm, qty56)</td>
<td>Beckman Coulter</td>
<td>361623</td>
</tr>
<tr>
<td>Ethanol 100%(ethyl alcohol, Pure 200 proof, for molecular biology)</td>
<td>Sigma Aldrich</td>
<td>E7023-1L; CAS 64-17-5</td>
</tr>
<tr>
<td>Ultrapure 1M TrisHCL PH 8</td>
<td>Thermo Fisher Scientific</td>
<td>Cat#15568-025</td>
</tr>
<tr>
<td>Corning Cell strainers 40µm</td>
<td>Sigma Aldrich</td>
<td>CLS 431750-50AE</td>
</tr>
<tr>
<td>Corning Cell strainers 70µm</td>
<td>Sigma Aldrich</td>
<td>CLS 431751-50AE</td>
</tr>
<tr>
<td>Fetal Calf Serum (FCS)</td>
<td>Hyclone RVM</td>
<td>35888</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE SOURCE IDENTIFIER</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
</tr>
<tr>
<td>Superscript III 1st strand Invitrogen</td>
<td>Cat#18080051</td>
</tr>
<tr>
<td>Platinum Taq PCR polymerase Invitrogen</td>
<td>Cat#10966083</td>
</tr>
<tr>
<td>E-Gel 96 well 1% agarose Thermo Fisher Scientific</td>
<td>Cat#G700801</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
</tr>
<tr>
<td>E20 F 5’GGGCCACACATGCCTGTGTACCCACAG 3’</td>
<td>Josefsson et al., 2012; von Stockenstrom et al., 2015</td>
</tr>
<tr>
<td>E115 5’ GAAAAATTCCCCTCCACAATTAA 3’</td>
<td>Josefsson et al., 2012; von Stockenstrom et al., 2015</td>
</tr>
<tr>
<td>E30 F 5’GTGTACCCACAGACCCCAGCCCACAAG 3’</td>
<td>Josefsson et al., 2012; von Stockenstrom et al., 2015</td>
</tr>
<tr>
<td>E125 R 5’CAATTTCTGGTCCCCCTCTGGAGG 3’</td>
<td>Josefsson et al., 2012; von Stockenstrom et al., 2015</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH642355</td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH642607</td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH642608</td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH643062</td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH643063</td>
<td></td>
</tr>
<tr>
<td>RNA and DNA sequencing Genbank  MH643573</td>
<td></td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
</tr>
<tr>
<td>Subtype analyses Smartgene, Lausanne, Switzerland</td>
<td><a href="https://www.smartgene.com/">https://www.smartgene.com/</a></td>
</tr>
<tr>
<td>Viral Tropism Web PSSM, Mullins lab, University of</td>
<td><a href="https://indra.mullins.microbiol.washington.edu/webpssm/">https://indra.mullins.microbiol.washington.edu/webpssm/</a></td>
</tr>
<tr>
<td>Flow Jo software BD Biosciences</td>
<td><a href="https://www.flowjo.com/">https://www.flowjo.com/</a></td>
</tr>
<tr>
<td>Perl script language Perl</td>
<td><a href="https://www.perl.org/">https://www.perl.org/</a></td>
</tr>
<tr>
<td>MUSCLE Edgar, 2004</td>
<td><a href="http://www.megasoftware.net/">http://www.megasoftware.net/</a></td>
</tr>
<tr>
<td>Aliview Larsson, 2014</td>
<td><a href="http://www.ormbunkar.se/aliview/downloads">www.ormbunkar.se/aliview/downloads</a></td>
</tr>
<tr>
<td>RDP4 Martin et al., 2015</td>
<td><a href="http://web.cbio.uct.ac.za/~darren/rdp.html">http://web.cbio.uct.ac.za/~darren/rdp.html</a></td>
</tr>
<tr>
<td>Translator X Abascal et al., 2010</td>
<td><a href="http://www.translatorx.co.uk/">http://www.translatorx.co.uk/</a></td>
</tr>
<tr>
<td>BaTS Parker et al., 2008</td>
<td><a href="http://evolve.zoo.ox.ac.uk/Evolve/BaTS.html">http://evolve.zoo.ox.ac.uk/Evolve/BaTS.html</a></td>
</tr>
<tr>
<td>Mr Bayes v3.6.2 Ronquist et al., 2012</td>
<td><a href="http://nibisweden.github.io/MrBayes/download.html">http://nibisweden.github.io/MrBayes/download.html</a></td>
</tr>
<tr>
<td>Tracer BEAST</td>
<td><a href="https://github.com/beast-dev/tracer/releases/tag/v1.7.1">https://github.com/beast-dev/tracer/releases/tag/v1.7.1</a></td>
</tr>
<tr>
<td>Biopython Cock et al., 2009</td>
<td><a href="https://biopython.org/wiki/Download">https://biopython.org/wiki/Download</a></td>
</tr>
<tr>
<td>Elim Dupes Los Alamos HIV Database</td>
<td><a href="http://hiv.lanl.gov">http://hiv.lanl.gov</a></td>
</tr>
<tr>
<td>R: A language and environment for statistical computing,</td>
<td>R script</td>
</tr>
<tr>
<td>General Liner hypothesis function (glht) R package multcomp</td>
<td><a href="https://cran.r-project.org/web/packages/multcomp/index.html">https://cran.r-project.org/web/packages/multcomp/index.html</a></td>
</tr>
<tr>
<td>FigTree BEAST</td>
<td><a href="http://beast.community/figtree">http://beast.community/figtree</a></td>
</tr>
<tr>
<td>iTol v3 Letunic and Bork, 2016</td>
<td><a href="https://itol.embl.de/">https://itol.embl.de/</a></td>
</tr>
<tr>
<td>NETWORK v5.0.0.3 Fluxus</td>
<td><a href="http://www.fluxus-engineering.com">http://www.fluxus-engineering.com</a></td>
</tr>
<tr>
<td>SPADS 1.0 Dellicour and Mardulyn, 2014</td>
<td><a href="http://ebe.uib.ac.be/ebe/SPADS.html">http://ebe.uib.ac.be/ebe/SPADS.html</a></td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
</tr>
<tr>
<td>Sanger sequencing AGRF Australian Genome Research Facility, Sydney, Australia</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Linos Vandekerckhove (linos.vandekerckhove@ugent.be). This study did not generate new unique reagents.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

The study protocol, experimental design, and recruitment strategy were approved by the Ethics Committee of the University Hospital of Ghent (Belgian registration number: B670201525474). Participants were recruited from the AIDS reference clinic at the Ghent University Hospital based on strict inclusion criteria and after intensive counseling (NCT02641756). Written informed consent was obtained from all participants. Twelve HIV-1 infected, long-term treated participants under chronic ART were included. The median age of the participants was 40y (range 32-56y). Two participants initiated cART during early infection (<6 months, STAR 3 and 5), while the remaining participants received ART during chronic infection (range 1-9 year after seroconversion). One participant was excluded from the second part of the study because of pre-existing resistance to ART, an exclusion criterion for treatment interruption. There were no drop-outs and no deviation from the protocol was reported, except for STAR 4 for which a post-hoc lower CD4 nadir was reported. The study population consisted of 11 male and 1 female subjects. Because of the lack of more female participants, we could not perform relevant analyses on the influence of gender on our study results. Participant characteristics are presented in Table S1.

**METHOD DETAILS**

**Inclusion-and Exclusion Criteria**

Inclusion and exclusion criteria are presented in NCT02641756 on clinicaltrial.gov. Twelve HIV-1 subtype B-positive individuals on long-term suppressed ART (2-11y), with stable immune reconstitution (normal CD4 count defined as > 500 CD4/µl or >30%) at time of inclusion and suppressed viral load (<20 copies/ml) for at least 2 years were recruited. All participants were treated with cART consisting of an integrase inhibitor combined with a solid backbone (in all cases 2 NRTIs). The switch to the required treatment regimen was done at least 3 months before the first sampling time point (T1, see Figure 1). Viral rebound was defined as 1 measurement with >1000 HIV RNA copies/ml or 2 consecutive measurements with >200 HIV RNA copies/ml. Randomization was not relevant for this study design.

HIV-1 subtype was defined using the Smartgene Integrated Database Network System’ (Smartgene, Lausanne, Switzerland) and viral tropism was assessed with the Web PSSM tool provided by the Mullins lab, University of Washington (Jensen et al., 2003).

**Clinical Study Design And Sample Collection**

A screening blood analysis (3 x 9ml EDTA tubes) was performed to assess eligibility based on general biological markers including CD4 count and viral load suppression before enrollment in the first part of the study. The total HIV-1 DNA levels in PBMCs was assessed using ddPCR (Kiselinova et al., 2016).

**Under cART (T1)** participants underwent in-depth sampling consisting of lumbar puncture (LP), bone marrow (BM) aspirate, inguinal lymph node resection, bronchoscopy with bronchoalveolar lavage (BAL), gastroscopy and colonoscopy, leukapheresis, urine and vaginal fluid/sperm collection. Participants were admitted at the day-clinic of Ghent University Hospital and procedures took place consecutively. LP with collection of 100 droplets cerebrospinal fluid was done before short narcosis. Under narcosis, a BM aspirate at the posterior superior iliac crest was performed collecting 12ml in a 20ml syringe prefilled with 1 ml of heparine to prevent clotting. Afterwards a gastroscopy and colonoscopy were performed collecting punch biopsies of 3-4mm each at 3 sites (10 duodenal biopsies, 10 ileum biopsies and 20 colon biopsies). An inguinal lymph node excision (1-2 nodes) was performed after ultrasound localization and a BAL concluded the sampling procedures. The lavage was done with physiological water aiming for a 50ml recuperation after aspiration. Leukapheresis was done within one week of the in-depth sampling (+/- 3 days) and in function of participants’ availability. The leukapheresis was not combined with the other sampling procedures because of an increased bleeding risk due to potential thrombocytopenia related to the procedure. We collected an average of 5 billion PBMCs and 100ml plasma. A urine sample and semen sample were collected. Vaginal fluid collection was done at the gynecology department.

If there were medical contraindications for any of the proposed interventions, they were left out or replaced (e.g. peripheral blood draw instead of leukapheresis). If abnormalities were reported during the exams, this was communicated to the participants and adequate follow-up was arranged for.

Approximately 3 months after the initial sampling, STAR 2 to STAR 12 underwent a monitored analytical treatment interruption (ATI). A second screening blood analysis was done before the ATI to ensure that the inclusion criteria were still met. Blood draws were organized twice per week with strict follow-up of viral load and CD4 count. Participants were also interrogated on clinical symptoms related to the treatment stop. A second leukapheresis was done between day 8 and day 15 after the treatment stop (T2). At viral rebound (>1000 copies/ml or 2 consecutive measurements with >200 HIV RNA copies/ml) a less extensive sampling of blood (12x 9ml EDTA tubes), cerebrospinal fluid, urine and genital fluids was performed (T4). Treatment was immediately restarted and a follow-up blood sample was taken after 1 month to assess viral response to re-initiation of the therapy. The study was finalized
with a last blood sample approx. 3 months after undetectable viral load. T3 was identified as the first detectable VL > 30 copies/ml.

Samples were stored on ice until processing in the appropriate collection reservoir and medium (see the Sample processing section). All samples were, on average, processed within 2 hours after collection.

**Sample Processing**

Experiments were not repeated due to the limited amount of sample available, however sampling processing was optimized on a trial participant (STAR 1).

**Gut Associated Lymphoid Tissue (GALT):** GALT samples were collected in 15ml Falcon tubes containing 5ml of RPMI complete (10%FCS, 5% L-glutamine 100x, 5% Penicillin/Streptomycine 100x) and additional antibiotics and antifungals (Piperacillin/tazobactam 500µg/ml and amphotericin B 1.25µg/ml). Cells were isolated following protocol as previously described (Morón-Lopez et al., 2017). Briefly, biopsies were rinsed and incubated in HBSS (Thermo Fisher) containing 1mM DTT (Sigma) and 1mM EDTA (Gibco) for 25min at room temperature and shaken to remove the epithelial layer. Then, biopsies were transferred to the complete media and cultured overnight (at 5%CO2) in 6-well low-binding plates (Costar) at 1-2 biopsies/well. Culture supernatants were collected to recover the released cells and the remaining tissue was disrupted by shaking (10min) and gently pipetting. Cell suspensions were separated from tissue debris through a 40µm mesh filter. Cells were stained for 30 min at 4°C with the following antibodies: CD45 (PercpCy5.5), CD3 (BB), CD8(PeCy7), CD45RO (PE) and CD27 (APC).

**Lymph Node (LN):** Lymph nodes were collected in a 50ml Falcon tube in 20ml of RPMI complete. They were transferred to a sterile plate and cleaned to remove as much fat as possible. A fraction was used for cell isolation and washed over a 70m mesh. Cells were isolated following protocol as previously described. Briefly, biopsies were rinsed and incubated in HBSS (Thermo Fisher) containing 1mM DTT (Sigma) and 1mM EDTA (Gibco) for 25min at room temperature and shaken to remove the epithelial layer. Then, biopsies were transferred to the complete media and cultured overnight (at 5%CO2) in 6-well low-binding plates (Costar) at 1-2 biopsies/well. Culture supernatants were collected to recover the released cells and the remaining tissue was disrupted by shaking (10min) and gently pipetting. Cell suspensions were separated from tissue debris through a 40µm mesh filter. Cells were stained for 30 min at 4°C with the following antibodies: CD45 (PercpCy5.5), CD3 (BB), CD8(PeCy7), CD45RO (PE) and CD27 (APC).

**Peripheral Blood Mononuclear Cells (PBMCs):** PBMCs were collected from blood samples using leukoSep tubes and lymphoprep density gradient separation following manufacturers protocol. PBMCs were collected and counted and a fraction was used for downstream experiments. CD4 negative selection using the BD CD4 enrichment kit was performed and CD4 cells were sorted based on the expression of the following markers: CD3 (BB), CD8 (PeCy7), CD45RO (PE) and CD27 (APC).

Cell sorting was performed using BD FACS Jazz flow cytometer, selecting the lymphocyte gate, singlets and live cells. For these experiments, we used the following sorted subsets (GALT CD45, CD45RO (PE-PERCPCY5.5)), CD3 (BB), CD8 (PeCy7), CD45RO (PE-PERCPCY5.5)), CD27 (APC), CD45RA (APC) and CCR7 (PE).

**Bone Marrow (BM):** A bone marrow aspirate was performed at the posterior iliac crest. Approximately 10-12 mL was collected in a 20ml syringe with 1 mL of heparine to prevent clothing. The aspirate was diluted with PBS and the cell fraction was isolated by ficoll gradient and washed over a 40µm mesh filter. The cells were counted and a fraction was stained for subsequent sorting of CD34+ Lin-cells and CD45+ Lin + cells.

**Cerebrospinal Fluid (CSF):** RNA extraction was performed after ultracentrifugation as described below.

**BAL, urine, and semen samples** were not included in further downstream analysis due to the low amount of cell numbers and/or viral DNA/RNA which did not allow for sequencing analysis.

We obtained a limited amount of sequences from CSF in 5/11 participants (Table 1). These sequences were not used for downstream analysis but were represented in Figure S1 panel D.

We sequenced HIV-1 env V1-V3 from CD4+ T cells from bone marrow samples in 2 participants to illustrate that they intermingle within the phylogenetic trees and do not constitute a separate reservoir (Figure S1 C) As expected, we were not able to sequence from HPC cells derived from bone marrow.

We did not obtain cells from LN in 1 participant (STAR 9). Therefore, more ALG MCM samples were taken and sorted not only into the CD45 subset but also into TEM and TCM. We did not obtain any good sequences from TEM but did obtain sequences from CD45+ Lin-cells and CD45+ Lin + cells.

**Single-Genome/Proviral Sequencing**

We performed single-genome/proviral sequencing of HIV-1 env (V1-V3) as previously described (Josefsson et al., 2012; von Stockenstrom et al., 2015). Briefly, cell pellets were lysed in 100uI home-made lysis buffer (10mM TRisHCl, 0.5%NP-40, 0.5%Tween-20 and proteinase K 20mg/ml) and DNA was extracted by incubating the samples at 55°C for 1 h and subsequently at 85°C for 15min.

Virus from plasma and CSF was first concentrated by ultracentrifugation. Briefly the sample was diluted using Tris buffered saline and centrifuged at 100,000 g for 1 h. The supernatant was discarded and the pellet was resuspended in 100uI of Tris buffered saline. This was repeated 3 times. The resulting pellets were resuspended in 100uI of Tris buffered saline and the samples were stored at -80°C until further analysis.
and transferred to ultracentrifuge tubes (OptiSeal), (Dahl et al., 2014; Palmer et al., 2003). After centrifugation the supernatant was removed and RNA was extracted using an in-house protocol. First TrisHCL pH8 and proteinase K were added to the pellet and incubated at 56°C for 30 min. After that Guanidium thiocyanate and Glycogen were added. After 15 min Isopropanol was added and the sample was centrifuged for 30 min. Supernatant was then removed and two washing steps with ethanol 70% followed. cDNA was generated using SuperScript III Reverse Transcriptase after adding dNTPs (PCR nucleotide mix) and gene-specific primers (E115 for env) to the dissolved pellet. All reactions without reverse transcriptase step (no RT controls) were negative, excluding contamination by cellular debris. DNA and cDNA were serially diluted and nested PCR primers specific to the HIV env region were used to obtain PCR products derived from single HIV-1 DNA molecules (i.e. at the first dilution for which ≤30% positive reactions per cell type are observed). Additional PCRs were performed at the target dilution to obtain 30–40 single proviral sequences where possible. The following primers were used for amplification: Round 1 forward (E20) 5’-GGGCCACACATGCCTGTATACACAG-3’ and reverse (E115) 5’-AGAAAAATTCCTCCACCACATTAA-3’; round 2, forward (E30) 5’-GTTACCCACAGCCCCAGCCAC-3’ and reverse (E125) 5’-CAATTTTCTGGTCCCCCTCTTGAGG-3’. The Platinum Taq PCR polymerase and reaction mix were used. For round 1 of PCR, the following thermocycler parameters were used: 94°C for 2 min, 94°C for 30 s, 50°C for 30 s, 72°C for 1 min, 44 cycles of steps 2–4 and 72°C for 3 min. For round 2 of PCR, the following thermocycler parameters were used: 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, 41 cycles of steps 1–3 and 72°C for 3 min. Read-out of positive reactions was done using E-gel 96 well 1% agarose. The PCR products representing single HIV-1 sequences were sequenced using Sanger sequencing (Australian Genome Research Facility, Sydney, Australia).

### QUANTIFICATION AND STATISTICAL ANALYSIS

**Phylogenetic Analysis**

Low-quality sequences (rejecting sequences using a mean quality score of Q20 (Ewing and Green, 1998)), potential mixtures and defective sequences were eliminated such that only high-quality contigs were used for further analysis (Table S3). Contigs were generated from the raw sequencing data using an in-house computer program written in Perl scripting language (available upon request). Sequences were aligned using MUSCLE (Edgar, 2004) in MEGA7 (http://www.megasoftware.net/) and several tools were used to generate a phylogenetic tree from the alignment (using a neighbor-joining approach). Neighbor joining trees were constructed, including a reference HIV-1 subtype B sequence and positive controls to exclude lab contaminations and inter-participant cross-contamination. Proviral sequences obtained from cell subsets at T2 and T4 and CSF derived sequence data were excluded from downstream analyses respectively because of the risk of reinfection bias after ATI and because of their low overall abundance.

For each participant, the sequence data were aligned using TranslatorX (Abascal et al., 2010) or MAFFT (Katoh and Standley, 2013) and manually edited in AllView (Larsson, 2014). There were clear indications for superinfection of one participant (STAR6), and the rebound dynamics were independently investigated for both lineages (hereafter named clades A and B), which clustered in monophyletic groups with 100% support. We only show the phylogenies of the variant that contributed to virus rebound. Recombinant sequences and their parents were identified with RDP4 (Martin et al., 2015). We attempted to preserve an as large as possible dataset by either excluding the recombinant sequences or the major or minor parental sequences. Recombinant groups that were removed from the main alignment but that included rebound viruses were interrogated separately. Evolutionary relationships between the virus isolates was estimated with maximum likelihood (ML) methods implemented in PhyML v3.0 (Guindon and Gascuel, 2003). The substitution process was modelled as an HKY + Γ process (Yang, 1994) and tree space was searched using the best of NNI and SPR rearrangements. To quantify the contribution of cell subsets and anatomical compartments to rebounding viruses, we computed estimates of the posterior expected number of times (Markov jumps) that each cell subset and compartment seeded rebound virus detected during ATI through stochastic mapping techniques (Minin and Suchard, 2008a; Minin and Suchard, 2008b) from the ML trees.

We performed phylogenetic association using BaTS (Parker et al., 2008) to assess the level of gene flow between anatomical compartments (blood, plasma, central nervous system, lymph node and gut) and between time points (T0, T1, T2, T3 and T4). This analysis is based on the phylogenetic association value (Wang et al., 2001) a measure of the degree of clustering by trait that can be re-scaled in to a phylogenetic association index by dividing them by phylogenetic association values obtained for randomized traits on trees. This ratio takes on values between 0 (absolute clustering by trait) and 1 (randomized clustering by trait). The expectations for the Association index (AI) under the null hypothesis of panmixia were obtained from 1000 taxon-character randomisations, and were scaled in to a phylogenetic association index by dividing them by phylogenetic association values obtained for randomized traits on ML trees.

Details on the data subsetting and filtering steps and the analyses performed on each data set are provided in Table S3. Phylogenies were visualized with FigTree v.1.4.3 (http://beast.community/figtree) or iTol v.3 (Letunic and Bork, 2016). As a graphical representation alternative to maximum likelihood trees, one haplotype network was created per participant using the median-joining method (Bandelt et al., 1999) in NETWORK v.5.0.0.3 (available at http://www.fluxus-engineering.com) and plotted
with the R function “networkGraph” available with the toolbox SPADS 1.0 (Dellicour and Mardulyn, 2014) (Figure S2). In these figures, each haplotype (i.e., unique sequence) is represented by a circle, the size of which is proportional to its overall sampling frequency in a given participant. The genetic relatedness between haplotypes are represented by lines. Each line segment in the networks represents a single mutational change, or multiple changes when the line is annotated by a number indicating the number of mutations. SPADS (Dellicour and Mardulyn, 2014) was also used to compute the nucleotide diversity (Nei and Li, 1979) within different sequence subsets (Figure S3).

**Statistical Analysis**

The sample size of this study (11 participants after exclusion, approximately >400 sequences per participant) was not determined by a sample size calculation. Rather the number of participants was limited by availability and eligibility, and the number of sequences per participant was cost and resource limited. A-priori power calculations were not possible primarily due to the unknown level of variability across participants.

For determining the sequences which were 100% genetically identical, we used Biopython to read and compare the pairwise distance between each HIV-1 env sequence within each participant’s sequence alignment (Cock et al., 2009). The sequence alignment is the input file for our in-house Biopython script and within this sequence alignment each HIV-1 env sequence has an unique identifier. The Biopython script generates an output file that lists those sequences with 100% genetic identity by their unique sequence identifier. We did not find any discrepancy when the results from the Biopython script were compared to the ElimDupes tool from Los Alamos HIV database using 100% genetic identity as the analysis parameter (https://www.hiv.lanl.gov/content/sequence/ELIMDUPES/ElimDupesExplain.html). A sequence was classified as identical if it was a 100% match against another sequences sampled from the same time point, anatomical compartment, cells subset and participant. Once identified, the proportion of identical sequences was calculated as the total number of sequences classified as identical for that group, divided by the total number of sequences for that group. ElimDupes 99% identity mentioned in this manuscript is in fact >99% which includes 99.5%, 99.9% and 100% genetic identity.

Multivariable logistic regression carried out in R version 3.4.1 using function glm, was used to compare the proportion of sequences that were intact, identical (clonal), and identical to rebound. The independent variables in these analyses were participant, cell subset, cell subset with TEM and TCM subsetted amongst anatomical locations, anatomical location, and time point of plasma collection (when appropriate). Anatomical location and cell subset frequently could not be included both due to codependency. Effect modification between an independent variable and participants was frequently strong, however models including this effect modification were unstable with poor standard errors on parameter estimates due to low sample sizes in some participant and cell subset combinations. Considering this, comparisons across subsets of the independent variable were made without considering the effect modification. Additionally, 11 participants is too limited to make population wide inferences. Therefore p-values comparing subsets of an independent variable should be interpreted as evidence for average effects within participants of this sample, rather than an indicator of evidence for population wide trends. To avoid the impact of large expansions on the proportion of identical sequences to rebound virus, clusters of identical sequences were reduced to a single sequence per subset. This reduces variability considerably, but also decreases the statistical power by reducing the number of sequences. In conducting this analysis, the TCM-derived sequences from the blood and LN were pooled and the TEM-derived sequences from the blood and LN were combined to increase sample size as there is no evidence for compartmentalization between these compartments.

The number of cells positive for HIV within a PCR/sequencing replicate was calculated as the proportion of wells positive for HIV by PCR multiplied by the proportion of sequences sequenced from the plate that were truly positive for HIV (defined as high quality forward and reverse sequences, excluding potential mixtures). For 96-well plates with no positive wells, the number of cells positive for HIV was considered as zero. Plates with positive wells of which none were sent for sequencing were excluded; these generally had too many wells positive to be informative. Estimating the true proportion of cells positive based on the Poisson distribution for these plates did not impact the results. The number of cells positive for HIV within a PCR/sequencing replicate, along with the total number of cells used within that dilution were used to estimate the proportion of cells infected with HIV. To compare the proportion of cells infected with HIV across independent variables, a logistic regression was carried out as above with the same strategy for effect modification. Odds ratios of comparisons and their confidence intervals were calculated using the General Linear Hypotheses function (glht) from R package multcomp (Hothorn Torsten and Peter, 2008);

“Presented p-values are not corrected for multiple comparisons. This is because calculated p-values (included and not included in this manuscript) are not independent, and so multiple comparison methods would be overly conservative and bias results towards null effects. To aid interpretation this manuscript has three primary outcomes (proportion of genetically identical virus sequences; infection frequency; and proportion of sequences identical to virus found in rebound plasma) which are compared across three variables (participants, cellular subsets, and anatomical locations) making a total of 5x3=9 primary comparisons. Importantly these comparisons are not independent as all outcome and predictor variables are frequently highly correlated. We advise the reader to interpret P-value results not on a statistical significance cut-off, but rather as one aspect of a level of evidence against null effects. Further assistance in interpreting p-values can be found in references (Betensky, 2019; Wasserstein et al., 2019).
DATA AND CODE AVAILABILITY

GenBank Accession Numbers
The sequences reported in this manuscript are available in GenBank with the following reference codes, MH642355; MH642607; MH642608; MH643062; MH643063; MH643573.

The datasets generated and/or analyzed from these sequences during the current study are available from the corresponding author on reasonable request.

Unreported custom computer code or algorithm used to generate results that are reported in the paper, are available upon request.

ADDITIONAL RESOURCES

Clinical trial registry number: NCT02641756.

Key resource table: further information on resources and materials are available in the Key resource table linked to this article.