

1 **Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia**
2 **before and during analytical treatment interruption**

3 Basiel Cole^{1,#}, Laurens Lambrechts^{1,2,#}, Zoe Boyer³, Ytse Noppe¹, Marie-Angélique De
4 Scheerder⁴, John-Sebastian Eden³, Bram Vrancken⁵, Timothy E. Schlub⁶, Sherry McLaughlin⁷,
5 Lisa M. Frenkel^{7,8}, Sarah Palmer^{3,†}, Linos Vandekerckhove^{1,4,†,*}

6 **Affiliations**

7 ¹HIV Cure Research Center, Department of Internal Medicine and Pediatrics, Ghent University
8 Hospital, Ghent University, Ghent 9000, Belgium.

9 ²BioBix, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience
10 Engineering, Ghent University, Ghent 9000, Belgium

11 ³Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney,
12 Sydney 2145, NSW, Australia.

13 ⁴Department of General Internal Medicine and Infectious Diseases, Ghent University
14 Hospital, Corneel Heymanslaan 10, Ghent 9000, Belgium.

15 ⁵Department of Microbiology, Immunology and transplantation, Rega Institute, Laboratory of
16 Evolutionary and Computational Virology, KU Leuven-University of Leuven, Leuven 3000, Belgium.

17 ⁶University of Sydney, Faculty of Medicine and Health, Sydney School of Public Health, Sydney 2000,
18 NSW, Australia.

19 ⁷Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle,
20 Washington, United States of America.

21 ⁸Departments of Global Health, Laboratory Medicine, Medicine, and Pediatrics, University of
22 Washington, Seattle, Washington, United States of America.

23

24 [#]These authors contributed equally

25 [†]These authors contributed equally

26 ^{*}Correspondence and requests for materials should be addressed to L.V.

27 (email:linos.vandekerckhove@ugent.be)

28 **Summary**

29 The HIV-1 reservoir is composed of cells harboring latent proviruses that are capable of
30 contributing to viremia upon antiretroviral treatment (**ART**) interruption. Although this reservoir is
31 known to be maintained by clonal expansion, the contribution of large, infected cell clones to
32 residual viremia and viral rebound remains underexplored. Here, we conducted an extensive analysis
33 on four ART-treated individuals who underwent an analytical treatment interruption (**ATI**). We
34 performed subgenomic (V1-V3 *env*), near full-length proviral and integration site sequencing, and
35 used multiple displacement amplification to sequence both the integration site and provirus from
36 single HIV-infected cells. We found eight proviruses that could phylogenetically be linked to plasma
37 virus obtained before or during the ATI. This study highlights a role for HIV-infected cell clones in the
38 maintenance of the replication-competent reservoir and suggests that infected cell clones can
39 directly contribute to rebound viremia upon ATI.

40 **Keywords**

41 HIV, latency, replication competency, antiretroviral therapy, full-length HIV sequencing, single
42 proviral sequencing, cellular proliferation, clonal expansion, rebound

43 **Introduction**

44 HIV-1 infection remains incurable due to the presence of a persistent viral reservoir, capable
45 of contributing to viral rebound upon treatment interruption (**TI**) (Chun and Fauci, 1999; Chun et al.,
46 1997, 1998; Finzi et al., 1997). Despite efforts to better understand the dynamics and persistence of
47 the HIV-1 viral reservoir, pinpointing the origins of rebounding viruses remains elusive (De Scheerder
48 et al., 2019). Previously, it was shown that infected CD4 T cells can undergo clonal expansion,
49 contributing to the long-term persistence of the HIV-1 viral reservoir during antiretroviral therapy
50 (**ART**) (Boritz et al., 2016; Cohn et al., 2015a; Einkauf et al., 2018; Hosmane et al., 2017; Maldarelli et
51 al., 2014; Salantes et al., 2018; Simonetti et al., 2016; Wagner et al., 2014a; Wang et al., 2018). The
52 observation that low-level viremia (**LLV**) under ART (Aamer et al., 2020; Bailey et al., 2006; Brennan
53 et al., 2009; Halvas et al., 2020; Tobin et al., 2005; Wagner et al., 2013) and rebound viremia upon TI
54 (Aamer et al., 2020; Kearney et al., 2016; Lu et al., 2018a; De Scheerder et al., 2019) often consist of
55 monotypic populations of viruses, suggest that HIV-1 infected cell clones are key contributors to
56 refueling viremia during TI. Clonality of infected cells has historically been demonstrated by
57 recovering identical proviral sequences or identical integration sites (**IS**) in multiple cells (Cohn et al.,

58 2015b; Hiener et al., 2017; Lee et al., 2017; Maldarelli et al., 2014; Pinzone et al., 2019;
59 Stockenstrom et al., 2015; Wagner et al., 2014b). While the former method allows for qualitative
60 assessment of the proviral genome, it is often not adequate to confidently predict clonal expansion
61 of HIV-1 infected cells, especially when evaluating a short subgenomic region (Lambrechts et al.,
62 2020; Laskey et al., 2016). On the other hand, analysis of IS provides direct proof of clonal expansion,
63 though it typically leaves the proviral sequence uncharacterized. Recently, three techniques to link
64 near full-length (**NFL**) proviral sequences to IS were developed by Einkauf *et al.* (Einkauf et al., 2018),
65 Patro *et al.* (Patro et al., 2019) and Artesi *et al.* (Artesi et al., 2021), respectively called Matched
66 Integration site and Proviral sequencing (**MIP-Seq**), Multiple Displacement Amplification Single-
67 Genome Sequencing (**MDA-SGS**) and Pooled CRISPR Inverse PCR sequencing (**PCIP-seq**). These
68 assays combine the qualitative strength of NFL HIV-1 sequencing with IS analysis, shedding light on
69 the integration profile of intact versus defective proviruses.

70 Analytical treatment interruption (**ATI**) studies allow for the investigation of the dynamics
71 and genetic makeup of rebounding viruses (Clarridge et al., 2018; Garner et al., 2017; Kearney et al.,
72 2016; Pannus et al., 2020). To identify the source of rebounding viruses, we previously conducted
73 the HIV-STAR (HIV-1 sequencing before analytical treatment interruption to identify the anatomically
74 relevant HIV reservoir) study (De Scheerder et al., 2019). During this study, in-depth sampling was
75 performed on 11 chronically treated HIV-1 infected participants prior to ATI. Cells were isolated from
76 different anatomical compartments and sorted into several CD4 T cell subsets. Subgenomic proviral
77 sequences (V1-V3 region of *env*) were recovered and phylogenetically linked to sequences from
78 rebounding plasma virus collected during different stages of the ATI. This study suggested that HIV-1
79 rebound is predominantly fueled by genetically identical viral expansions, highlighting the potentially
80 important role of clonal expansion in the maintenance of the HIV-1 reservoir. While this study
81 yielded a total of 4329 V1-V3 *env* sequences from peripheral blood mononuclear cells (**PBMCs**),
82 lymph node (**LN**) and gut-associated lymphoid tissue (**GALT**), enabling a detailed investigation of the
83 viral reservoir and its relation to rebound viremia, it left some questions unanswered. Most
84 importantly, the evaluation of a short subgenomic region (V1-V3 *env*) to link proviral sequences to
85 rebounding plasma virus made it impossible to investigate the entire genome structure of proviruses
86 linked to rebound. Furthermore, the lack of integration site (**IS**) analysis did not allow for the study
87 of the chromosomal location of the rebounding proviruses.

88 To address these points, we performed a combination of multiple displacement
89 amplification (**MDA**), IS analysis and NFL proviral sequencing on four participants that were enrolled
90 in the HIV-STAR study, with special attention to clonally expanded HIV-1 infected cells. We

91 demonstrate that HIV-1 proviral sequences and corresponding IS of clonally expanded infected cells
92 could be retrieved, and in rare cases these could be linked to low-level viremia during ART and
93 rebound viremia upon ATI, highlighting the clinical relevance of large, infected cell clones.

94 **Results**

95 **Experimental set-up**

96 To investigate the genetic composition and chromosomal location of proviruses within
97 clonally expanded cells, and their relationship to rebound viremia, several qualitative assays were
98 performed on samples from chronically treated HIV-1 infected individuals undergoing an ATI (Figure
99 1A, Table S1). Samples from these individuals were obtained longitudinally before (T1) and during
100 the ATI (T2, T3, T4), as summarized in Figure 1B.

101 First, the overall landscape of HIV-1 infected cell clones prior to ATI (T1, Figure 1B) was
102 determined by subgenomic single-genome sequencing (**SGS**) and Full-length Individual Proviral
103 Sequencing (**FLIPS**) of proviruses, and by Integration Site Loop Amplification (**ISLA**) at the integration
104 site level (Figure 1A, Table S1). This yielded three datasets that were used independently to identify
105 potential clonally expanded infected cell populations.

106 Next, to find links between the different datasets, MDA was performed at limiting dilution
107 on sorted cell lysates from peripheral blood obtained before ATI (T1). MDA wells were subjected to
108 V1-V3 *env* SGS and ISLA, and MDA reactions that yielded a V1-V3 *env* sequence and/or an IS
109 corresponding to a suspected cellular clone were further investigated. This was determined by an
110 exact link to ISLA/FLIPS/SGS data generated in the first step, or by identical V1-V3 *env* sequences
111 and/or IS shared between several MDA wells. The NFL genomes of the proviruses in these selected
112 MDA wells were amplified and sequenced using either a one-amplicon, four-amplicon, or five-
113 amplicon approach, or a combination thereof (see Methods). These MDA-derived NFL sequences
114 were subsequently mapped back to proviral FLIPS sequences and historic V1-V3 *env* proviral
115 sequences from PBMCs, GALT and LN subsets prior to ATI (T1, Figure 1B), as well as V1-V3 *env*
116 plasma-derived RNA sequences retrieved during the ATI (T2-T4, Figure 1B). In addition, remaining T4
117 plasma samples from all four participants were subjected to 5'- and 3'-half genome amplification to
118 complement existing V1-V3 *env* plasma SGS data.

119 This set-up allowed for the assessment of the genetic structure of proviruses in clonally
120 expanded infected cells, their placement across cellular subsets and anatomical compartments, and
121 their contribution to residual viremia on ART and rebound viremia during an ATI.

122 **Integration site analysis and full-length proviral sequencing**

123 To gain insight into the composition of the viral reservoirs of the four STAR participants,
124 especially in terms of clonal expansion of infected cells, we initially performed bulk NFL proviral
125 sequencing and IS analysis.

126 ISLA was performed on endpoint-diluted non-amplified cell lysate and on MDA-amplified cell
127 lysates of central memory/transitional memory (**TCM/TTM**) and effector memory (**TEM**) CD4 T cell
128 subsets from peripheral blood (T1) from three of the four participants: STAR 9, STAR 10 and STAR 11
129 (Figure 2A, Tables S2 and S3). Analysis of IS revealed a significantly higher degree of clonally
130 expanded HIV-1 infected cells in the TEM subset (mean 55%) compared to the TCM/TTM fraction
131 (mean 16%) in the peripheral blood ($P < 0.001$ for STAR 9 and STAR 11; $P < 0.05$ for STAR 10), as
132 previously reported (Hiener et al., 2017). Identical IS between different subsets, suggestive of linear
133 differentiation from an originally infected TCM/TTM into a TEM, was observed in rare instances, with
134 IS at 8 unique positions in the human genome shared between subsets out of 328 IS recovered (171
135 in TCM/TTM and 157 in TEM).

136 Near full-length proviral sequencing (spanning 92% of the proviral genome) was performed
137 on the TCM/TTM and TEM subsets in the peripheral blood and on CD45+ cells from the GALT for all
138 four participants (T1, Figure 1B). In addition, based on sample availability, other cell subsets from the
139 peripheral blood and LN were assayed with FLIPS for some participants, as listed in Table S1. This
140 yielded a total number of 479 proviral genomes with a mean of 120 genomes per participant (Figure
141 2B, Table S3). Across all participants, 29 (6%) intact proviral genomes were retrieved, with a majority
142 of proviral sequences (68%, $n=327$) displaying large internal deletions (Figures S1 and S2A). Overall,
143 intact proviral genomes were found more frequently in the TEM fraction compared to the TCM/TTM
144 fraction (Figure S2D). In addition, the overall HIV-1 infection frequency differed significantly between
145 lymphocyte cell subsets from the peripheral blood ($P < 0.001$), with the TEM subset having the
146 highest infection frequency, except for participant STAR 4 (Figure S2C). Expansions of identical
147 sequences (**EIS**), suggestive of clonal expansion of infected cells, were observed in all participants
148 (Figure 2C). In accordance with the ISLA results, EIS were more frequent in the TEM fraction
149 compared to the TCM/TTM fraction ($P < 0.05$ for STAR 4; $P = 0.23$ for STAR 9; $P = 0.061$ for STAR 10;
150 $P < 0.001$ for STAR 11, Figure 2C).

151 **Multiple displacement amplification-mediated characterization of near full-length proviruses**

152 MDA-mediated HIV-1 provirus and IS sequencing offers the unique opportunity of linking
153 NFL proviral sequences to their precise location in a chromosome. Applying this technique to three
154 of the four study participants, we detected several expanded clones from which NFL sequences and
155 linked IS were sequenced (n=12), as shown in Figure 3A (Table S3).

156 In total, 8 clonal cell populations with a defective provirus were identified: 1 with
157 hypermutation, 1 with a frameshift, and 6 with a defect in the packaging signal and/or major splice
158 donor site (Figure 3A). Somewhat less prevalent were clones with a putatively intact proviral
159 genome (n=4), 3 of these were retrieved in STAR 11 (Figure 3A). Interestingly, 5 out of 8 defective
160 clones were found in both the TCM/TTM and the TEM fractions, which suggests linear differentiation
161 of clonal cell populations (Figures 3B-3D). In contrast, all the clones with intact proviruses were
162 found exclusively in the TEM fraction (Figures 3C and 3D).

163 Of note, we identified one cell clone with an IS in *STAT5B*, harboring a provirus with a 25-bp
164 deletion in the packaging signal (Figure 3A). Previous work has shown that proviral integration into
165 the first intron of this gene in the *cis* configuration can lead to HIV LTR-driven dysregulation of
166 *STAT5B*, resulting in cellular proliferation (Cesana et al., 2017). However, in the present study, the
167 provirus was integrated in the first intron in the *trans* configuration, suggesting that integration in
168 the *trans* configuration can also result in aberrant transcription of *STAT5B*, or that the clone was
169 expanded by mechanisms unrelated to the IS.

170 We conclude that a large fraction of the clonally expanded infected cell populations we
171 identified harbor defective proviruses and that these are frequently found across different cellular
172 subsets, suggesting differentiation of infected cell populations.

173 **Large discrepancies between suspected clonal HIV-1 infected cell populations identified with ISLA,** 174 **SGS and FLIPS**

175 ISLA, SGS and FLIPS can independently be used to assess clonality of infected cells, the
176 former based on the integration site and the two latter on a (subgenomic) sequence of the proviral
177 genome. To investigate whether the methods appear biased in their ability to detect specific clones,
178 we used V1-V3 *env* or NFL sequences to assess overlap between assays (Figure 4). In the case of IS
179 data, only those IS that were associated with a corresponding V1-V3 *env* sequence (identified
180 through MDA) could be linked to other assays. For FLIPS sequences, proviruses that had an internal

181 deletion covering the V1-V3 *env* region could not be linked to SGS data. Matches between IS (upper
182 panel) and FLIPS data (bottom panel) were based on NFL sequences, whereas other links were based
183 on V1-V3 *env* sequences (IS and SGS; SGS and FLIPS). For matches between NFL sequences, up to 3-
184 bp differences were allowed to account for PCR-induced errors and sequencing errors, while 100%
185 concordance was required for V1-V3 *env* matches. To assess the capacity of the V1-V3 *env* region to
186 distinguish different NFL proviruses, the clonal prediction score (**CPS**) was calculated for each
187 participant (Laskey et al., 2016) (Table S5). In addition, the nucleotide diversity of the V1-V3 *env*
188 region was calculated (Table S5).

189 Upon comparison of EIS present in SGS data and FLIPS data from STAR 4, one clear overlap
190 could be found in the peripheral blood TCM/TTM fraction. All other proviral sequences retrieved
191 with SGS could not be linked unequivocally to EIS identified using FLIPS, indicating a significant bias.
192 However, one V1-V3 *env* sequence found with SGS in the TEM and the TCM/TTM fractions matched
193 two different unique FLIPS sequences (Figure 4, green arrows). This is an example of a presumed
194 clonal EIS as detected by SGS that consists of two or more proviruses sharing the same V1-V3 *env*
195 region, although differing elsewhere in their genome. This discrepancy is reflected by a CPS of 96%,
196 with a V1-V3 *env* nucleotide diversity of 0.01 (Table S5).

197 A similar picture was observed for STAR 9, with only a single overlap between IS and FLIPS
198 data. One major clone, with a hypermutated provirus integrated at an intergenic region on
199 chromosome 11, was detected with both MDA and FLIPS. In both assays, this clone was
200 predominantly found in the peripheral blood TEM fraction (23% and 29% respectively), but also
201 appeared in the peripheral blood TCM/TTM fraction. However, this provirus was never amplified
202 with SGS, which can be explained by the fact that V1-V3 *env* primers did not anneal to this
203 hypermutated sequence. The CPS for STAR 9 was 100%, with a V1-V3 *env* nucleotide diversity of
204 0.017 (Table S5).

205 Participant STAR 10 displayed a clear discrepancy between the assays. One small EIS from
206 the TEM SGS data was shared an identical V1-V3 *env* with two different EIS found in the FLIPS data,
207 one of which can be linked to the *CASC5* cell clone (Figure 4, blue arrows). This is most likely the
208 result of multiple distinct proviruses sharing an identical V1-V3 *env* sequence but integrated at
209 different sites. This resulted in a lower CPS of 95%, with a V1-V3 *env* nucleotide diversity of 0.014
210 (Table S5).

211 In contrast with the previous participants, remarkable consistency between assays was
212 observed for STAR 11, with all the clonal NFL sequences identical to those from both SGS and MDA-

213 derived IS data (Figure 4). However, the largest clone based on ISLA data, integrated in the *ZFC3H1*
214 gene (Figure 2A), could not be linked to SGS and FLIPS data, which was probably the result of large
215 internal deletions spanning the entire length of the genome. In fact, out of ten MDA wells that
216 yielded this integration site, the proviral sequence could not be amplified by V1-V3 *env* SGS, or by a
217 one-, four- or five-amplicon approach NFL sequencing (Table S3). The CPS for STAR 11 was 100%,
218 with a V1-V3 *env* nucleotide diversity of 0.017 (Table S5).

219 In conclusion, we demonstrate that for two out of four participants, the CPS is lower than
220 100%, suggesting linkage inaccuracies when using the V1-V3 *env* to predict clonality of infected cells.
221 Furthermore, we show compartmentalization between the viral populations identified by the V1-V3
222 *env* SGS method versus the FLIPS method. This could either result from primer bias or from sampling
223 bias, caused by the high frequency of *env*-deleted proviruses sampled by FLIPS.

224 **Plasma viral sequences match intact proviruses and proviruses with large internal deletions or** 225 **defects in the packaging signal**

226 In our previous study on HIV-STAR participants, proviral V1-V3 *env* SGS sequences from
227 multiple lymphocyte subsets and anatomical tissues collected prior to ATI were linked to plasma
228 RNA sequences of rebounding viruses (De Scheerder et al., 2019). Yet, no conclusions about the
229 genomic structure of the NFL proviruses and their associated IS could be inferred, since these
230 subgenomic sequences did not allow for such analysis. Data generated in the present study allowed
231 for a deeper characterization of the proviral landscape through linkage of NFL proviral sequences
232 and their corresponding IS to plasma RNA sequences of rebounding viruses.

233 To investigate how the new FLIPS- and MDA-derived NFL sequences compare to the historic
234 plasma and proviral V1-V3 *env* sequences generated during the original HIV-STAR study (De
235 Scheerder et al., 2019), the trimmed V1-V3 *env* region from the MDA- and FLIPS- derived NFL
236 sequences were aligned with SGS-derived and MDA-derived V1-V3 *env* sequences. Subsequently,
237 phylogenetic trees were constructed for each participant (Figure 5). As seen before in Figure 4, the
238 V1-V3 *env* regions of some NFL proviruses are an identical match to historic proviral V1-V3 *env* SGS
239 sequences and cluster together on the same branch (Figures 5 and S4). Interestingly, identical
240 sequences were detected in different anatomical compartments. For example, participant STAR 9
241 displays an intact FLIPS provirus which falls into a cluster of proviral peripheral blood and GALT V1-
242 V3 *env* sequences (Figure 5B). Furthermore, participant STAR 10 has an MDA-derived provirus
243 integrated in *STAT5B* that matches the V1-V3 *env* proviral sequences from the peripheral blood and
244 LN, suggesting that cells of this clone traffic between these compartments (Figure 5C).

245 A total of 8 matches between plasma virus and NFL proviral sequences could be found,
246 which are highlighted corresponding to their NFL category (Figure 5, inner circles). To further
247 investigate the relationship between these matches in more detail, phylogenetic trees including NFL
248 proviral sequences and V1-V3 *env* plasma sequences were constructed (Figure 6). In addition to the
249 latter, the alignment was complemented with trimmed 3'-half viral plasma sequences obtained at
250 T4. This analysis revealed that most of the trimmed 3'-half plasma sequences intermingled with the
251 previously obtained V1-V3 *env* sequences, showing good accordance between the methods (Figure
252 6). In addition, alignments using proviral sequences and either of the 5'- or 3'- plasma viruses were
253 constructed and checked for the potential of recombination, yet no evidence for such events was
254 observed.

255 Out of the 8 V1-V3 *env* matches between a NFL provirus and a plasma virus, 3 involved a
256 plasma virus obtained either when ART-suppressed before the ATI (T1) or during the ATI but before
257 a detectable viral load (T2). All of these links were found in participant STAR 11, which displays a CPS
258 of 100% for the V1-V3 *env* region (Figure 6D). The first match was between a defective, MDA-derived
259 provirus with a frameshift (integrated at chr17:7545670) found in both the TCM/TTM and TEM
260 peripheral blood subsets, and a plasma virus at T1 (Figure S4D). The other matches were found
261 between a FLIPS-derived intact provirus from the TCM/TTM peripheral blood subset and a T1 plasma
262 virus, and an MDA-derived intact provirus from the TEM peripheral blood subset and a T2 plasma
263 virus, respectively (Figure 6D and Figure S4D). The latter provirus was found to be integrated in the
264 *ZNF141* gene, which belongs to the Krüppel-associated box domain (KRAB) containing zinc finger
265 nuclease family. Interestingly, this viral sequence was not detected in the plasma from rebounding
266 timepoints T3 and T4, however matches T0 plasma sequences (prior to ART timepoint), suggesting a
267 phylogenetic relationship to a founder virus (Figure 5D).

268 Five out of the 8 plasma to proviral V1-V3 *env* matches include plasma sequences from the
269 rebounding timepoints (T3 and T4 during ATI), but only one of them involved an intact proviral
270 genome (Figure 6A-C). This unique FLIPS provirus in STAR 9 matched plasma sequences found at T2
271 prior to rebound (3/4 plasma sequences from that timepoint) and T4 during rebound. Although this
272 provirus, located in the peripheral TEM subset, was recovered only once using FLIPS, it matched to a
273 cluster of identical SGS V1-V3 *env* sequences found in cells from the peripheral blood and GALT,
274 suggesting that it is part of a clone (Figure 5B and Figure S4B).

275 For participants STAR 10 and 4, one or more plasma sequences from rebound were identical
276 to defective proviral genomes (Figure 6A-B). In fact, STAR 10 had three matches between

277 rebounding V1-V3 *env* sequences and defective proviruses: one matched a unique FLIPS-derived
278 provirus with a large internal deletion originating from peripheral TCM/TTM subset, one matched a
279 provirus with a PSI/MSD deletion belonging to a clonal cell population found in the peripheral TEM
280 subset located in an intergenic region on chromosome 8 and one matched an MDA-derived provirus
281 from the peripheral TEM subset with a probable deletion, integrated in the *ZBTB20* gene. Note that
282 the latter also shared a similar V1-V3 *env* sequence with two 3'-half T4 plasma sequences, yet when
283 the alignment was extended over the entire 3'-half genome, regions with genetic differences were
284 observed (Figure S3D). In STAR 4, a unique FLIPS-derived provirus with a PSI/MSD deletion matched
285 to plasma V1-V3 *env* sequences from all three timepoints during the ATI (T2, T3, T4) in addition to 2
286 different sets of T4 3'-half genomes with identical an V1-V3 *env* (Figure 6B). This FLIPS provirus was
287 derived from GALT, but was identical to two other V1-V3 *env* SGS sequences from the bone marrow
288 and peripheral blood (Figure 5A). When looking at extended alignments based on either 5'- or 3'-half
289 T4 plasma genomes and proviral sequences, a close relatedness between this provirus and the
290 plasma half-genomes could be observed (Figure S3A). At the 5' region, the sequences were identical
291 beside a 105 bp deletion in the PSI/MSD region of the proviral sequence while the two sets of 3'
292 plasma genomes matching the V1-V3 *env* are either 8 bp or 19 bp different to the proviral sequence,
293 suggesting a common ancestor.

294 In conclusion, by performing MDA-mediated NFL and IS analysis, we identified several
295 proviruses with linked IS, predominantly belonging to peripheral blood CD4 memory subsets, that
296 matched sequences from plasma before and/or during an ATI. Interestingly, most of these
297 proviruses were classified as defective, raising the question whether these are still capable of
298 producing viremia. Furthermore, some intact identical proviral sequences were detected in
299 specimens from multiple anatomical compartments, suggesting that certain clones that harbor
300 genome-intact proviruses can traffic between different compartments.

301 Discussion

302 Integration of HIV-1 genomes into the DNA of host cells leads to the establishment of a
303 persistent HIV-1 reservoir. While most of these integrated proviruses are defective, a small
304 proportion are genetically intact and fully capable of producing infectious virions (Hiener et al.,
305 2017). The proportion of genetically intact HIV-1 proviruses, as measured by the Intact Proviral DNA
306 Assay (IPDA), has been shown to decay slowly, with an estimated average half-life of 4 years during
307 the first 7 years of suppression, and 18.7 years thereafter (Peluso et al., 2020). This long half-life can
308 in part be explained by continuous clonal expansion of infected cells harboring these genetically
309 intact HIV-1 proviruses (Liu et al., 2020; Patro et al., 2019). While this phenomenon is well-
310 established, the contribution of clonally expanded HIV-1 infected cells to residual viremia under ART
311 and rebound viremia upon ATI remains underexplored. Previously, others have tried to characterize
312 rebounding viruses by phylogenetically linking these to proviral sequences and viral sequences
313 obtained by viral outgrowth assays (VOA), with limited success. While two studies were unable to
314 find links between rebounding sequences and viral sequences recovered by VOA (Lu et al., 2018b;
315 Vibholm et al., 2019), three other groups did find several links using similar techniques (Aamer et al.,
316 2020; Cohen et al., 2018; Salantes et al., 2018). However, two of these latter studies were performed
317 in the context of interventional clinical trials and the IS of these viruses remained unknown. In
318 addition, two groups were able to link proviral sequences to rebound sequences, though only a small
319 part of the proviral genome was queried (Barton et al., 2016; Kearney et al., 2016). We previously
320 conducted the HIV-STAR study, where SGS on the V1-V3 *env* region was used to link proviral
321 sequences to plasma sequences (De Scheerder et al., 2019). We found multiple links between
322 proviral sequences and rebounding plasma sequences, however, this study was limited by the
323 sequencing of a small subgenomic region of the proviruses and plasma viruses. In the current study,
324 we used a combination of NFL sequencing, IS analysis and MDA-mediated IS/NFL sequencing to
325 more accurately define the source of rebounding virus detected during ATI in a subset of HIV-STAR
326 participants.

327 We first showed large discrepancies between different techniques to assess clonal
328 expansion of HIV-1 infected cells. These discrepancies are often the result of primer biases, dictating
329 which proviruses are amplified. This has important implications for HIV-1 reservoir research, as some
330 assays will be unable to detect potentially relevant proviruses. In addition, we demonstrated that
331 the use of a short subgenomic region of the HIV-1 genome (V1-V3 *env*) to assess clonality of infected
332 cells can lead to inaccurate results. This was shown by the recovery of distinct NFL proviruses,
333 integrated at different sites, displaying identical V1-V3 *env* sequences. Similar observations were

334 made in a recently published study, where P6-PR-RT sequences were compared to matched NFL/IS
335 sequences (Patro et al., 2019). They found multiple instances of identical proviral P6-PR-RT
336 sequences, with distinct IS. Taken together, we conclude that evaluating clonality of HIV-1 infected
337 cells based on the assessment of a subgenomic region should be done with caution.

338 We next set out to find links between NFL proviral sequences and sequences found in the
339 plasma during different stages of an ATI. First, we identified several identical V1-V3 *env* sequences in
340 defective proviruses and rebounding plasma viruses. Interestingly, for participant STAR 4,
341 phylogenetic trees suggest that both V1-V3 *env* and 3'-half genome plasma SGS cluster with a
342 provirus containing a 105 bp packaging signal deletion (including stem-loop 1 and stem-loop 2). It
343 has been shown previously that proviruses with such defects are still capable of producing infectious
344 virions, though with significantly lower efficiency (Pollack et al., 2017). Yet, in the 5'-half plasma
345 dataset, a sequence was found that was completely identical to the PSI-deleted provirus in STAR 4,
346 except that this sequence had an intact PSI, suggesting a close phylogenetic relationship.

347 Three other defective proviruses linked to rebound viruses, all in participant STAR 10,
348 contain large internal deletions, making it unlikely that these are the actual source of the virus
349 rebounding during ATI. Rather, these are probably phylogenetically related proviruses, as they share
350 an identical V1-V3 *env* sequence. Two previous studies that tried to link proviral sequences to
351 rebound sequences, based on full *env* sequences, concluded that while they were not able to
352 directly link the proviral sequences to the rebounding ones, the rebounding sequences could often
353 be accounted for by recombination (Cohen et al., 2018; Vibholm et al., 2019). Because we assessed
354 only a small portion of the *env* gene (V1-V3 region) at timepoints T1-T3, we were not able to
355 comprehensively study recombination events, though we hypothesize that recombination may be a
356 probable cause of identical overlap between defective proviral sequences and rebounding virus
357 sequences. At T4, we recovered half-genome plasma sequences, though these did not show any
358 signs of recombination.

359 We further identified two links between genetically intact NFL proviruses and plasma viruses
360 emerging upon ATI. On both occasions the intact NFL provirus was located in the peripheral TEM
361 blood subset, suggesting these might be easier to reactivate due to higher activation status.
362 Alternatively, this could reflect the higher degree of clonality observed in the TEM subset compared
363 to other memory subsets, which in turn increases the chances of detecting links. The first link was
364 found in participant STAR 9, where an intact provirus obtained with FLIPS could be linked to plasma
365 virus at T2 and T4. Because this provirus was not retrieved in an MDA reaction, the IS remains

366 unknown. Interestingly, this virus was first sampled at T2 and persisted into T4, which suggests that
367 this virus emerged during the phase of an ATI when the viral load was still undetectable. In
368 participant STAR 11, an intact provirus integrated in the *ZNF141* gene could be linked to plasma virus
369 at T2 during an ATI. Another recent publication found a clonal infected cell population with IS in the
370 *ZNF721/ABCA11P* gene, that contributed to persistent residual viremia which was not suppressed by
371 ART (Halvas et al., 2020). This gene is located on chromosome 4 and belongs to the KRAB-containing
372 zinc finger nuclease (ZNF) family. This integration event shows great similarities with the provirus we
373 identified in the *ZNF141* gene, which also belongs to the KRAB-ZNF family and which is located on
374 chromosome 4, just upstream of the *ZNF721/ABCA11P* gene. Interestingly, three other studies also
375 described infected cell clones harboring a genetically-intact provirus integrated in the
376 *ZNF721/ABCA11P* gene, suggesting that this region is a particular hotspot for the persistence of
377 genetically intact proviruses (Einkauf et al., 2018; Halvas et al., 2020; Jiang et al., 2020). Because the
378 plasma virus that was linked to our *ZNF141* clone stems from T2, the latest timepoint with
379 undetectable viral load during the ATI, but did not persist in the later timepoints (T3 and T4), we
380 cannot exclude that the virus we sampled emerged as a result of continuous virus shedding, as
381 described by Aamer *et al.* (Aamer et al., 2020), rather than ‘true’ rebounding virus. Previously, it was
382 suggested that the origin of rebounding plasma viruses includes clonally expanded infected cells that
383 are transcriptionally active before TI (Kearney et al., 2016). Similarly, a recent study found several
384 overlaps between monotypic viremia that was not suppressible by ART and large proviral clones
385 (Halvas et al., 2020). These two findings, together with the observations by Aamer *et al.* (Aamer et
386 al., 2020), leads to the expectation that the provirus integrated in the *ZNF141* gene is a prime
387 candidate to contribute to viral rebound, however, our current data does not support this. Of
388 course, we cannot exclude that this viral strain was not identified at T3 and T4 because it was
389 obscured by other rebound viruses, and thus not included among the variants we sequenced.

390 In a recent study it was observed that ‘elite controllers’ (EC), individuals that control HIV-1
391 infection spontaneously, often carry genetically intact proviral sequences integrated at spots
392 associated with ‘deep latency’, which persist over time and are not cleared by the immune
393 system (Jiang et al., 2020). In one EC, they described a persistently infected cell population with an
394 intact provirus integrated in the *ZNF274* gene, which is associated with highly condensed chromatin.
395 Interestingly, we also observed a clonally expanded infected cell population in the peripheral blood
396 TEM fraction from STAR11, with a genetically intact provirus integrated in the *ZNF274* gene. Despite
397 the rather large size of the clone, we did not observe the emergence of the corresponding viral
398 sequence in the plasma during the ATI, which is in agreement with its presumed ‘deep latent’ state.
399 In fact, it is possible that because of the heterochromatin state of the DNA at this spot, this provirus

400 would tend to remain latent. Alternatively, we cannot exclude that this virus was not identified
401 during the ATI due to timing of our specimen collection. Indeed, it is possible that this virus would be
402 detected if the ATI would have been prolonged and if the participant was sampled at later time-
403 points, especially knowing that transcription at this specific IS could be diminished and, if possible at
404 all, would need more time to complete. These findings add to the current understanding that not all
405 genetically intact proviruses contribute to the ‘replication competent HIV-1 viral reservoir’, as some
406 are unlikely to rebound due to an unfavorable IS, though they may possess all the necessary
407 attributes to rebound under specific conditions.

408 A study by Bertagnolli *et al.* showed that the outgrowth of a substantial fraction of viruses of
409 the latent reservoir is blocked by autologous IgG antibodies against HIV-1 envelope (Bertagnolli *et*
410 *al.*, 2020). This mechanism might explain the discrepancy between proviruses recovered *ex vivo* and
411 viruses recovered from the plasma. Indeed, the population of viruses that rebound might have been
412 shaped by immune pressure, which is absent when assessing proviral sequences recovered from
413 extracted DNA. This phenomenon further complicates finding links between proviruses and plasma
414 viruses.

415 We acknowledge several limitations in this study. The first is the limited sampling from
416 tissue compartments, possibly causing us to miss important rebound lineages. Indeed, it has been
417 shown that tissues, including lymph nodes and GALT, harbor most of the HIV-1 latent reservoir,
418 orders of magnitude higher than the peripheral blood compartment (Estes *et al.*, 2017). Whether
419 there is compartmentalization between different anatomical compartments is under debate. Several
420 studies, including our previously conducted HIV-STAR study, have suggested that there is limited
421 compartmentalization between the HIV-1 proviral sequences recovered from lymph nodes and from
422 peripheral blood (Josefsson *et al.*, 2013; Mcmanus *et al.*, 2019; De Scheerder *et al.*, 2019;
423 Stockenstrom *et al.*, 2015; Vibholm *et al.*, 2019), based on identical proviral sequences and/or IS
424 shared between both compartments. In addition, our previous HIV-STAR study did not show
425 evidence of any enrichment of rebounding sequences stemming from specific anatomical
426 compartments (De Scheerder *et al.*, 2019), justifying our decision to focus the current study
427 primarily on the peripheral blood compartment. The second limitation of the current study is that
428 the link to plasma sequences at T1-T3 is based on the V1-V3 *env* region, rather than on NFL plasma
429 sequences. This means that we cannot exclude the possibility that links between proviral sequences
430 and rebounding plasma sequences are the result of matches in V1-V3 *env* but with genetic variation
431 outside of this region, however the CPS for the V1-V3 *env* region for participants STAR 9 and STAR
432 11, which display links between intact proviral sequences and plasma rebound sequences, was

433 calculated at 100%. Furthermore, the lack of matching sequences between half-genome plasma
434 sequences and proviral sequences from T1 might be because at T4 the plasma typically was
435 dominated by a genetically oligoclonal pool of viruses, which might have obscured less fit
436 rebounding viruses that match T1 proviruses.

437 In conclusion, our data show that reservoir characterization using multiple methods,
438 including IS analysis, NFL proviral sequencing and a combination of both, enables the identification
439 of matches between proviral sequences and plasma sequences recovered before and/or during an
440 ATI, however these matches are rare. While our findings confirm that expanded HIV infected cell
441 clones present in the peripheral blood can contribute to both residual and rebound plasma viremia,
442 the origins of a large fraction of rebounding viruses remained unknown. Future studies should focus
443 on in-depth characterization of tissue reservoirs to further investigate their relative contribution to
444 rebound viremia.

445 **Methods**

446 **Samples**

447 A total of four HIV-1 infected, ART treated participants were included in this study. All had an
448 undetectable viral load (<20 copies/ml) for at least 1 year prior to treatment interruption, and all
449 initiated ART during the chronic phase of infection. The participants characteristics are summarized
450 in Table S6. Participants were sampled longitudinally, prior to and during an ATI (Figure 1B).
451 Anatomical compartments that were sampled, and corresponding cell subsets sorted from these, are
452 summarized in Table S1.

453 **CD4 T cell subset sorting**

454 Cryopreserved PBMCs were thawed and CD4 T cell enrichment was carried out with negative
455 magnet-activated cell sorting (Beckton Dickinson, BD IMag™, #557939). CD4 T cells were stained
456 with the following monoclonal antibodies: CD3 (Becton Dickinson, #564465), CD8 (Becton Dickinson,
457 #557746), CD45RO (Becton Dickinson, #555493), CD27 (Becton Dickinson, #561400), CCR7 (Becton
458 Dickinson, #560765) and a fixable viability stain (Becton Dickinson, #565388). Fluorescence-activated
459 cell sorting was used to sort stained peripheral blood-derived CD4 T cells into naïve CD4 T cells
460 (CD45RO-, CD45RA+), central memory/transitional memory CD4 T cells (CD3+ CD8- CD45RO+ CD27)
461 and effector memory CD4 T cells (CD3+ CD8- CD45RO+ CD27-), GALT cells into CD45+ cells and cells
462 from lymph nodes into central memory/transitional memory CD4 T cells (CD3+ CD8- CD45RO+
463 CD27+) and effector memory CD4 T cells (CD3+ CD8- CD45RO+ CD27-), using a BD FACSJazz cell
464 sorter machine, as previously described (De Scheerder et al., 2019). The gating strategy used for the
465 aforementioned sorts can be found in Figure S5. A small fraction of each sorted cell population was
466 analyzed by flow cytometry to check for purity, which was over 95% on average. Flow cytometry
467 data was analyzed using FlowJo software (Tree-Star).

468 **Droplet digital PCR (ddPCR)**

469 Sorted cells were pelleted and lysed in 100µL lysis buffer (10mM TrisHCl, 0.5% NP-40, 0.5% Tween-
470 20 and proteinase K at 20 mg/ml) by incubating for 1 hour at 55°C and 15 min at 85°C. HIV-1 copy
471 number was determined by a total HIV-1 DNA assay on droplet digital PCR (Bio-Rad, QX200 system),
472 as described previously (Rutsaert et al., 2019). PCR amplification was carried out with the following
473 cycling program: 10 min at 98°C; 45 cycles (30 sec at 95°C, 1 min at 58°C); 10 min at 98°C. Droplets

474 were read on a QX200 droplet reader (Bio-Rad). Analysis was performed using ddpcRquant software
475 (Trypsteen et al., 2015).

476 **Whole genome amplification (WGA)**

477 Cell lysates were diluted according to ddPCR HIV-1 copy quantification, so that less than 30% of
478 reactions contained a single proviral genome. Whole genome amplification was performed by
479 multiple displacement amplification with the REPLI-g single cell kit (Qiagen, #150345), according to
480 manufacturer's instructions. The resulting amplification product was split for downstream IS
481 analysis, single genome/proviral sequencing, and, for selected reactions, near full-length HIV-1
482 sequencing.

483 **Single genome/proviral sequencing**

484 Single genome/proviral sequencing (SGS) of the V1-V3 region of *env* was performed as described
485 before (Josefsson et al., 2012; Von Stockenstrom et al., 2015), with a few adaptations. The
486 amplification consists of a nested PCR with the following primers: Round 1 forward (E20) 5'-
487 GGGCCACACATGCCTGTGTACCCACAG-3' and reverse (E115) 5'-AGAAAAATCCCCTCCACAATTAA-3';
488 round 2, forward (E30) 5'-GTGTACCCACAGACCCAGCCCAAG-3' and reverse (E125) 5'-
489 CAATTTCTGGGTCCCCTCTGAGG-3'. The 25 µL PCR mix for the first round is composed of: 5 µL 5X
490 Mytaq buffer, 0.375 µL Mytaq polymerase (Bioline, #BIO-21105), 400 nM forward primer, 400 nM
491 reverse primer and 1 µL REPLI-g product. The mix for the second round has the same composition
492 and takes 1 µL of the first-round product as an input. Thermocycling conditions for first and second
493 PCR rounds are as follows: 2 min at 94°C; 35 cycles (30 sec at 94°C, 30 sec at 60°C, 1 min at 72°C); 5
494 min at 72°C. Resulting amplicons were visualized on a 1% agarose gel and Sanger sequenced
495 (Eurofins Genomics, Ebersberg, Germany) from both ends, using second round PCR primers.

496 Both 5'- and 3'-half genome amplicons were generated from T4 plasma samples. RNA was extracted
497 from the virions and cDNA was generated as follows: 1) Plasma samples were thawed at 37°C. 2)
498 Remove debris by centrifuging the plasma for 10 min at 3600 rpm and discarding the pellet. 3)
499 Transfer supernatant to ultracentrifuge tube and adjust volume to 9 mL with PBS. 4) Centrifuge at
500 >85.000 g for 70 min at 4°C. 5) 240 µL of the supernatant is subjected to viral RNA extraction using
501 the QIAamp Viral RNA Mini Kit (Qiagen, #52904), according to manufacturer's instructions. 6) Half of
502 the RNA was used to generate cDNA for 5'-half sequencing using the R5968 primer (5'-
503 TGTCTYCKCTTCTTCTGCCATAG-3'), while the other half was used to generate cDNA for 3'-half
504 sequencing using the primer R9665 (5'- GTCTGAGGGATCTCTAGWTACCAGA-3'). Two mastermixes

505 were prepared. Mix 1 consisted of 25 μ L RNA, 2.5 μ L 10mM dNTP, 1.25 μ L 20 μ M oligo-dT
506 (SuperScript III First Strand synthesis system, Invitrogen, #18080051) and 0.5 μ L 50 μ M primer. Mix 2
507 consisted of 0.75 μ L RNase free water, 5X RT buffer (Invitrogen, #18080051), 2.5 μ L 100 mM DTT,
508 2.5 μ L 40U/ μ L RNase inhibitor (Takara, #2313B), 2.5 μ L SuperScript III Reverse Transcriptase
509 (Invitrogen, #18080051) and 2 μ L ThermaSTOP RT (Sigma Aldrich, #TSTOPRT-250). Mix 1 was heated
510 to 65°C for 5 min and then snap-chilled on ice for at least 2 min. Mix 2 was pre-warmed to 50°C and
511 then added to the chilled mix 1. The mixture was incubated at 50°C for 90 min. 1 μ L SuperScript III
512 Reverse Transcriptase was added to the reaction, followed by another incubation of 90 min at 50°C
513 and then 70°C for 15 min. Finally, 1 μ L RNase H (Invitrogen, #18080051) was added, followed by an
514 incubation of 20 min at 37°C. Subsequently, the cDNA was used as template for half-genome long-
515 range PCRs, as described previously (Cole et al., 2021). The 25 μ L PCR mix for the first round was
516 composed of: 5 μ L 5X Prime STAR GXL buffer, 0.5 μ L PrimeStar GXL polymerase (Takara
517 Bio, #R050B), 0.125 μ L ThermaStop (Sigma Aldrich, #TSTOP-500), 250 nM forward and reverse
518 primers, and 1 μ L MDA product. The mix for the second round had the same composition and took 1
519 μ L of the first-round product as an input. Thermocycling conditions for first and
520 second PCR rounds were as follows: 2 min at 98°C; 35 cycles (10 sec at 98°C, 15 sec at 62°C, 5 min at
521 68°C); 7 min at 68°C. Reactions without reverse transcriptase were negative, ensuring that the RNA
522 extracts were not contaminated by DNA. PCR products were checked on a 1% agarose gel and
523 positives were sequenced by Illumina sequencing, as described below.

524 **Integration site loop amplification (ISLA)**

525 Integration site sequencing was carried out by integration site loop amplification (ISLA), as described
526 by Wagner *et al.* (Wagner et al., 2014a), but with a few modifications. Firstly, the *env* primer used
527 during the linear amplification step was omitted, as it was not necessary to recover the *env* portion
528 of the provirus at a later stage. Therefore, the reaction was not split after the linear amplification,
529 and the entire reaction was used as an input into subsequent decamer binding and loop formation.
530 For some proviruses, an alternative set of primers were used to retrieve the IS from the 5' end (Table
531 S7). Resulting amplicons were visualized on a 1% agarose gel and positives were sequenced by
532 Sanger sequencing. Analysis of the generated sequences was performed using the 'Integration Sites'
533 webtool developed by the Mullins lab;
534 <https://indra.mullins.microbiol.washington.edu/integrationsites/>.

535 **Full-length individual proviral sequencing assay**

536 Proviral sequences from the genomic DNA of sorted subsets were recovered by the Full-length
537 Individual Proviral Sequencing (FLIPS) assay as first described by Hiener *et al.* (Hiener et al., 2017)
538 with some minor alterations. Briefly, the assay consists of two rounds of nested PCR at an end-point
539 dilution where 30% of the wells are positive. This yields proviral fragments of up to 9 kb using the
540 following primers for the first round BLOuterF (5'-AAATCTCTAGCAGTGGCGCCCGAACAG-3') and
541 BLOuterR (5'-TGAGGGATCTCTAGTTACCAGAGTC-3') followed by a second round using primers 275F
542 (5'-ACAGGGACCTGAAAGCGAAAG-3') and 280R (5'-CTAGTTACCAGAGTCACACAACAGACG-3'). The
543 cycling conditions are 94°C for 2 m; then 94°C for 30 s, 64°C for 30 s, 68°C for 10 m for 3 cycles; 94°C
544 for 30 s, 61°C for 30 s, 68°C for 10 m for 3 cycle; 94°C for 30 s, 58°C for 30 s, 68°C for 10 m for 3
545 cycle; 94°C for 30 s, 55°C for 30 s, 68°C for 10 m for 21 cycle; then 68°C for 10 m. For the second
546 round, 10 extra cycles at 55°C are included. The PCR products were visualized using agarose gel
547 electrophoresis. Amplified proviruses from positive wells were cleaned using AMPure XP beads
548 (Beckman Coulter), followed by a quantification of each cleaned provirus with Quant-iT PicoGreen
549 dsDNA Assay Kit (Invitrogen). Next, an NGS library preparation using the Nextera XT DNA Library
550 Preparation Kit (Illumina) with indexing of 96-samples per run was used according to the
551 manufacturer's instructions, except that input and reagents volumes were halved and libraries were
552 normalized manually. The pooled library was sequenced on a MiSeq Illumina platform via 2x150 nt
553 paired-end sequencing using the 300 cycle v2 kit.

554 **Near full-length provirus amplification from MDA reactions**

555 MDA reactions containing a potentially clonal proviral sequence were subjected to near full-length
556 proviral sequencing, using either a single-amplicon approach (Hiener et al., 2017), a four-amplicon
557 approach (Patro et al., 2019), or a five-amplicon approach (Einkauf et al., 2018), as previously
558 described. In case of the multiple amplicon approaches, amplicons were pooled equimolarly and
559 sequenced as described above.

560 ***De Novo* assembly of HIV-1 proviruses and analysis**

561 The generated sequencing data from either FLIPS or multiple amplicon approaches was
562 demultiplexed and used to *de novo* assemble individual proviruses. The code used to perform *de*
563 *novo* assembly can be found at the following GitHub page:
564 https://github.com/laulambr/virus_assembly. In short, the workflow consists of following steps: (i)
565 check of sequencing quality for each library using FastQC
566 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and removal of Illumina adaptor
567 sequences and trimming of 5' and 3' terminal ends using BBtools

568 (sourceforge.net/projects/bbmap/). (ii) The trimmed reads are *de novo* assembled using MEGAHIT
569 (Li et al., 2015) generating contigs for each library. (iii) Per library, all *de novo* contigs were checked
570 using blastn against the HXB2 reference virus as a filter to exclude non-HIV-1 contigs in the following
571 analysis steps. (iv) Subsequently, the trimmed reads were mapped against the *de novo* assembled
572 HIV-1 contigs to enable the calling of the final majority consensus sequence of each provirus using
573 bbmap. Alignments of proviral sequences for each participant were made via MAFFT (Katoh et al.,
574 2002) and manually inspected via MEGA7 (Kumar et al., 2016). The generated HIV-1 proviruses were
575 categorized as intact or defective as described previously (Hiener et al., 2017). NFL proviruses and
576 half-genome plasma sequences were screened for recombination by the “DualBrothers” software
577 (Minin et al., 2005) and the “Recombinant Identification Program” webtool from the Los Alamos
578 National Laboratory HIV sequence database (<https://www.hiv.lanl.gov>). Phylogenetic trees were
579 constructed using PhyML v3.0 (Guindon et al., 2010) (best of NNI and SPR rearrangements) and 1000
580 bootstraps. MEGA7 (Kumar et al., 2016) and iTOL v5 (Letunic and Bork, 2019) were used to visualize
581 phylogenetic trees.

582 **Statistical analysis**

583 P-values in Figure 2A and Figure 2C test for a difference in the proportion of respectively unique IS or
584 unique proviruses between TCM/TTM and TEM subsets. P-values were calculated using “prop.test”
585 command in R versions 3.6.2 (“R Core Team,” 2020). Infection frequencies for FLIPS data were
586 calculated by expressing the total number of identified HIV positive cells as a proportion of all cells
587 analysed. The infection frequency was compared across cellular subsets using a logistic regression on
588 the number of cells positive for HIV and total number of cells using “glm” function in R. Interaction
589 between participant and cellular subset was detected ($P < 0.001$) and included in the logistic
590 regression. P-values were calculated using the “Anova” function from the “car” package in R (John
591 and Sanford, 2019).

592 **Data availability statement**

593 Data will be uploaded to public repositories upon acceptance of the manuscript.

594 **Study approval**

595 This study was approved by the Ethics Committee of the Ghent University Hospital (Belgian
596 registration number: B670201525474). Written informed consent was obtained from all study
597 participants.

598 **Acknowledgements and funding sources**

599 We would like to acknowledge and thank all participants who donated samples to the HIV-STAR
600 study, and all the MDs and study nurses that assisted with the sample collection. We would also like
601 to thank Marion Pardons, Tine Struyve and Sofie Rutsaert for providing guidance during initial data
602 analyses, for the constructive discussions and for critically reading the manuscript. We are grateful
603 for the discussions with and input from James Mullins, Rafick Sékaly, Susan Pereira Ribeiro, Hadega
604 Aamer, Sam Kint, Oleg Denisenko, Katie Fisher and Bethany Horsburgh. In addition, we would like to
605 thank Kim De Leeneer, Céline Helmoortel and Bram Parton for their assistance in performing MiSeq
606 sequencing at UZ Ghent. This current research work was supported by the NIH (R01-AI134419, MPI:
607 LV and LF) and the Research Foundation Flanders (S000319N and G0B3820N). LV was supported by
608 the Research Foundation Flanders (1.8.020.09.N.00) and the Collen-Francqui Research Professor
609 Mandate. SP was supported by the Delaney AIDS Research Enterprise (DARE) to Find a Cure
610 (1U19AI096109 and 1UM1AI126611-01) and the Australian National Health and Medical Research
611 Council (APP1061681 and APP1149990). The sample collection at UZ Ghent was supported by an
612 MSD investigator grant (ISS 52777). BC and LL were supported by FWO Vlaanderen (1S28918N,
613 1S29220N). BV was supported by a postdoctoral grant (12U7121N) of the Research Foundation -
614 Flanders (Fonds voor Wetenschappelijk Onderzoek).

615 **Author contributions**

616 BC, LL, LF, SP and LV conceptualized the experiments. MADS processed the samples from the initial
617 HIV STAR study, including cell isolation from peripheral blood and tissue, and she performed cell
618 sorting and single-genome sequencing. BC and YN performed experiments involving cell sorting,
619 multiple displacement amplification, single-genome sequencing and integration site sequencing. LL
620 and ZB performed experiments involving near full-length proviral sequencing. BC, LL, BV, JSE and TS
621 analyzed data and performed associated analyses. BC, LL, TS and BV made figures and tables. BC and
622 LL wrote the manuscript. All co-authors edited and approved the manuscript.

623 **Declaration of interests**

624 The authors declare that no conflict of interest exists.

625 **Figure legends**

626

627 **Figure 1: Overview of the workflow for HIV-1 reservoir characterization and viral loads at each**
628 **timepoint of sample collection.** (A) Workflow of HIV-1 reservoir characterization by single-genome
629 sequencing (**SGS**), full-length individual proviral sequencing (**FLIPS**), integration site loop
630 amplification (**ISLA**) and multiple displacement amplification (**MDA**). In a first step, potentially clonal
631 HIV-1 infected cells were identified by SGS, FLIPS and ISLA on lysed sorted CD4 T-cell subsets. In a
632 second step, MDA with subsequent SGS and ISLA was performed on selected sorted cell lysates. In
633 the final step, MDA reactions containing a potentially clonal provirus were identified and the near
634 full-length (**NFL**) genome of the according provirus was amplified and sequenced. (B) Viral load
635 (copies/mL) at each time of sample collection for all participants. The day of analytical treatment
636 interruption (**ATI**) initiation is indicated with a vertical red line. The plasma was sampled during ART
637 (timepoint 1, T1), 8 to 14 days after ATI (timepoint 2, T2), at the first detectable viral load (timepoint
638 3, T3), and at later rebound (timepoint 4, T4). Note that T1 is not shown to scale. The horizontal
639 dashed lines indicate the limit of detection at 20 copies/mL.

640

641 **Figure 2: Classification of HIV-1 integration sites and proviral near full-length genome sequences**
642 **across different cell subsets before ATI.** (A) Proportions of integration sites (**IS**) retrieved by
643 integration site loop amplification (**ISLA**) for participants STAR 9, STAR 10 and STAR 11 from
644 TCM/TTM and TEM subsets from peripheral blood. IS found more than once are defined as “clonal”
645 and are shown in color as proportion of all IS. Identical IS found in both subsets are linked with
646 dashed lines. P-values test was used for a difference in the proportion of unique IS between
647 TCM/TTM and TEM by “prop.test” in R. (B) Proportions of intact and defective near full-length
648 sequences from Full-Length Individual Provirus sequencing (**FLIPS**) within all sequenced proviruses
649 from peripheral blood, gut-associated lymphoid tissue (**GALT**) and lymph nodes for each participant.
650 (C) Proportions of identical sequences (**EIS**) found in TCM/TTM and TEM peripheral blood subsets
651 based on FLIPS data for each participant. EIS consisting of a defective and intact provirus are shown
652 in shades of red and green respectively while unique proviruses are grouped in the unique provirus
653 category. P-values test was used for a difference in the proportion of unique proviruses between
654 TCM/TTM and TEM by “prop.test” in R. TCM/TTM = central/transitional memory CD4 T cell, TEM =
655 effector memory CD4 T cell.

656

657

658

659 **Figure 3: Near full-length proviral HIV-1 genomes and associated integration sites recovered from**
660 **the peripheral blood by MDA.** (A) For each participant, the recovered proviral genome structures
661 are shown aligned to the HXB2 reference genome and corresponding integration sites, if available,
662 are listed on the right. Each provirus is colored according to their structural category. (B-D) For each
663 participant, the number of integration site (IS) and near full-length (NFL) proviruses linked to each
664 multiple displacement amplification (MDA) clone are shown together with their corresponding
665 cellular subset. TCM/TTM = central/transitional memory CD4 T cell, TEM = effector memory CD4 T
666 cell, NA = not available

667
668 **Figure 4: Comparison of assays to identify potentially clonal HIV-1 infected cell populations.** The
669 total number of examined integration sites (IS), V1-V3 *env* sequences and near full-length (NFL)
670 proviral sequences is noted in the middle of each donut plot. Sequences found multiple times within
671 the same assay are colored by a shade of grey, purple or blue (for integration site loop amplification
672 (ISLA), single-genome sequencing (SGS) and full-length individual proviral sequencing (FLIPS)
673 respectively). When NFL or V1-V3 *env* sequences could be linked to an identified multiple
674 displacement amplification (MDA) cell clone, they were given a distinct standout color and
675 chromosome designation as indicated in the legend. Populations of identical FLIPS or ISLA sequences
676 that are not associated with a V1-V3 *env* sequence (due to deletions and/or primer mismatches) are
677 shaded. Arrows are used to indicate discrepancies between the different assays. EIS = expansion of
678 identical sequences.

679
680 **Figure 5: Integration sites linked to circular HIV-1 V1-V3 *env* maximum likelihood phylogenetic**
681 **trees for each participant using all generated proviral and plasma V1-V3 *env* sequences before and**
682 **during different stages of the ATI.**

683 The plasma and proviral sequences were obtained either prior ART initiation (timepoint 0, T0),
684 during ART (timepoint 1, T1), 8 to 14 days after analytical treatment interruption (ATI) (timepoint 2,
685 T2), at the first detectable viral load (timepoint 3, T3), and at later rebound (timepoint 4, T4). The
686 inner circle represents the sequence type, either obtained through single-genome sequencing (SGS)
687 of the V1-V3 *env* region shown in grey and V1-V3 *env* trimmed near full-length (NFL) genomes in
688 colors indicating their intactness category. The middle circle shows the integration site associated
689 with multiple displacement amplification (MDA) derived proviruses if available. The integration sites
690 in the legend are shown in order of appearance on the circle. The outer circle displays the origin
691 (sampling timepoint and/or anatomical compartment) of each plasma and proviral sequence.
692 Matches (placed on same branch) of identical V1-V3 *env* regions between plasma and proviral NFL

693 sequences are shown in bold lines, where the line color reflects the intactness category of the
694 matching NFL virus.

695 NA = not available, GALT = gut-associated lymphoid tissue.

696

697 **Figure 6: Maximum-likelihood phylogenetic trees of V1-V3 *env* sequences derived from trimmed**
698 **FLIPS and/or MDA proviral sequences before ATI and (trimmed) rebounding plasma viruses before**
699 **and during different stages of ATI.** Proviral sequences derived from Full-Length Individual Provirus
700 sequencing (**FLIPS**) and multiple displacement amplification (**MDA**) are shown as squares and circles
701 respectively. The integration sites (**IS**) associated with MDA-derived proviruses are noted if available.
702 Plasma sequences are shown as triangles (V1-V3 *env*) or diamonds (3'-half genome) where the color
703 indicates the timepoint during analytical treatment interruption (**ATI**). Arrows indicate identical
704 matches between proviral and plasma V1-V3 *env* sequences. All trees are rooted to the HXB2
705 reference sequence. (A) In participant STAR 10, three identical matches between defective proviral
706 and plasma rebound sequences were found. For two, the corresponding IS *ZBTB20* and
707 Chr8:100792121 could be recovered. (B) In participant STAR 4, only one match between a unique
708 major splice donor (**MSD**) deleted provirus and plasma sequences was observed. (C) In STAR 9, a
709 match between a unique intact provirus and multiple plasma sequences from different timepoints
710 were found. (D) In STAR 11, a rebounding plasma sequence could be linked to an expansion of
711 identical intact near full-length (**NFL**) genomes located in the *ZNF141* gene. One unique intact
712 provirus can be linked to a residual plasma sequence from T1. SGS = single-genome sequencing,

713 **References**

- 714 Aamer, H.A., McClure, J., Ko, D., Maenza, J., Collier, A.C., Mullins, J.I., and Frenkel, L.M. (2020). Cells
715 producing residual viremia during antiretroviral treatment appear to contribute to rebound viremia
716 following interruption of treatment. *PLOS Pathogens* *16*, e1008791.
- 717 Artesi, M., Hahaut, V., Cole, B., Lambrechts, L., Ashrafi, F., Marçais, A., Hermine, O., Griebel, P., Arsic,
718 N., van der Meer, F., et al. (2021). PCIP-seq: simultaneous sequencing of integrated viral genomes
719 and their insertion sites with long reads. *Genome Biol* *22*, 97.
- 720 Bailey, J.R., Sedaghat, A.R., Kieffer, T., Brennan, T., Lee, P.K., Wind-rotolo, M., Haggerty, C.M.,
721 Kamireddi, A.R., Liu, Y., Lee, J., et al. (2006). Residual Human Immunodeficiency Virus Type 1 Viremia
722 in Some Patients on Antiretroviral Therapy Is Dominated by a Small Number of Invariant Clones
723 Rarely Found in Circulating CD4+ T Cells. *Journal of Virology* *80*, 6441–6457.
- 724 Barton, K., Hiener, B., Winckelmann, A., Rasmussen, T.A., Shao, W., Byth, K., Lanfear, R., Solomon, A.,
725 McMahon, J., Harrington, S., et al. (2016). Broad activation of latent HIV-1 in vivo. *Nature*
726 *Communications* *7*, 1–8.
- 727 Bertagnolli, L.N., Varriale, J., Sweet, S., Brockhurst, J., Simonetti, F.R., White, J., Beg, S., Lynn, K.,
728 Mounzer, K., Frank, I., et al. (2020). Autologous IgG antibodies block outgrowth of a substantial but
729 variable fraction of viruses in the latent reservoir for HIV-1. *Proc Natl Acad Sci U S A* *117*, 32066–
730 32077.
- 731 Boritz, E.A., Darko, S., Swaszek, L., Wolf, G., Wells, D., Wu, X., Henry, A.R., Laboune, F., Hu, J.,
732 Ambrozak, D., et al. (2016). Multiple Origins of Virus Persistence during Natural Control of HIV
733 Infection. *Cell* *166*, 1004–1015.
- 734 Brennan, T.P., Woods, J.O., Sedaghat, A.R., Siliciano, J.D., Siliciano, R.F., and Wilke, C.O. (2009).
735 Analysis of Human Immunodeficiency Virus Type 1 Viremia and Provirus in Resting CD4+ T Cells
736 Reveals a Novel Source of Residual Viremia in Patients on Antiretroviral Therapy. *Journal of Virology*
737 *83*, 8470–8481.
- 738 Cesana, D., Santoni de Sio, F.R., Rudilosso, L., Gallina, P., Calabria, A., Beretta, S., Merelli, I., Bruzzesi,
739 E., Passerini, L., Nozza, S., et al. (2017). HIV-1-mediated insertional activation of STAT5B and BACH2
740 trigger viral reservoir in T regulatory cells. *Nature Communications* *8*, 498.
- 741 Chun, T., and Fauci, A.S. (1999). Perspective Latent reservoirs of HIV: Obstacles to the eradication
742 of virus. *Proceedings of the National Academy of Sciences of the United States of America* *96*,
743 10958–10961.
- 744 Chun, T.W., Stuyver, L., Mizell, S.B., Ehler, L. a, Mican, J. a, Baseler, M., Lloyd, a L., Nowak, M. a, and
745 Fauci, a S. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral
746 therapy. *Proceedings of the National Academy of Sciences of the United States of America* *94*,
747 13193–13197.
- 748 Chun, T.W., Engel, D., Berrey, M.M., Shea, T., Corey, L., and Fauci, A.S. (1998). Early establishment of
749 a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proceedings of the*
750 *National Academy of Sciences of the United States of America* *95*, 8869–8873.
- 751 Clarridge, K.E., Blazkova, J., Einkauf, K., Petrone, M., Refsland, W., Justement, J.S., Shi, V., Huiting,
752 E.D., Seamon, C.A., Lee, G.Q., et al. (2018). Effect of analytical treatment interruption and reinitiation

753 of antiretroviral therapy on HIV reservoirs and immunologic parameters in infected individuals. *PLoS*
754 *Pathogens* *14*, e1006792.

755 Cohen, Y.Z., Lorenzi, J.C.C., Krassnig, L., Barton, J.P., Burke, L., Pai, J., Lu, C.L., Mendoza, P., Oliveira,
756 T.Y., Sleckman, C., et al. (2018). Relationship between latent and rebound viruses in a clinical trial of
757 anti – HIV-1 antibody 3BNC117. *Journal of Experimental Medicine* *215*, 2311–2324.

758 Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L.,
759 McElrath, M.J., Lehmann, C., et al. (2015a). HIV-1 integration landscape during latent and active
760 infection. *Cell* *160*, 420–432.


761 Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L.,
762 McElrath, M.J., Lehmann, C., et al. (2015b). HIV-1 integration landscape during latent and active
763 infection. *Cell* *160*, 420–432.

764 Cole, B., Lambrechts, L., Gantner, P., Noppe, Y., Bonine, N., Witkowski, W., Chen, L., Palmer, S.,
765 Mullins, J.I., Chomont, N., et al. (2021). In-depth single-cell analysis of translation-competent HIV-1
766 reservoirs identifies cellular sources of plasma viremia. *Nat Commun* *12*, 3727.

767 Einkauf, K., Lee, G.Q., Gao, C., Sharaf, R., Sun, X., Hua, S., Chen, S., Jiang, C., Lian, X., Chowdhury, F.Z.,
768 et al. (2018). Distinct chromosomal positions of intact HIV-1 proviruses. *Journal of Clinical*
769 *Investigation* *129*, 988–998.

770 Estes, J.D., Kityo, C., Ssali, F., Swainson, L., Makamdop, K.N., Prete, G.Q. Del, Deeks, S.G., Luciw, P.A.,
771 Chipman, J.G., Beilman, G.J., et al. (2017). Defining total-body AIDS-virus burden with implications
772 for curative strategies. *Nature Medicine* *23*, 1271–1276.

773 Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick,
774 K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on
775 highly active antiretroviral therapy. *Science* *278*, 1295–1300.

776 Garner, S.A., Rennie, S., Ananworanich, J., Dube, K., Margolis, D.M., Sugarman, J., Tressler, R.,
777 Gilbertson, A., and Dawson, L. (2017). Interrupting antiretroviral treatment in HIV cure research: :
778 scientific and ethical considerations. *Journal of Virus Eradication* *3*, 82–84.

779 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New
780 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance
781 of PhyML 3.0. *Systematic Biology* *59*, 307–321.

782 Halvas, E.K., Hughes, S.H., Mellors, J.W., Halvas, E.K., Joseph, K.W., Brandt, L.D., Guo, S., Sobolewski,
783 M.D., Jacobs, J.L., Tumiotto, C., et al. (2020). HIV-1 viremia not suppressible by antiretroviral therapy
784 can originate from large T cell clones producing infectious virus. *Journal of Clinical Investigation* *130*,
785 5847–5857.

786 Hiener, B., Horsburgh, B.A., Eden, J.S., Barton, K., Schlub, T.E., Lee, E., von Stockenstrom, S., Odevall,
787 L., Milush, J.M., Liegler, T., et al. (2017). Identification of Genetically Intact HIV-1 Proviruses in
788 Specific CD4+T Cells from Effectively Treated Participants. *Cell Reports* *21*, 813–822.

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

- 793 Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K.B., Chevalier, J.M., Chen, S.M.Y., Hua, S., Rhee, B.,
794 Chang, K., et al. (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1.
795 *Nature* *585*, 261–267.
- 796 John, F., and Sanford, W. (2019). *An R Companion to Applied Regression* (Thousand Oaks CA: Sage).
- 797 Josefsson, L., Eriksson, S., Sinclair, E., Ho, T., Killian, M., Epling, L., Shao, W., Lewis, B., Bacchetti, P.,
798 Loeb, L., et al. (2012). Hematopoietic Precursor Cells Isolated From Patients on Long-term
799 Suppressive HIV Therapy Did Not Contain HIV-1 DNA. *Journal of Infectious Diseases* *206*, 28–34.
- 800 Josefsson, L., von Stockenstrom, S., Faria, N.R., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Tan, A.,
801 Ho, T., Lemey, P., et al. (2013). The HIV-1 reservoir in eight patients on long-term suppressive
802 antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National*
803 *Academy of Sciences of the United States of America* *110*, E4987-96.
- 804 Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple
805 sequence alignment based on fast Fourier transform. *Nucleic Acids Research* *30*, 3059–3066.
- 806 Kearney, M.F., Wiegand, A., Shao, W., Coffin, J.M., Mellors, J.W., Lederman, M., Gandhi, R.T., Keele,
807 B.F., and Li, J.Z. (2016). Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That
808 Are Transcriptionally Active before Stopping of Antiretroviral Therapy. *Journal of Virology* *90*, 1369–
809 1376.
- 810 Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis
811 Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* *33*, 1870–1874.
- 812 Lambrechts, L., Cole, B., Rutsaert, S., Trypsteen, W., and Vandekerckhove, L. (2020). Emerging PCR-
813 Based Techniques to Study HIV-1 Reservoir Persistence. *Viruses* *12*, 1–12.
- 814 Laskey, S.B., Pohlmeier, C.W., Bruner, K.M., and Siliciano, R.F. (2016). Evaluating Clonal Expansion of
815 HIV-Infected Cells: Optimization of PCR Strategies to Predict Clonality. *PLOS Pathogens* *12*,
816 e1005689.
- 817 Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.-H., Hua, S.,
818 Chen, H.-R., Ouyang, Z., et al. (2017). Clonal expansion of genome-intact HIV-1 in functionally
819 polarized Th1 CD4+ T cells. *Journal of Clinical Investigation* *127*, 2689–2696.
- 820 Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new
821 developments. *Nucleic Acids Research* *47*, W256–W259.
- 822 Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node
823 solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*
824 *31*, 1674–1676.
- 825 Liu, R., Simonetti, F.R., and Ho, Y. (2020). The forces driving clonal expansion of the HIV-1 latent
826 reservoir. *Virology Journal* *17*.
- 827 Lu, C., Pai, J.A., Nogueira, L., Mendoza, P., Gruell, H., Oliveira, T.Y., and Barton, J. (2018a).
828 Relationship between intact HIV-1 proviruses in circulating CD4 + T cells and rebound viruses
829 emerging during treatment interruption. *Proceedings of the National Academy of Sciences of the*
830 *United States of America* *115*, 11341–11348.

- 831 Lu, C.-L., Pai, J.A., Nogueira, L., Mendoza, P., Gruell, H., Oliveira, T.Y., Barton, J., Lorenzi, J.C.C.,
832 Cohen, Y.Z., Cohn, L.B., et al. (2018b). Relationship between intact HIV-1 proviruses in circulating
833 CD4 + T cells and rebound viruses emerging during treatment interruption. *Proceedings of the*
834 *National Academy of Sciences of the United States of America* *115*, 11341–11348.
- 835 Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W.,
836 Kearney, M.F., et al. (2014). Specific HIV integration sites are linked to clonal expansion and
837 persistence of infected cells. *Science* *345*, 179–183.
- 838 Mcmanus, W.R., Coffin, J.M., Kearney, M.F., Mcmanus, W.R., Bale, M.J., Spindler, J., Wiegand, A.,
839 Musick, A., Patro, S.C., Sobolewski, M.D., et al. (2019). HIV-1 in lymph nodes is maintained by cellular
840 proliferation during antiretroviral therapy. *Journal of Clinical Investigation* *129*, 4629–4642.
- 841 Minin, V.N., Dorman, K.S., Fang, F., and Suchard, M.A. (2005). Dual multiple change-point model
842 leads to more accurate recombination detection. *Bioinformatics* *21*, 3034–3042.
- 843 Pannus, P., Rutsaert, S., De Wit, S., Allard, S.D., Vanham, G., Cole, B., Nescoi, C., Aerts, J., De
844 Spiegelare, W., Tsoumanis, A., et al. (2020). Rapid viral rebound after analytical treatment
845 interruption in patients with very small HIV reservoir and minimal on-going viral transcription.
846 *Journal of the International AIDS Society* *23*, e25453.
- 847 Patro, S.C., Brandt, L.D., Bale, M.J., Halvas, E.K., Joseph, K.W., Shao, W., and Wu, X. (2019).
848 Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows
849 reconstruction of replicating viral ancestors. *Proceedings of the National Academy of Sciences of the*
850 *United States of America* *116*, 25891–25899.
- 851 Peluso, M.J., Laird, G.M., Deeks, S.G., Peluso, M.J., Bacchetti, P., Ritter, K.D., Beg, S., Lai, J., Martin,
852 J.N., Hunt, P.W., et al. (2020). Differential decay of intact and defective proviral DNA in HIV-1 –
853 infected individuals on suppressive antiretroviral therapy. *JCI Insight* *5*, e132997.
- 854 Pinzone, M.R., Vanbelzen, D.J., Weissman, S., Bertuccio, M.P., Cannon, L., Venanzi-rullo, E., Migueles,
855 S., Jones, R.B., Mota, T., Joseph, S.B., et al. (2019). Longitudinal HIV sequencing reveals reservoir
856 expression leading to decay which is obscured by clonal expansion. *Nature Communications* *10*.
- 857 Pollack, R.A., Jones, R.B., Perteza, M., Bruner, K.M., Martin, A.R., Thomas, A.S., Capoferri, A.A., Beg,
858 S.A., Huang, S.H., Karandish, S., et al. (2017). Defective HIV-1 Proviruses Are Expressed and Can Be
859 Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host and Microbe*
860 *21*, 494–506.
- 861 “R Core Team” (2020). R: A language and environment for statistical computing.
- 862 Rutsaert, S., Spiegelare, W. De, Clercq, L. De, and Vandekerckhove, L. (2019). Evaluation of HIV-1
863 reservoir levels as possible markers for virological failure during boosted darunavir monotherapy.
864 *Journal of Antimicrobial Chemotherapy* *74*, 3030–3034.
- 865 Salantes, D.B., Tebas, P., Bar, K.J., Salantes, D.B., Zheng, Y., Mampe, F., Srivastava, T., Beg, S., Lai, J.,
866 Li, J.Z., et al. (2018). HIV-1 latent reservoir size and diversity are stable following brief treatment
867 interruption. *Journal of Clinical Investigation* *128*, 3102–3115.
- 868 De Scheerder, M.-A., Vrancken, B., Dellicour, S., Schlub, T., Lee, E., Shao, W., Rutsaert, S.,
869 Verhofstede, C., Kerre, T., Malfait, T., et al. (2019). HIV Rebound Is Predominantly Fueled by
870 Genetically Identical Viral Expansions from Diverse Reservoirs. *Cell Host & Microbe* *26*, 347–358.

- 871 Simonetti, F.R., Sobolewski, M.D., Fyne, E., Shao, W., Spindler, J., Hattori, J., Anderson, E.M.,
872 Watters, S.A., Hill, S., Wu, X., et al. (2016). Clonally expanded CD4+ T cells can produce infectious
873 HIV-1 in vivo. *Proceedings of the National Academy of Sciences of the United States of America* *113*,
874 1883–1888.
- 875 Stockenstrom, S. Von, Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Shao, W.,
876 Hoh, R., Ho, T., et al. (2015). Longitudinal Genetic Characterization Reveals That Cell Proliferation
877 Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *Journal of Infectious*
878 *Diseases* *212*, 596–607.
- 879 Von Stockenstrom, S., Odevall, L., Lee, E., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Shao, W.,
880 Hoh, R., Ho, T., et al. (2015). Longitudinal Genetic Characterization Reveals That Cell Proliferation
881 Maintains a Persistent HIV Type 1 DNA Pool during Effective HIV Therapy. *Journal of Infectious*
882 *Diseases* *212*, 596–607.
- 883 Tobin, N.H., Learn, G.H., Holte, S.E., Wang, Y., Melvin, A.J., McKernan, J.L., Pawluk, D.M., Mohan,
884 K.M., Lewis, P.F., Mullins, J.I., et al. (2005). Evidence that low-level viremias during effective highly
885 active antiretroviral therapy result from two processes: expression of archival virus and replication
886 of virus. *Journal of Virology* *79*, 9625–9634.
- 887 Trypsteen, W., Vynck, M., Neve, J. De, Bonczkowski, P., Vandekerckhove, L., Spiegelaere, W. De, De
888 Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., et al. (2015). ddpcRquant: threshold
889 determination for single channel droplet digital PCR experiments. *Analytical and Bioanalytical*
890 *Chemistry* *407*, 5827–5834.
- 891 Vibholm, L.K., Lorenzi, J.C.C., Pai, J.A., Cohen, Y.Z., Oliveira, T.Y., Barton, J.P., Garcia Noceda, M., Lu,
892 C.-L., Ablanedo-Terrazas, Y., Del Rio Estrada, P.M., et al. (2019). Characterization of Intact Proviruses
893 in Blood and Lymph Node from HIV-Infected Individuals Undergoing Analytical Treatment
894 Interruption. *Journal of Virology* *93*, e01920-18.
- 895 Wagner, T.A., McKernan, J.L., Tobin, N.H., Tapia, K.A., Mullins, J.I., Frenkel, M., and Frenkel, L.M.
896 (2013). An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment
897 suggests proliferation of HIV-infected cells. *Journal of Virology* *87*, 1770–1778.
- 898 Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y.K., Larsen, B.B., Styrchak, S., Huang, H.C.,
899 Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014a). Proliferation of cells with HIV integrated into
900 cancer genes contributes to persistent infection. *Science* *345*, 570–573.
- 901 Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y.K., Larsen, B.B., Styrchak, S., Huang, H.C.,
902 Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014b). Proliferation of cells with HIV integrated into
903 cancer genes contributes to persistent infection. *Science* *345*, 570–573.
- 904 Wang, Z., Gurule, E.E., Brennan, T.P., Gerold, J.M., Kwon, K.J., Hosmane, N.N., Kumar, M.R., Beg, S.A.,
905 Capoferri, A.A., Ray, S.C., et al. (2018). Expanded cellular clones carrying replication-competent HIV-
906 1 persist, wax, and wane. *Proceedings of the National Academy of Sciences of the United States of*
907 *America* *115*, E2575–E2584.

908

Figure 1

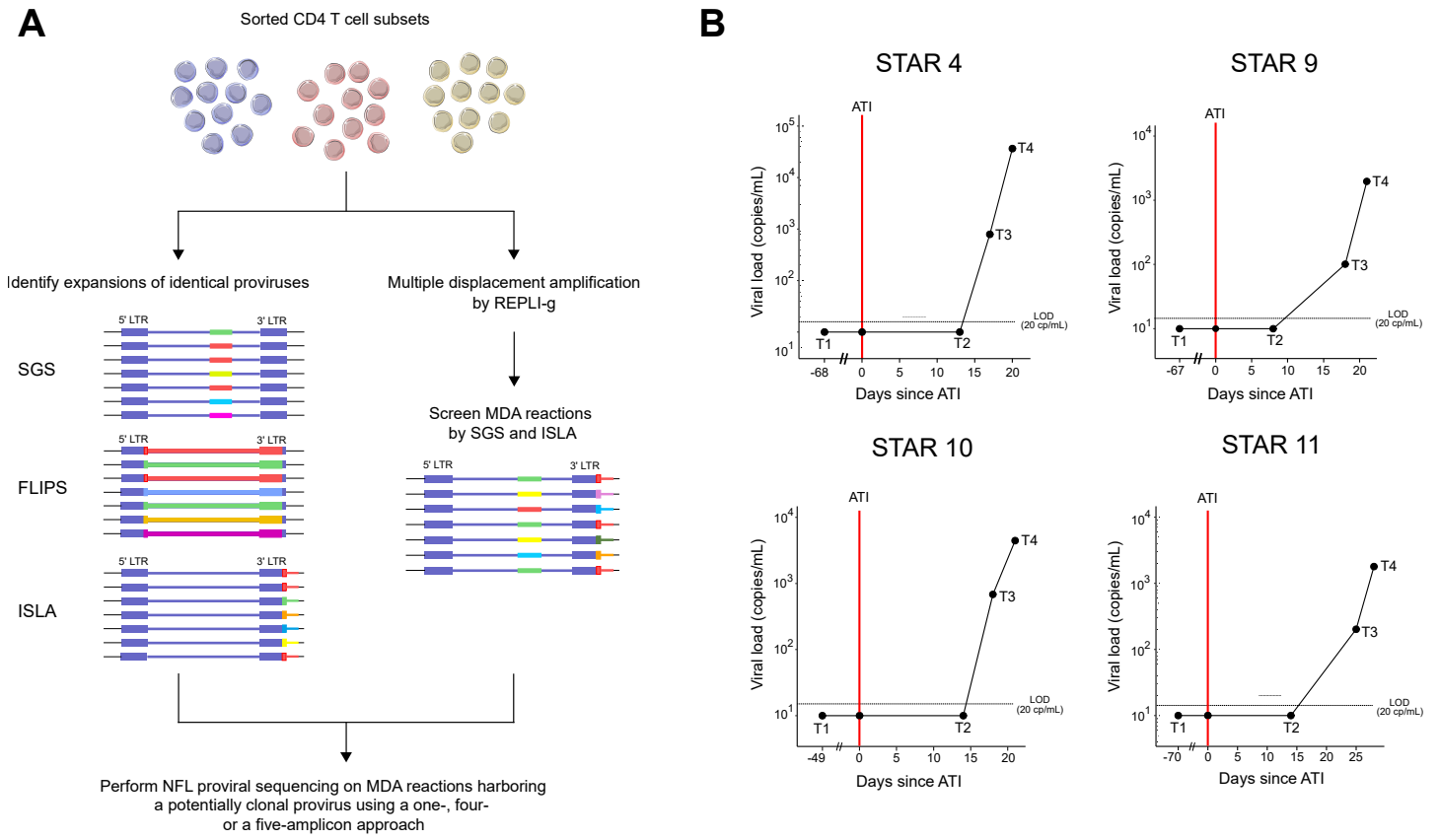


Figure 2

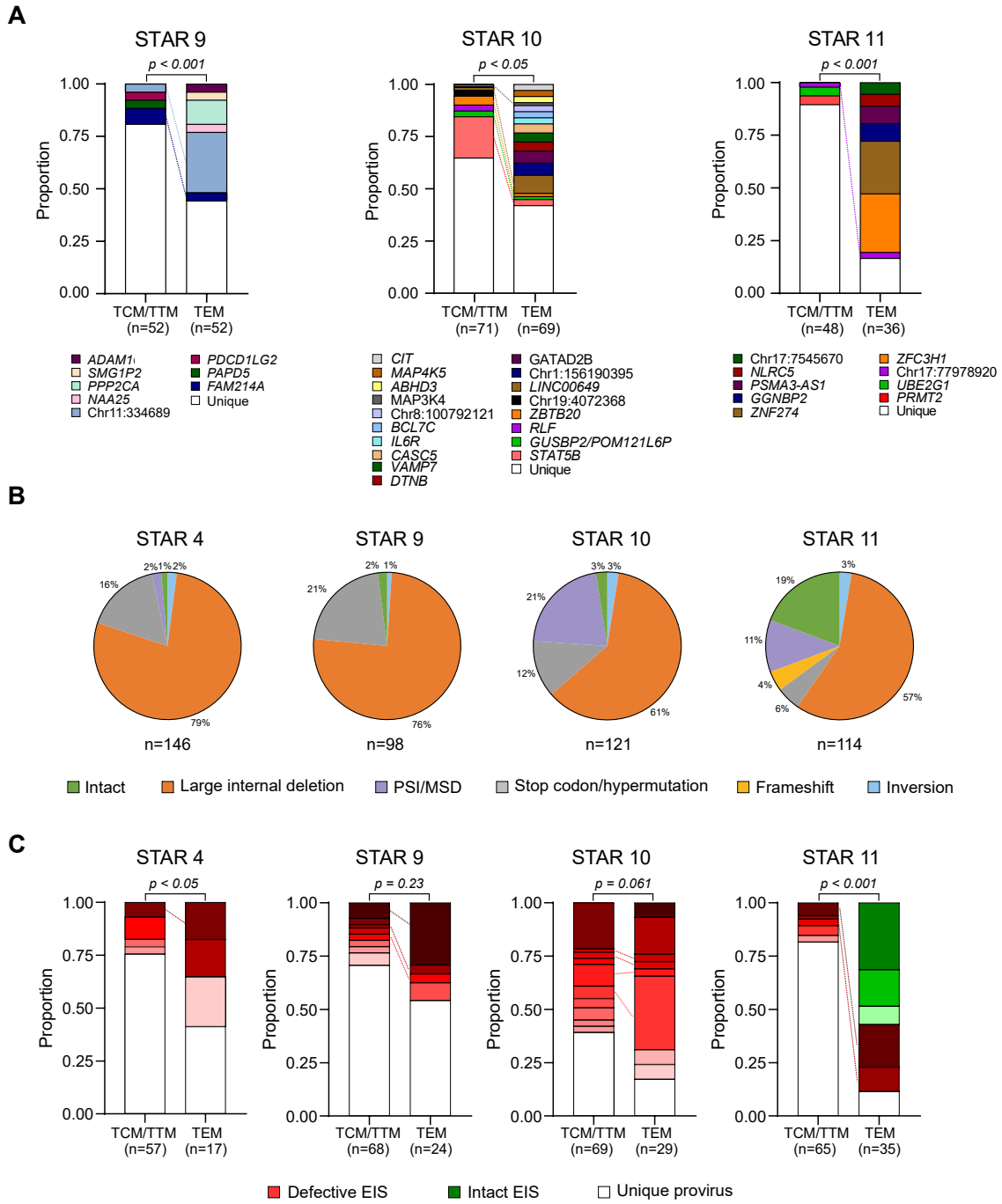


Figure 3

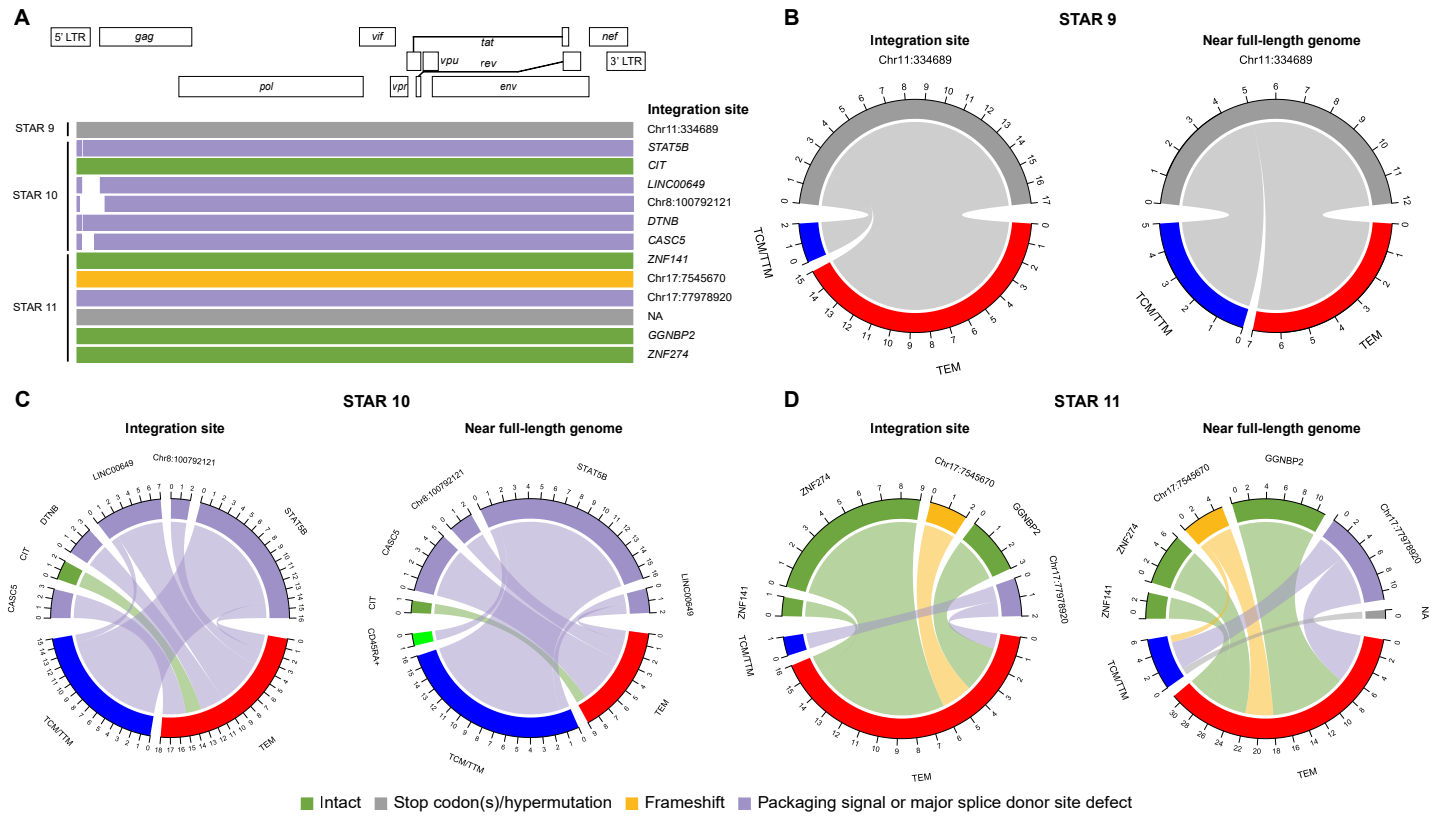


Figure 4

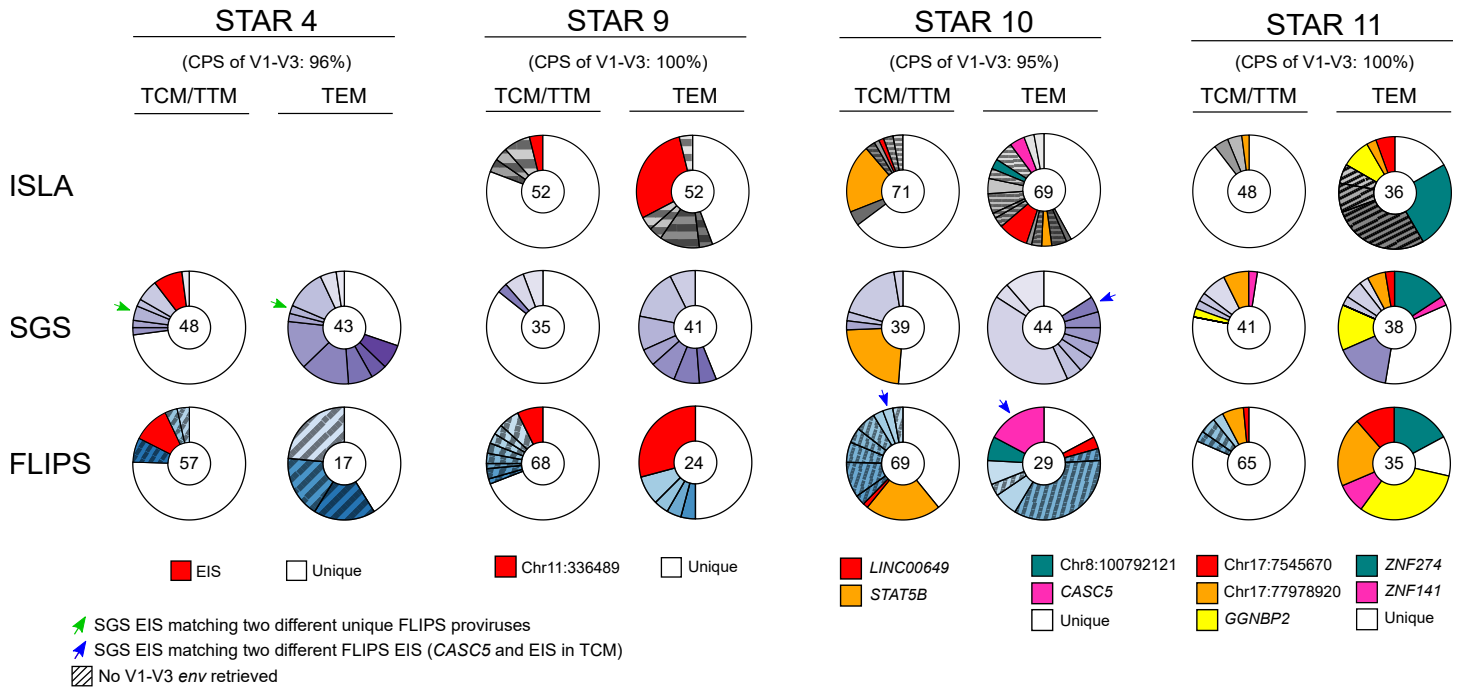


Figure 5

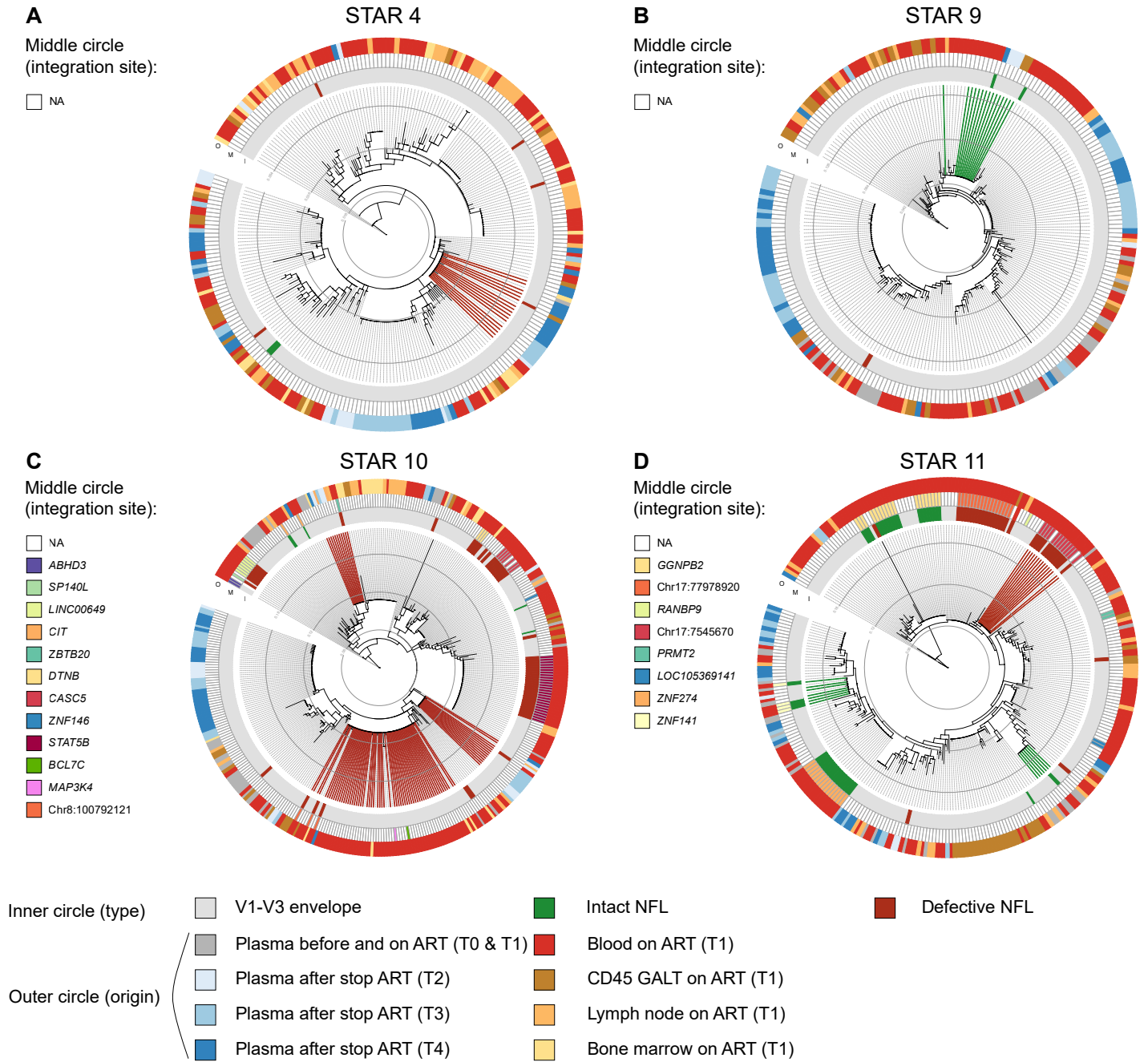


Figure 6

