

1 **Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses that are**
2 **Transcriptionally Active Before Stopping Antiretroviral Therapy**

3

4 Mary F. Kearney^{1*}, Ann Wiegand¹, Wei Shao², John M. Coffin³, John W. Mellors⁴, Michael Lederman⁵,
5 Rajesh T. Gandhi⁶, Brandon F. Keele², Jonathan Z. Li^{7*}

6

7 ¹HIV Dynamics and Replication Program, National Cancer Institute, Frederick, MD, USA. ²Leidos
8 Biomedical Research, Inc, Frederick National Laboratories for Cancer Research, Frederick, MD, USA.

9 ³Department of Molecular Biology and Microbiology, Tufts University, Boston MA. ⁴Department of
10 Medicine, University of Pittsburgh, Pittsburgh, PA. ⁵Case Western Reserve University, Cleveland, OH,
11 USA. ⁶Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. ⁷Brigham and
12 Women's Hospital, Harvard Medical School, Boston, MA, USA.

13

14 **Corresponding Authors***

15

16 Mary F. Kearney, PhD

17 HIV Drug Resistance Program

18 National Cancer Institute at Frederick

19 1050 Boyles Street, Building 535, Room 109

20 Frederick, Maryland 21702-1201

21 tel: 301 846 6796

22 fax: 301 846 6013

23 email: kearney@ncifcrf.gov

24

25 Jonathan Z. Li, MD
26 Brigham and Women's Hospital
27 Harvard Medical School
28 65 Landsdowne Street, Rm 421
29 Cambridge, MA 02139
30 Tel: 617-768-8476
31 Fax: 617-768-8738
32 Email: jli@bwh.harvard.edu

33
34

35 **Key words:** HIV, treatment interruption, monitored antiretroviral pause, ATI, HIV reservoir, active
36 reservoir, viral rebound, rebound viremia, expanded clones

37

38 **Running Title:** Predicting HIV Sequences that Arise after Analytic Treatment Interruption

39

40 **Conflicts of Interest:** JZL has received research support from Gilead Sciences and served as a consultant
41 for Quest Diagnostics. JWM is a consultant to Gilead Sciences and a shareholder of Co-Crystal, Inc. The
42 remaining authors have no potential conflicts.

43

44 **Financial Support:**

45 This work was supported by National Institutes of Health (NIH) grants AI100699 (to Dr. Li), U01AI068636
46 (to the AIDS Clinical Trials Group), UM1AI106701 (to the Pittsburgh Virology Specialty Laboratory, and by
47 Leidos Biomedical Research, Inc. (subcontract 12XS547 to JWM and 13SX110 to JMC through the
48 National Cancer Institute), and by intramural NIH funding including the National Cancer Institute, the

49 AIDS Targeted Antiviral Program Award to MFK, and NIH Contract No. HHSN261200800001E. JMC was a
50 Research Professor of the American Cancer Society, with support from the FM Kirby Foundation The
51 content is solely the responsibility of the authors and does not necessarily represent the official views of
52 the National Cancer Institute or the National Institutes of Health

53

54 **Presented in part:** Conference on Retroviruses and Opportunistic Infections, Seattle, WA, February 23-
55 26, 2015.

56

57 **ABSTRACT**

58 Understanding the origin of HIV variants during viral rebound may provide insight into the composition
59 of the HIV reservoir and has implications for the design of curative interventions. HIV single-genome
60 sequences were obtained from ten AIDS Clinical Trials Group participants who underwent analytic ART
61 interruption (ATI). Rebounding variants were compared with those in pre-ART plasma in all 10
62 participants and with on-ART PBMC-associated DNA and RNA (DNA and CA-RNA) in 7/10 participants.
63 The highest viral diversities were found in the DNA and CA-RNA populations. In 3 of 7 participants, we
64 detected multiple, identical DNA and CA-RNA sequences during suppression on ART that exactly
65 matched rebounding plasma HIV sequences. Hypermutated DNA and CA-RNA were detected in four
66 participants, contributing to diversities in these compartments that were higher than in the pre-ART and
67 post-ART plasma. Shifts in the viral rebound populations could be detected in some participants over the
68 2-3 month observation period. These findings suggest that a source of initial rebound viremia could be
69 populations of infected cells that clonally expanded either prior to and/or during ART, some of which
70 were already expressing HIV RNA before treatment was interrupted. These clonally-expanding
71 populations of HIV-infected cells may represent an important target for strategies aimed at achieving
72 reservoir reduction and sustained virologic remission.

73

74 **IMPORTANCE**

75 Antiretroviral therapy alone cannot eradicate the HIV reservoir and viral rebound is generally rapid after
76 treatment interruption. It has been suggested that clonal expansion of HIV-infected cells is an important
77 mechanism of HIV reservoir persistence, but the contribution of these clonally-proliferating cells to the
78 rebounding virus is unknown. We report a study of AIDS Clinical Trials Group participants who
79 underwent treatment interruption and compared rebounding plasma virus with that found within cells
80 prior to treatment interruption. We found several incidences where plasma HIV variants exactly
81 matched that of multiple proviral DNA copies from infected blood cells sampled before treatment
82 interruption. In addition, we found that these cells were not dormant, but were generating unspliced
83 RNA transcripts before treatment was interrupted. Identification of the HIV reservoir and determining
84 its mechanisms for persistence may aid in the development of strategies towards a cure for HIV.

85

86 **INTRODUCTION**

87 The discovery of strategies that target the HIV reservoir and induce sustained ART-free remission is one
88 of the highest priorities of the HIV research field. The evaluation of such therapeutic strategies will
89 require demonstration of effectiveness in analytic treatment interruption (ATI) studies, either through a
90 significant delay in the timing of viral rebound or a reduction in viral load set point. Our knowledge of
91 the source and diversity of rebounding HIV after ATI is still limited, but such information will be crucial
92 to design and evaluate interventions aimed at eliminating the HIV reservoir.

93 During suppressive antiretroviral therapy (ART), there is little to no evidence of active HIV
94 evolution in plasma viremia suggesting an absence of on-going viral replication in PBMCs during optimal
95 ART (1). However, ART alone is unable to eradicate infected cells and without life-long treatment, HIV
96 plasma viremia almost invariably rebounds (2, 3). One study of individuals undergoing multiple short
97 ATIs detected generally homogeneous populations of rebounding virus with different lineages present at
98 different ATI cycles, suggesting stochastic reactivation of mono- or small oligo-clonal populations of
99 latently infected cells (4). However, other studies have shown a relatively large number of rebounding
100 founder variants, suggesting that viral reactivation occurs from many latently infected cells, possibly
101 from multiple anatomic sites (5). There is evidence that virus after treatment interruption is similar to
102 that found in the pre-ART plasma quasispecies (1, 6), but the exact cellular origin of these rebounding
103 variants is still unclear, especially after prolonged suppressive ART (7, 8).

104 ART results in rapid 3-4 \log_{10} declines in levels of HIV plasma RNA but only ~10-fold declines in
105 the number of infected PBMCs (9). Clonal proliferation of HIV-infected cells may be a key contributor to
106 persistence of infected cells on ART (10, 11), however, it is not known if these cells are a source of viral
107 rebound after stopping ART. This uncertainty has been fueled by a study interpreted to mean that
108 clonally-expanded PBMCs contain only replication-incompetent proviruses (12). The evaluation of HIV
109 proviral DNA and actively transcribed intracellular HIV RNA may provide a better predictor of the

110 rebounding HIV variants, especially as the number of actively expressing HIV-infected cells has been
111 associated with the timing of viral rebound (13). Understanding the origin of HIV variants during early
112 and late rebound would provide insight into the composition of the HIV reservoir and has implications
113 for the design of curative interventions. In this study, we evaluated a cohort of ten participants who
114 initiated ART a median of 4 years earlier as part of an AIDS Clinical Trials Group (ACTG) study and then
115 enrolled in a subsequent ACTG study with an ATI component. These participants all had plasma samples
116 available both before ART initiation and shortly after the ATI. In addition, seven of these participants had
117 available PBMC samples shortly before the ATI. Using single-genome sequencing (SGS), we compared
118 the post-ATI rebounding virus to the plasma virus from pre-ART and to the on-ART PBMC-associated
119 DNA and RNA in an attempt to link infected cells and their transcriptional activity to rebound viremia
120 after stopping ART.

121

122 **MATERIALS AND METHODS**

123 **Study population**

124 Ten participants of prior ACTG ATI studies with available pre-ART and post-ATI plasma samples
125 were included. Seven of these participants had available on-ART PBMCs samples from shortly prior to
126 ATI and while virologically suppressed (HIV RNA <50 copies/mL). These participants initiated first-line
127 ART as part of either ACTG trial 384 (14, 15) or A5095 (16) and underwent an ATI in one of the following
128 four ACTG studies: A5024 (17), A5068 (18), A5170 (2), or A5197 (19). The pre-ART sample was collected
129 within 2 weeks prior to ART initiation and the duration of ART treatment prior to the ATI was a median
130 of 4 years (Table 1). The on-ART (pre-ATI) PBMC sample was collected immediately prior to the ATI with
131 the exception of one participant (PID 7), whose sample was collected 7 months prior to the ATI. The first
132 post-ATI viremic time point was a median of 4.5 weeks after ATI at a median viral load of 6,386 HIV RNA
133 copies/mL. The second post-ATI viremic time point was within 4-8 weeks of the first time point (median
134 12 weeks) at a median viral load of 18,128 HIV RNA copies/mL. Written informed consent was provided
135 by all study participants for use of stored samples in HIV-related research. This study was approved by
136 the Partners Institutional Review Board.

137

138 **Single-genome sequencing (SGS) and genetic analyses**

139 SGS of HIV-1 p6-PR-RT was performed with modifications to previously described methods (20, 21)
140 including the plasma viral RNA and total PBMC DNA and RNA being extracted using Qiagen kits,
141 according the manufacturer's protocol. RNA was reverse transcribed into cDNA using Superscript III
142 reverse transcriptase according to manufacturer's recommendations (Life Technologies). In brief, a
143 cDNA reaction of 1× RT buffer, 0.5 mM of each deoxynucleoside triphosphate, 5mM dithiothreitol,
144 2U/ml RNaseOUT (RNase inhibitor), 10U/ml of Superscript III reverse transcription, and 0.25mM
145 antisense primer 3500-R1 (5'-CTA TTA AGT ATT TTG ATG GGT CAT AA-3') was incubated at 50°C for 60

146 min, 55°C for 60 min then heat-inactivated at 70°C for 15 min followed by treatment with 1 U of RNase
147 H at 37°C for 20 min. DNA or newly synthesized cDNA was serially diluted prior to PCR such that the
148 majority of wells contain no template and the wells with template most likely contain only a single copy
149 (<25% PCR positive). Nested PCR was then performed with 1× PCR buffer, 2mM MgCl₂, 0.2mM of each
150 deoxynucleoside triphosphate, 0.2μM of each primer, and 0.025U/μl Platinum Taq polymerase (Life
151 Technologies) in a 20-μl reaction. First round PCR was performed with sense primer 1849-F1 5'-GAT GAC
152 AGC ATG TCA GGG AG-3' and antisense primer 3500-R1 under the following conditions: 1 cycle of 94°C
153 for 2 min, 35 cycles at 94°C for 15 sec, 55°C for 30 sec, and 72°C for 2 min, followed by a final extension
154 of 72°C for 10 min. Next, 1 μl from the first-round PCR product was added to a second-round PCR
155 reaction that included the sense primer 1870-F2 5'-GAG TTT TGG CTG AGG CAA TGA G-3' and antisense
156 primer 3410-R2 5'-CAG TTA GTG GTA TTA CTT CTG TTA GTG CTT-3' performed under the same
157 conditions used for first-round PCR, but with a total of 45 cycles. Correct sized amplicons were identified
158 by agarose gel electrophoresis and directly sequenced with BigDye Terminator based Sanger sequencing
159 (Life Technologies). To confirm PCR amplification from a single template, chromatograms were manually
160 examined for double peaks, indicative of the presence of amplicons resulting from PCR-generated
161 recombination events, Taq polymerase errors, or multiple templates. Sequences with evidence of mixed
162 bases were excluded from analysis. We sequenced ~20-50 SGS from each plasma and PBMC sample,
163 which provided sensitivity to detect variants present in 10% of the population with ≥90% certainty.

164 The resulting single-genome sequences were aligned using ClustalW. To ensure that there was
165 no cross-contamination of patient samples, we generated a NJ tree including all the sequences obtained
166 for the study and confirmed that all SGS clustered appropriately (Fig. 1). Population genetic diversity was
167 calculated as average pairwise difference (APD) using an in-house program (1, 22). In brief, the in-house
168 program tabulates the number of nucleotide difference between all possible pairs of aligned sequences,
169 divides the number of differences between each pair by the sequence length and sums the total. The

170 total is then divided by the number of sequence pairs, $N(N-1)/2$, where N is the number of sequences in
171 the population. Indels were not counted as differences between pairs of sequences. This program is
172 available by request. Shifts in population structure were calculated using a subdivision test for panmixia
173 with a significance cut off level of $p < 10^{-3}$ (23). The probability of 10^{-3} for assigning a significant change in
174 viral populations obtained from SGS was derived statistically taking into consideration the very large
175 numbers of comparisons across sequences (1, 23). Neighbor-joining phylogenetic analyses and
176 divergence calculations were done using MEGA6. Trees were rooted on the subtype B consensus
177 sequence (www.HIV-1.lanl.gov).

178

179 **Statistical analysis**

180 HIV-1 sequence diversity and divergence between time points were assessed by Spearman correlation
181 and the non-parametric Friedman test. Paired sample comparisons were performed by the Wilcoxon
182 signed-rank test. The Friedman test excluded the post-ATI plasma #2 time point as one of the
183 participants did not have a sample available.

184

185 RESULTS

186 Patients and study characteristics

187 Participants had a median pre-ART CD4 count of 474 cells/mm³ and a median on-ART CD4 count of 742
188 cells/mm³ at the time of ATI (Table 1). The median [Q1, Q3] plasma HIV RNA at the first and second
189 post-ATI time point were 6,386 [2,050, 33,165] and 18,128 [6,022, 22,616] HIV RNA copies/mL,
190 respectively. A median of 27 [24, 31] SGS were obtained from each sample type and time point (Table
191 2). There was no evidence of cross-participant sequence contamination in phylogenetic analysis (Fig. 1).

192

193 Intracellular and plasma HIV diversity before, during, and after ART

194 Viral diversity was evaluated within each sample by average pairwise distance (APD). The highest levels
195 of viral diversity were observed in the PBMC DNA populations (Fig. 2). Levels of DNA diversity (APD
196 1.5%) were significantly higher than those of the CA-RNA (1.3%, Wilcoxon signed rank P=0.03), pre-ART
197 plasma (1.1%, P=0.02), and post-ATI plasma (0.5%, P=0.02) populations. CA-RNA had the second highest
198 median levels of diversity and it was lowest in the post-ATI plasma (Fig. 2). Plasma HIV RNA diversity did
199 not significantly change over the 1-2 months between the first and second post-ATI time points (APD
200 post-ATI time point #1 and #2: 0.5% vs. 0.5%, Wilcoxon signed rank P=1.0) but was significantly higher in
201 pre-ART compared to post-ATI (1.1% vs. 0.5%, P=0.04). The lower diversity of the rebound virus
202 compared with pre-ART virus indicates that initial rebound virus represents only a subset of the virus
203 population pre-ART.

204

205 Divergence between pre-, during, and post-ART HIV populations

206 We measured the divergence of the initial post-ATI plasma HIV sequences as compared to viral
207 sequences from pre-ART, on-ART DNA and CA-RNA, and the second post-ATI plasma time point (Fig. 3).
208 The rebounding virus was overall more similar to pre-ART plasma HIV than to on-ART DNA or CA-RNA

209 HIV populations (Fig. 3). This is also supported by the finding that early post-ATI plasma HIV diversity
210 was closely associated with pre-ART plasma virus diversity (Spearman $r=0.72$, $P=0.02$), but not
211 significantly correlated with either on-ART DNA or CA-RNA viral diversity. These results are consistent
212 with a minor subset of HIV-infected cells being responsible for HIV viremia before ART, persisting
213 despite ART, and subsequently contributing to viral variants detected after ATI.

214

215 **Probable expanded clonal populations may be a source of rebound HIV**

216 We used phylogenetics to investigate the relationship between sequences detected in pre-ART, on-ART,
217 and post-ATI (Fig. 4). In three of seven participants (PID 1, 2, and 7, Fig. 4a, b, g), we detected multiple,
218 identical DNA and CA-RNA *gag-pol* sequences that *exactly* matched sequences from rebound plasma HIV
219 (indicated with red arrows), suggesting that this expanded population could be a source of rebound
220 viremia. These clonal plasma sequences represented 15-41% of all plasma sequences identified at the
221 first post-ATI viremic time point. In one of these cases (PID 2, Fig. 4b), the pre-ART plasma viral diversity
222 was low (0.24%) suggesting that this individual initiated ART in early infection and, therefore, the
223 detection of identical proviral sequences may not have been due to cellular proliferation but rather to
224 infection of multiple cells with the same variant. However, the pre-ART plasma virus populations in the
225 other two cases (PID 1 and 7, Fig. 4a, 4g) were more diverse (1.16% and 0.74%), (the typical *pol* APD in
226 chronically-infected individuals is $\sim 1.0\%$ and consists of all unique variants when sampled by SGS)
227 indicating that these individuals likely initiated treatment in chronic infection (22). This result suggests
228 that the identical proviral sequences in the PBMCs may have resulted from clonal expansion and not
229 from a single HIV variant infecting multiple cells prior to initiating ART. Importantly, in both cases,
230 matching CA-RNA sequences were found in the on-ART (pre-ATI) PBMC populations, suggesting that
231 expanded cell populations with identical proviruses may be the source of rebound viremia. PID 7 is the
232 best example of rebound from a presumably expanded, transcriptionally active cell population persisting

233 during ART (Fig. 4g, boxed sequences). However, most rebounding variants detected did not match DNA
234 or CA-RNA from pre-ATI suggesting that these arose from different reservoirs or tissues or that their
235 presence in blood was below the level of detection by SGS.

236

237 **Rebound viremia can result from cells that were infected prior to initiating ART**

238 In three cases (PID 2, 6, 9; Fig 4b, f, i) , rebound sequences were found to exactly match a pre-ART
239 plasma HIV variant (indicated with black arrows) and in one of these cases (PID 2), PBMCs were
240 available for analyses and matching DNA/RNA were also detected (Fig. 4b, black arrows). Two of these
241 participants had low levels of viral diversity prior to ART (PID 2 = 0.24%, PID 6 = 0.16%), suggesting that
242 ART was initiated during early infection. Although it has been shown previously that viral evolution is
243 inhibited by ART, halting viral diversification when ART is initiated early (1, 24), this study is the first to
244 report that at least some of these infected cells or their descendants, can continue to express HIV RNA
245 during treatment for at least 4 years. This finding provides evidence that some PBMCs infected prior to
246 initiating ART can express unspliced HIV RNA during ART, which can result in rebound viremia when ART
247 is interrupted.

248

249 **Hypermutated proviruses are expressed during ART**

250 We detected *both* hypermutated PBMC DNA and CA-RNA sequences in three of seven participants
251 where sequences were obtained from PBMCs during ART (Fig 4b, e, g). This result confirms findings by
252 others that cells infected with hypermutated proviruses persist for years on ART (25), but also shows
253 that at least some proviruses can be expressed in cells during treatment despite being hypermutated.
254 This finding is important because it explains, at least in part, the weakness of the association between
255 levels of HIV CA-RNA and residual plasma viremia during ART (26). Since G to A hypermutants are

256 defective (27), this finding also suggests that HIV proviruses that encode non-functional viruses due to
257 other mutations, insertions, or deletions may be expressed as well.

258

259 **HIV populations sometimes shift after treatment interruption**

260 Nine of the ten participants had plasma HIV sampled at a second post-ATI time point 1-2 months after
261 the first post-ATI time point. Although there were no significant differences in viral diversity between
262 the first and second post-ATI time points (Fig. 2) and the overall levels of divergence were also low (Fig.
263 3), in 3 individuals (PIDs 1, 3 and 4), there were significant population shifts between the rebounding
264 viral populations across the two time points that were detected by a test for panmixia ($p < 0.001$) and by
265 phylogenetic analyses (Fig. 4a, c, d indicated with blue arrows). Rebounding populations from the
266 temporally spaced plasma samples both contained highly homogenous, but distinct, subpopulations. In
267 all cases, the two sets of rebounding variants appear to have originated from different cellular sources
268 that may have been consecutively activated over a 4 week period (PID 1 and 3 shown as examples of this
269 in Fig. 4a, c). This observation suggests that latently infected cells persisting during ART could be
270 activated sequentially during or after stopping ART, leading to the rebound of multiple HIV variants that
271 continues for at least several weeks. However, one shifting rebound virus populations in PID 3 (Fig. 4c)
272 may have occurred by a different mechanism. In this case, one population of variants detected in the
273 second rebound time point (variants near the purple arrow) are present on the same main node but on
274 longer branches on the phylogenetic tree than variants from the first rebounding time point (green
275 arrow), making it a possibility that the variants in the later rebound sample evolved from those in the
276 initial rebounding population.

277

278 **DISCUSSION**

279 Although the HIV reservoir is often referred to as the pool of infected cells that persist in patients during
280 ART, it is now known that the vast majority of these cells carry defective proviruses (28) and, hence,
281 cannot be responsible for rebound viremia. As such, these cells are not a “true” reservoir for HIV. The
282 true reservoir is more accurately defined as those cells that carry replication-competent proviruses and
283 have the potential to result in rebound of HIV plasma RNA when treatment is interrupted. Such cells are
284 clearly only a tiny fraction of all of the infected cells that persist during therapy (12, 28). Clonal
285 expansion of HIV infected cells have been identified as a potential key mechanism in the maintenance
286 and persistence of the HIV reservoir (10, 11). While there is evidence that clonally-expanded populations
287 of HIV-infected cells may contribute to residual viremia during ART (29), it has also been claimed that
288 that much of the clonally-expanded cellular populations harbor replication-defective variants (12). It is
289 not known if clonally expanded populations within PBMCs that persist during ART are the source of
290 rebound viremia when treatment is interrupted and, therefore, represent one of the true reservoirs of
291 HIV. As the HIV field turns towards the search for successful curative strategies, it will be essential that
292 we identify this replication-competent reservoir of HIV and elucidate its mechanisms for renewal and
293 persistence.

294 To identify the sources of rebound viremia, we performed single-genome sequencing on
295 samples from 10 participants, including plasma from pre- and post-ART time points as well as cells
296 obtained immediately prior to rebound. We identified examples of identical *gag-pol* proviruses that
297 matched both intracellular HIV RNA during ART and plasma HIV RNA during rebound, suggesting that
298 clonally expanded, infected cells may be a source of rebound viremia, and hence, a true reservoir for
299 HIV. The *gag-pol* region that we selected is one that has previously been shown (30) to result in virtually
300 all unique sequences when sampled by SGS during chronic HIV infection and therefore, identical
301 sequences in this region (when detected during chronic infection) likely result from proliferation of

302 infected cells rather than from viral replication. However, it is important to note that while we show that
303 some HIV DNA and RNA sequences from PBMCs match variants that rebound after treatment
304 interruption, most of the viral rebound sequences did not have exact matches to the PBMC DNA or RNA.
305 This discrepancy could reflect limited sampling of infected PBMCs, viral evolution after treatment
306 interruption, or the presence of anatomic reservoirs of HIV that were not sampled but may also
307 contribute to rebound viremia like those described by Rothenberger, *et. al.* [5]. Despite this limited
308 sampling, the fact that we found *gag-pol* matches between multiple proviral variants and rebound
309 viremia suggests that these apparently proliferating cell populations are present in the blood at high
310 levels, consistent with a report from Wagner, *et. al.* (31). While the detection of identical HIV *gag-pol*
311 proviruses in this study suggests the presence of clonally-expanded cellular populations, we cannot
312 definitively rule-out the alternative explanation that such populations were created by the infection of
313 multiple cells with a homogeneous population of viruses, perhaps during acute infection. The fact that
314 the majority of participants were in the chronic phase of HIV infection pre-ART, as reflected by their
315 relatively high levels of viral diversity, would argue against the latter hypothesis. However, integration
316 site analyses that include their corresponding HIV *gag-pol* sequences need to be developed in order to
317 definitively show that these identical HIV sequences are derived from clonally-expanded cellular
318 populations and full-length HIV sequencing is needed to verify that these variants are not defective
319 elsewhere in the genome.

320 We also found examples of rebound virus identical to proviral sequences in cells that appeared
321 to express HIV RNA prior to interrupting treatment and possibly to proviruses that may have become
322 activated after treatment was interrupted, resulting in a shift in the plasma virus population over the
323 course of treatment interruption. Furthermore, we found that the intracellular HIV RNA populations
324 during ART were diverse, were not structurally different from the proviral populations (sequences
325 intermingled in phylogenetic trees) or the pretherapy virus, and most did not result in rebound viremia.

326 This finding, along with the detection of hypermutated HIV RNA sequences, implies that defective HIV
327 proviruses are expressed during ART and may explain previous reports on the weak association between
328 levels of HIV cellular and plasma RNA compartments (26), as most defective HIV proviruses would not
329 lead to successful viral protein production (28). This finding also highlights the uncertainty surrounding
330 the use of CA-RNA as the main efficacy outcome in the evaluation of latency-reversing agents (32, 33)
331 since a subset of the induced CA-RNA may be replication incompetent and subject to considerable
332 stochastic fluctuation from one patient to the next. These considerations highlight the limits of PCR-
333 based assays for HIV reservoir assessment. It is interesting that despite this finding, CA-RNA was found
334 to be predictive of time to virus rebound (13). The likely explanation for this apparent discrepancy lies in
335 the probability that the higher the fraction of cells that express HIV RNA during ART, the more likely that
336 the population will include a replication-competent HIV variant.

337 Overall, our results revealed several notable observations that need to be verified in larger
338 cohorts: 1) As suggested previously, rebound viremia appeared to result from only a small subset of the
339 total infected cells that persisted during treatment (28), 2) A subset of the infected cell populations that
340 likely resulted in rebound viremia, at least in some cases, appear to have been maintained through
341 clonal proliferation, 3) Some cell populations that carried proviruses matching plasma sequences in
342 rebound were already expressing HIV RNA *prior* to stopping ART and appear to have contributed to the
343 rebound viremia, 4) While viral diversity remained unchanged in the 2-3 months after treatment
344 interruption, shifts in viral population structure could be detected that suggests stochastic reactivation
345 of latently-infected cells and possibly, viral evolution after rebound, and 5) Replication-defective HIV
346 proviruses can be expressed and result in a diverse pool of CA-RNA variants, some containing
347 hypermutated sequences. The results of this study contribute to our understanding of the reservoir for
348 HIV during ART and lead to the hypothesis that the HIV reservoir includes populations of proliferating
349 cells, some of which are already expressing unspliced HIV RNA during ART. Larger studies, including host

350 integration site analyses, are critically needed to test the hypotheses that clonally-expanded cells are a
351 reservoir for HIV during ART.
352
353

354 **ACKNOWLEDGEMENTS**

355 We thank the participants, principal investigators, and staff of the ACTG studies 384 (Robert Shafer,
356 Gregory Robbins), A5095 (Roy Gulick, Cecilia Shikuma), A5024 (J. Michael Kilby, Ronald Mitsuyasu),
357 A5068 (Jeffrey Jacobson, Ian Frank, Michael Saag, Joseph Eron), A5170 (Daniel Skiest, David Margolis,
358 Diane Havlir), A5187 (Daniel Barouch, Eric Rosenberg, Daniel Kuritzkes), and A5197 (Robert Schooley,
359 Michael Lederman, Diane Havlir). We appreciate the assistance of Evgenia Aga and Ronald Bosch with
360 sample identification.

361

362

363

364 **FIGURE LEGENDS**

365

366 **FIG 1.** Neighbor-joining tree of all participant sequences showing no evidence of cross-participant
367 sequence contamination.

368

369 **FIG 2.** HIV sequence diversity measured by average pairwise distance in the pre-ART plasma, on-ART
370 PBMC, and post-ATI plasma. Post-ATI plasma #1 and #2 refer to the initial and subsequent sampling of
371 plasma HIV, respectively. Each symbol represents a different patient. The non-parametric Friedman test
372 excluded the post-ATI plasma #2 time point as one of the participants did not have a sample available at
373 that time point. Paired sample comparisons were performed with the Wilcoxon signed-rank test.

374

375 **FIG 3.** Sequence divergence compared to the initial post-ATI plasma time point. Each symbol represents
376 a different patient. Paired sample comparisons were performed by Wilcoxon signed-rank test.

377

378 **FIG 4.** Neighbor-joining trees from the ten participants. The dates for each participant time point is
379 labeled. For the post-ART plasma time points, the number of weeks after treatment discontinuation is
380 also indicated. The red arrows highlight the multiple identical on-ART CA-RNA and DNA sequences that
381 exactly matched post-ART plasma HIV sequences. Black arrows indicate exact matches between pre-ART
382 plasma sequences and either on-ART CA-RNA and DNA sequences, or post-ART plasma sequences. PID 3
383 (c) illustrates the shifts in divergent viral populations after HIV rebound and the possible evolution of
384 variants from a later rebound time point (purple arrows) from an early rebound viral population (green
385 arrows). The presence of G-to-A hypermutations in both the DNA and CA-RNA sequences were detected
386 and are labeled.

387

References

- 389
390
391 1. **Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, Rehm C, Poethke C, Kovacs N,**
392 **Mellors JW, Coffin JM, Maldarelli F.** 2014. Lack of detectable HIV-1 molecular evolution during
393 suppressive antiretroviral therapy. *PLoS Pathog* **10**:e1004010.
- 394 2. **Skiest DJ, Su Z, Havlir DV, Robertson KR, Coombs RW, Cain P, Peterson T, Krambrink A, Jahed**
395 **N, McMahon D, Margolis DM.** 2007. Interruption of antiretroviral treatment in HIV-infected
396 patients with preserved immune function is associated with a low rate of clinical progression: a
397 prospective study by AIDS Clinical Trials Group 5170. *J Infect Dis* **195**:1426-1436.
- 398 3. **Steingrover R, Pogany K, Fernandez Garcia E, Jurriaans S, Brinkman K, Schuitemaker H,**
399 **Miedema F, Lange JM, Prins JM.** 2008. HIV-1 viral rebound dynamics after a single treatment
400 interruption depends on time of initiation of highly active antiretroviral therapy. *AIDS* **22**:1583-
401 1588.
- 402 4. **Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Boni J, Hirschel B, Weber R, Trkola A, Gunthard**
403 **HF.** 2008. HIV rebounds from latently infected cells, rather than from continuing low-level
404 replication. *Proc Natl Acad Sci U S A* **105**:16725-16730.
- 405 5. **Rothenberger MK, Keele BF, Wietgreffe SW, Fletcher CV, Beilman GJ, Chipman JG, Khoruts A,**
406 **Estes JD, Anderson J, Callisto SP, Schmidt TE, Thorkelson A, Reilly C, Perkey K, Reimann TG,**
407 **Utay NS, Nganou Makamdop K, Stevenson M, Douek DC, Haase AT, Schacker TW.** 2015. Large
408 number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic
409 tissues after treatment interruption. *Proc Natl Acad Sci U S A*.
- 410 6. **Imamichi H, Crandall KA, Natarajan V, Jiang MK, Dewar RL, Berg S, Gaddam A, Bosche M,**
411 **Metcalfe JA, Davey RT, Jr., Lane HC.** 2001. Human immunodeficiency virus type 1 quasi species
412 that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral
413 quasi species present before initiation of therapy. *J Infect Dis* **183**:36-50.
- 414 7. **Chun TW, Davey RT, Jr., Ostrowski M, Shawn Justement J, Engel D, Mullins JI, Fauci AS.** 2000.
415 Relationship between pre-existing viral reservoirs and the re-emergence of plasma viremia after
416 discontinuation of highly active anti-retroviral therapy. *Nat Med* **6**:757-761.
- 417 8. **Lerner P, Guadalupe M, Donovan R, Hung J, Flamm J, Prindiville T, Sankaran-Walters S,**
418 **Syvanen M, Wong JK, George MD, Dandekar S.** 2011. The gut mucosal viral reservoir in HIV-
419 infected patients is not the major source of rebound plasma viremia following interruption of
420 highly active antiretroviral therapy. *J Virol* **85**:4772-4782.
- 421 9. **Besson GJ, Lalama CM, Bosch RJ, Gandhi RT, Bedison MA, Aga E, Riddler SA, McMahon DK,**
422 **Hong F, Mellors JW.** 2014. HIV-1 DNA Decay Dynamics in Blood During More Than a Decade of
423 Suppressive Antiretroviral Therapy. *Clin Infect Dis*.
- 424 10. **Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW,**
425 **Kearney MF, Coffin JM, Hughes SH.** 2014. HIV latency. Specific HIV integration sites are linked to
426 clonal expansion and persistence of infected cells. *Science* **345**:179-183.
- 427 11. **Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT,**
428 **Mullins JI, Frenkel LM.** 2014. Proliferation of cells with HIV integrated into cancer genes
429 contributes to persistent infection. *Science*.
- 430 12. **Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL,**
431 **McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M,**
432 **Nussenzweig MC.** 2015. HIV-1 Integration Landscape during Latent and Active Infection. *Cell*
433 **160**:420-432.
- 434 13. **Etamad B, Ahmed H, Aga E, Bosch RJ, Mellors J, Kuritzkes DR, Para M, Gandhi RT, Li JZ.**
435 February 23-26 (abstract 110LB). The Size of the Active HIV Reservoir Predicts Timing of Viral
436 Rebound, Conference on Retroviruses and Opportunistic Infections 2015, Seattle, WA.

- 437 14. **Robbins GK, De Gruttola V, Shafer RW, Smeaton LM, Snyder SW, Pettinelli C, Dube MP, Fischl**
438 **MA, Pollard RB, Delapenha R, Gedeon L, van der Horst C, Murphy RL, Becker MI, D'Aquila RT,**
439 **Vella S, Merigan TC, Hirsch MS.** 2003. Comparison of sequential three-drug regimens as initial
440 therapy for HIV-1 infection. *N Engl J Med* **349**:2293-2303.
- 441 15. **Shafer RW, Smeaton LM, Robbins GK, De Gruttola V, Snyder SW, D'Aquila RT, Johnson VA,**
442 **Morse GD, Nokta MA, Martinez AI, Gripshover BM, Kaul P, Haubrich R, Swingle M, McCarty**
443 **SD, Vella S, Hirsch MS, Merigan TC.** 2003. Comparison of four-drug regimens and pairs of
444 sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med* **349**:2304-
445 2315.
- 446 16. **Gulick RM, Ribaud HJ, Shikuma CM, Lalama C, Schackman BR, Meyer WA, 3rd, Acosta EP,**
447 **Schouten J, Squires KE, Pilcher CD, Murphy RL, Koletar SL, Carlson M, Reichman RC, Bastow B,**
448 **Klingman KL, Kuritzkes DR.** 2006. Three- vs four-drug antiretroviral regimens for the initial
449 treatment of HIV-1 infection: a randomized controlled trial. *JAMA* **296**:769-781.
- 450 17. **Kilby JM, Bucy RP, Mildvan D, Fischl M, Santana-Bagur J, Lennox J, Pilcher C, Zolopa A,**
451 **Lawrence J, Pollard RB, Habib RE, Sahner D, Fox L, Aga E, Bosch RJ, Mitsuyasu R.** 2006. A
452 randomized, partially blinded phase 2 trial of antiretroviral therapy, HIV-specific immunizations,
453 and interleukin-2 cycles to promote efficient control of viral replication (ACTG A5024). *J Infect*
454 *Dis* **194**:1672-1676.
- 455 18. **Jacobson JM, Pat Bucy R, Spritzler J, Saag MS, Eron JJ, Jr., Coombs RW, Wang R, Fox L, Johnson**
456 **VA, Cu-Uvin S, Cohn SE, Mildvan D, O'Neill D, Janik J, Purdue L, O'Connor DK, Vita CD, Frank I.**
457 2006. Evidence that intermittent structured treatment interruption, but not immunization with
458 ALVAC-HIV vCP1452, promotes host control of HIV replication: the results of AIDS Clinical Trials
459 Group 5068. *J Infect Dis* **194**:623-632.
- 460 19. **Schooley RT, Spritzler J, Wang H, Lederman MM, Havlir D, Kuritzkes DR, Pollard R, Battaglia C,**
461 **Robertson M, Mehrotra D, Casimiro D, Cox K, Schock B.** 2010. AIDS clinical trials group 5197: a
462 placebo-controlled trial of immunization of HIV-1-infected persons with a replication-deficient
463 adenovirus type 5 vaccine expressing the HIV-1 core protein. *J Infect Dis* **202**:705-716.
- 464 20. **Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT,**
465 **Jr., Dewar RL, Metcalf JA, Hammer S, Mellors JW, Coffin JM.** 2005. Multiple, linked human
466 immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are
467 missed by standard genotype analysis. *J Clin Microbiol* **43**:406-413.
- 468 21. **Kearney M, Palmer S, Maldarelli F, Shao W, Polis MA, Mican J, Rock-Kress D, Margolick JB,**
469 **Coffin JM, Mellors JW.** 2008. Frequent polymorphism at drug resistance sites in HIV-1 protease
470 and reverse transcriptase. *Aids* **22**:497-501.
- 471 22. **Kearney M, Maldarelli F, Shao W, Margolick JB, Daar ES, Mellors JW, Rao V, Coffin JM, Palmer**
472 **S.** 2009. Human immunodeficiency virus type 1 population genetics and adaptation in newly
473 infected individuals. *J Virol* **83**:2715-2727.
- 474 23. **Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, Wakeley J.** 2004. A robust
475 measure of HIV-1 population turnover within chronically infected individuals. *Mol Biol Evol*
476 **21**:1902-1912.
- 477 24. **Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, Epling L, Tan A, Ho**
478 **T, Lemey P, Shao W, Hunt PW, Somsouk M, Wylie W, Douek DC, Loeb L, Custer J, Hoh R, Poole**
479 **L, Deeks SG, Hecht F, Palmer S.** 2013. The HIV-1 reservoir in eight patients on long-term
480 suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc Natl Acad*
481 *Sci U S A* **110**:E4987-4996.
- 482 25. **Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF.** 2005. G->A hypermutation in
483 protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in
484 resting CD4+ T cells in vivo. *J Virol* **79**:1975-1980.

- 485 26. **Li JZ, Heisey A, Ahmed H, Wang H, Zheng L, Carrington M, Wrin T, Schooley RT, Lederman MM,**
486 **Kuritzkes DR.** 2014. Relationship of HIV reservoir characteristics with immune status and viral
487 rebound kinetics in an HIV therapeutic vaccine study. *AIDS*.
- 488 27. **Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D.** 2003. Broad antiretroviral defence
489 by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**:99-103.
- 490 28. **Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano**
491 **JD, Siliciano RF.** 2013. Replication-competent noninduced proviruses in the latent reservoir
492 increase barrier to HIV-1 cure. *Cell* **155**:540-551.
- 493 29. **Simonetti FR, Sobolowski MD, Hill S, Shao W, Fyne E, Wu X, Coffin J, Hughes S, Mellors JW,**
494 **Maldarelli F.** February 23-26 (abstract 105). Residual viremia caused by clonally expanded
495 tumor-infiltrating CD4+ cells, Conference on Retroviruses and Opportunistic Infections 2015,
496 Seattle, WA.
- 497 30. **Maldarelli F, Kearney M, Palmer S, Stephens R, Mican J, Polis MA, Davey RT, Kovacs J, Shao W,**
498 **Rock-Kress D, Metcalf JA, Rehm C, Greer SE, Lucey DL, Danley K, Alter H, Mellors JW, Coffin**
499 **JM.** 2013. HIV populations are large and accumulate high genetic diversity in a nonlinear
500 fashion. *J Virol* **87**:10313-10323.
- 501 31. **Wagner TA, McKernan JL, Tobin NH, Tapia KA, Mullins JI, Frenkel LM.** 2013. An increasing
502 proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests
503 proliferation of HIV-infected cells. *J Virol* **87**:1770-1778.
- 504 32. **Archin NM, Liberty AL, Kashuba AD, Choudhary SK, Kuruc JD, Crooks AM, Parker DC, Anderson**
505 **EM, Kearney MF, Strain MC, Richman DD, Hudgens MG, Bosch RJ, Coffin JM, Eron JJ, Hazuda**
506 **DJ, Margolis DM.** 2012. Administration of vorinostat disrupts HIV-1 latency in patients on
507 antiretroviral therapy. *Nature* **487**:482-485.
- 508 33. **Rasmussen RA, Tolstrup M, Brinkman CR, Olesen R, Erikstrup C, Solomon A, Winckelmann A,**
509 **Palmer S, Dinarello C, Buzon MJ, Lichterfeld M, Lewin SR, Ostergaard L, Sogaard OS.** 2014.
510 Panobinostat, a histone deacetylase inhibitor, for latent-virus reactivation in HIV-infected
511 patients on suppressive antiretroviral therapy: a phase 1/2, single group, clinical trial. *Lancet HIV*
512 **1**:e13-e21.
513
514
515
516

Table 1. Characteristics of study participants.

PID	Gender	Race	Pre-ART CD4 (cells/mm³)	Pre-ATI CD4 (cells/mm³)	Years on ART	Post-ATI #1 Week	Post-ATI #2 Week
1	Male	White	203	704	4.0	4	8
2	Male	White	416	543	1.5	8	12
3	Male	Hispanic	544	1075	4.9	5	8
4	Female	Hispanic	563	800	4.0	8	16
5	Male	White	541	1202	5.8	4	12
6	Male	White	447	749	4.4	4	12
7	Male	White	554	1355	6.1	4	
8	Female	Black	496	734	3.4	4	11
9	Female	Hispanic	451	591	4.0	12	16
10	Female	White	378	530	3.7	8	12
Median			474	742	4.0	4.5	12
[Q1, Q3]			[424-543]	[619-1006]	[3.9-4.5]	[4-8]	[11-12]

*On-ART (pre-ATI) PBMC sample was collected immediately prior to the ATI with the exception of one participant (PID 7), whose sample was collected 7 months prior to the ATI.

Table 2. Number of single-genome sequences obtained for each participant sample.

PID	Pre-ART Plasma	On-ART DNA	On-ART CA-RNA	Post-ATI Plasma #1	Post-ATI Plasma #2
1	24	28	34	53	23
2	28	19	29	29	24
3	25	26	8	27	25
4	28	38	8	26	26
5	13	53	33	4	25
6	28	-	-	42	27
7	33	81	31	42	-
8	24	-	-	24	36
9	26	-	-	21	17
10	25	30	27	25	29
Median [Q1, Q3]	26 [24, 28]	30 [27, 30]	29 [18, 29]	27 [24, 27]	25 [24-25]

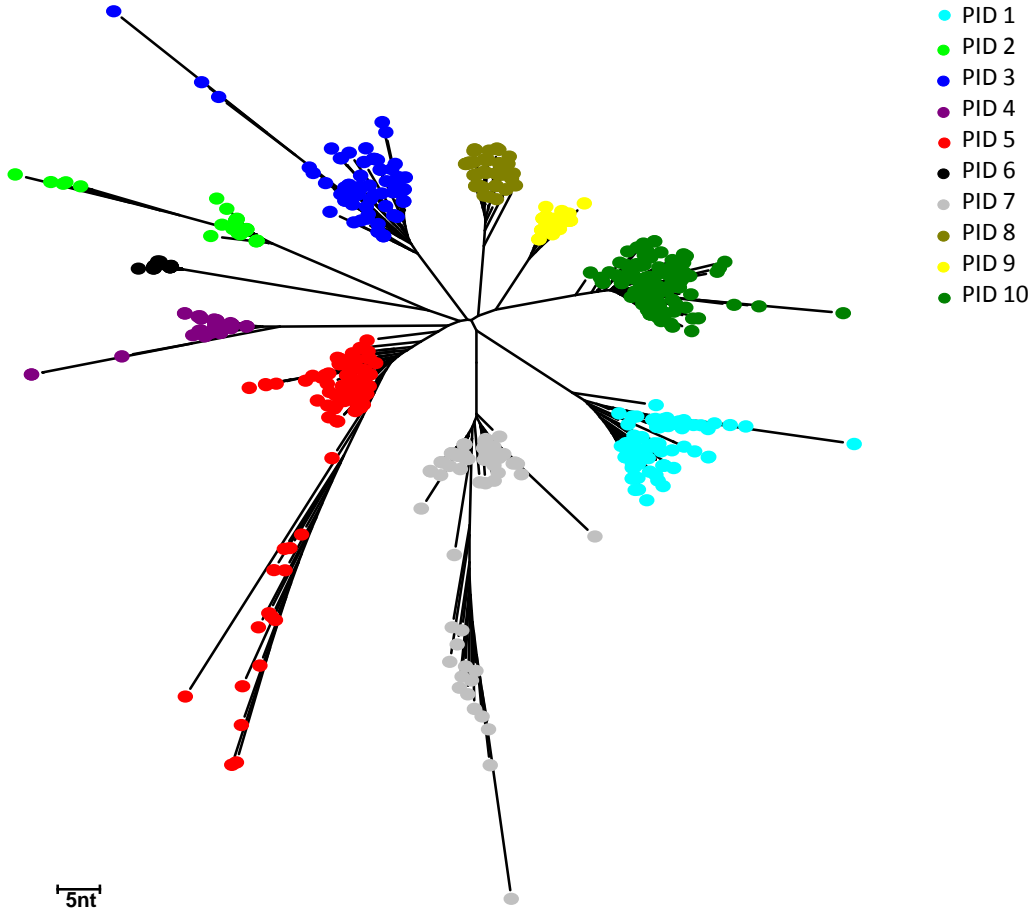


Figure 1.

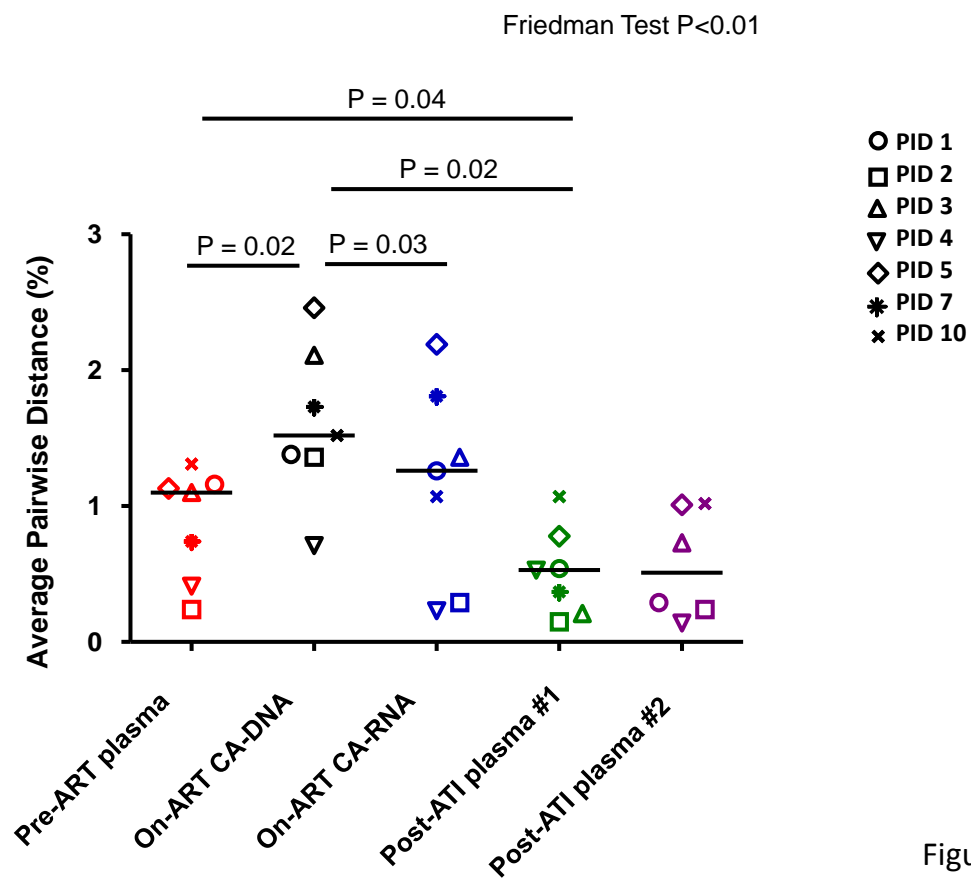


Figure 2.

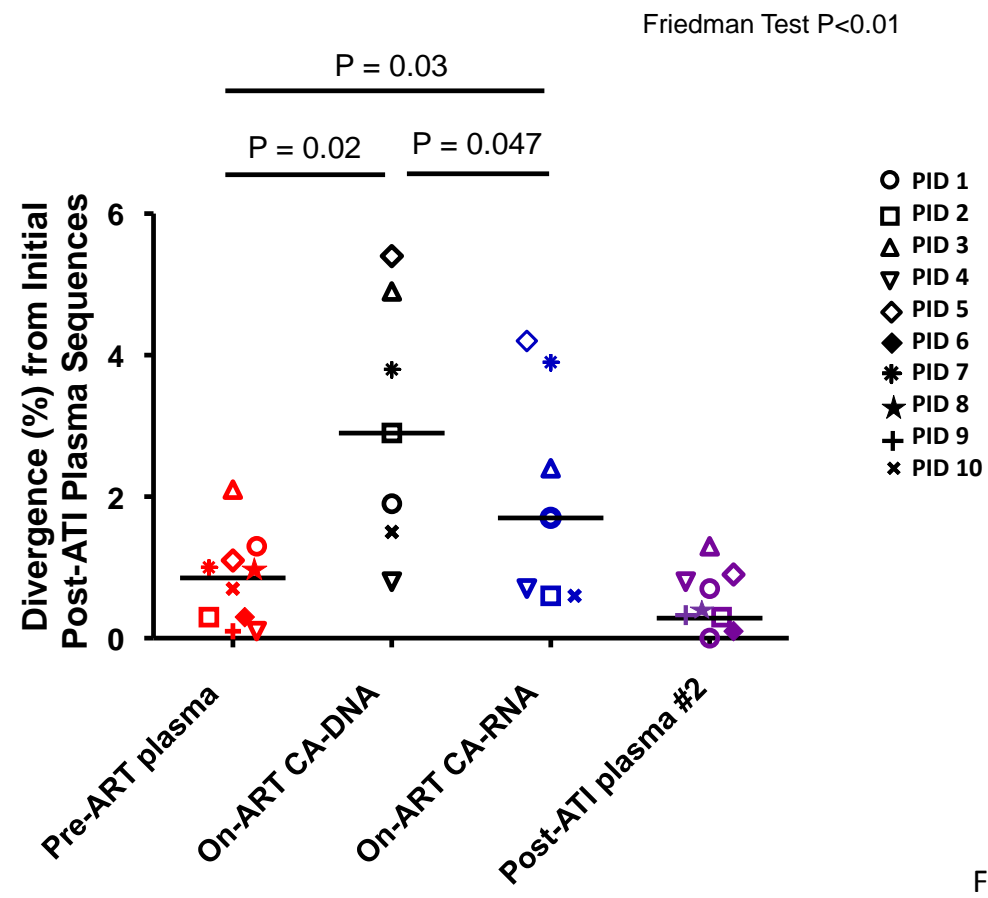


Figure 3.

PID 1

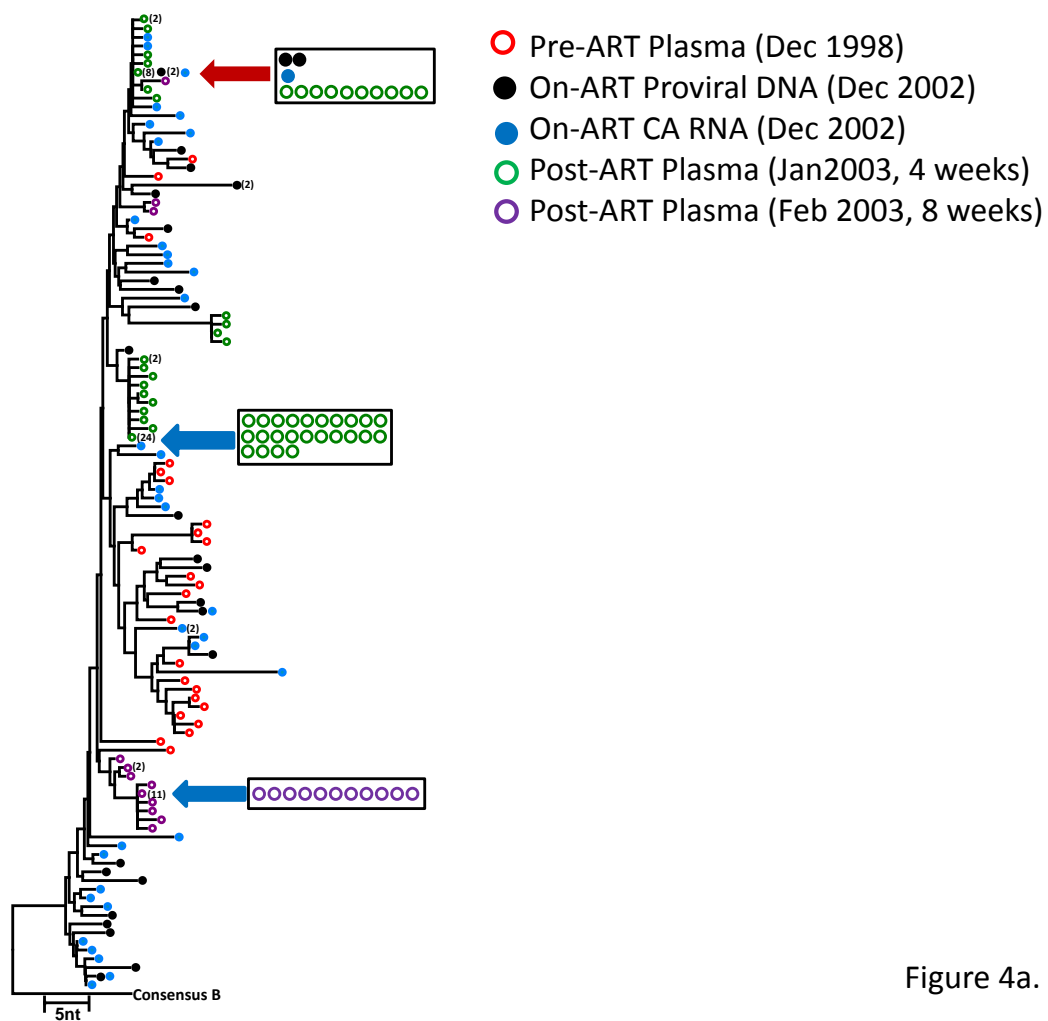


Figure 4a.

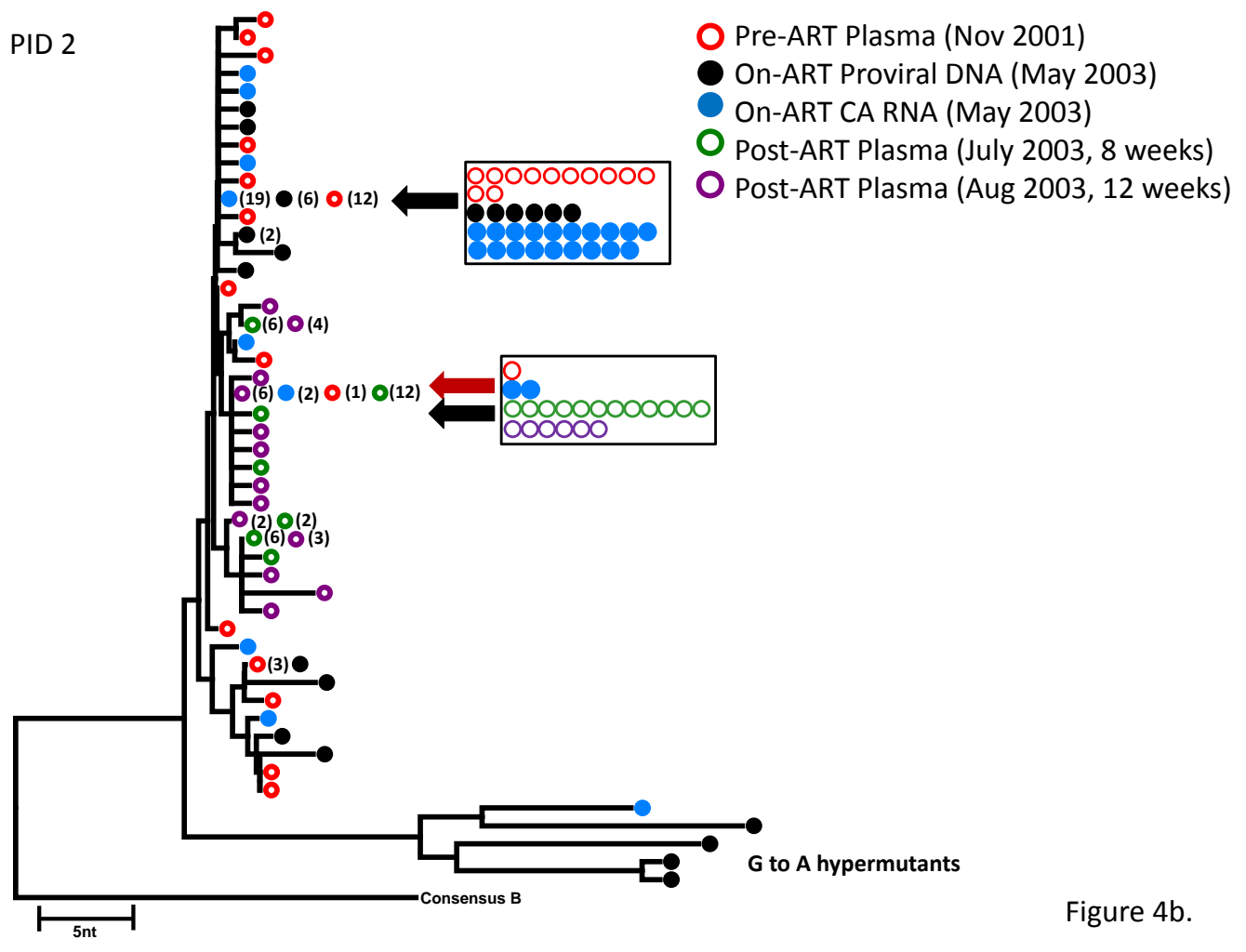


Figure 4b.

PID 3

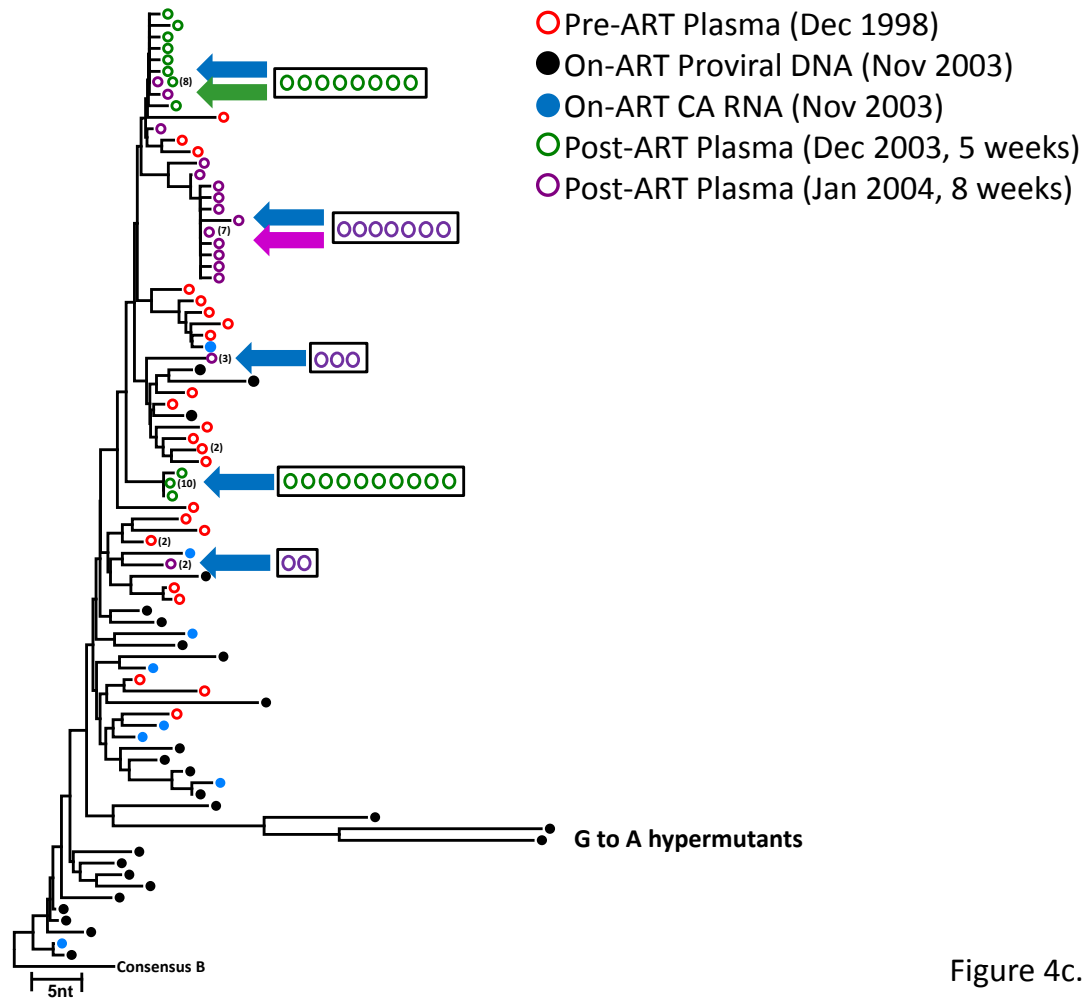


Figure 4c.

PID 4

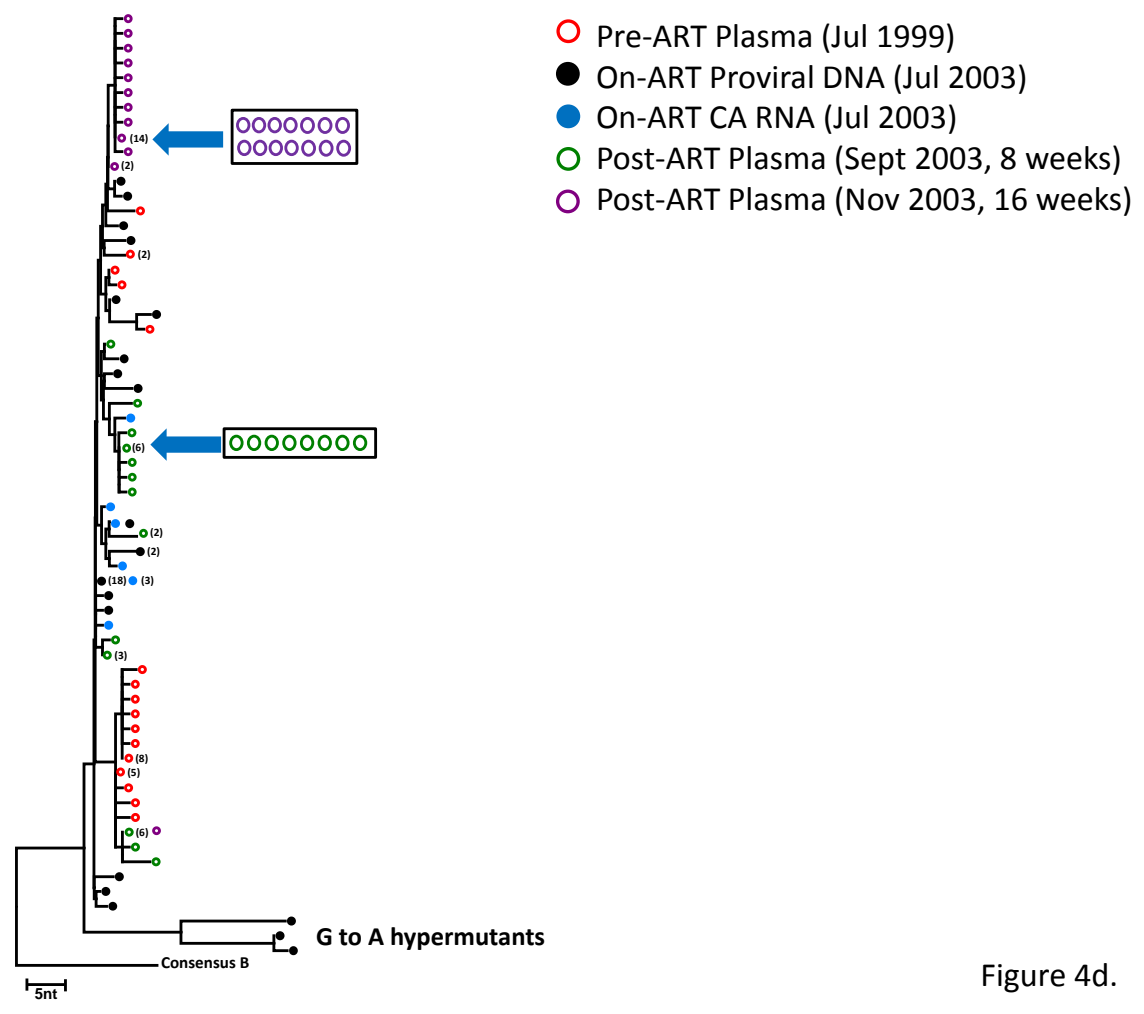


Figure 4d.

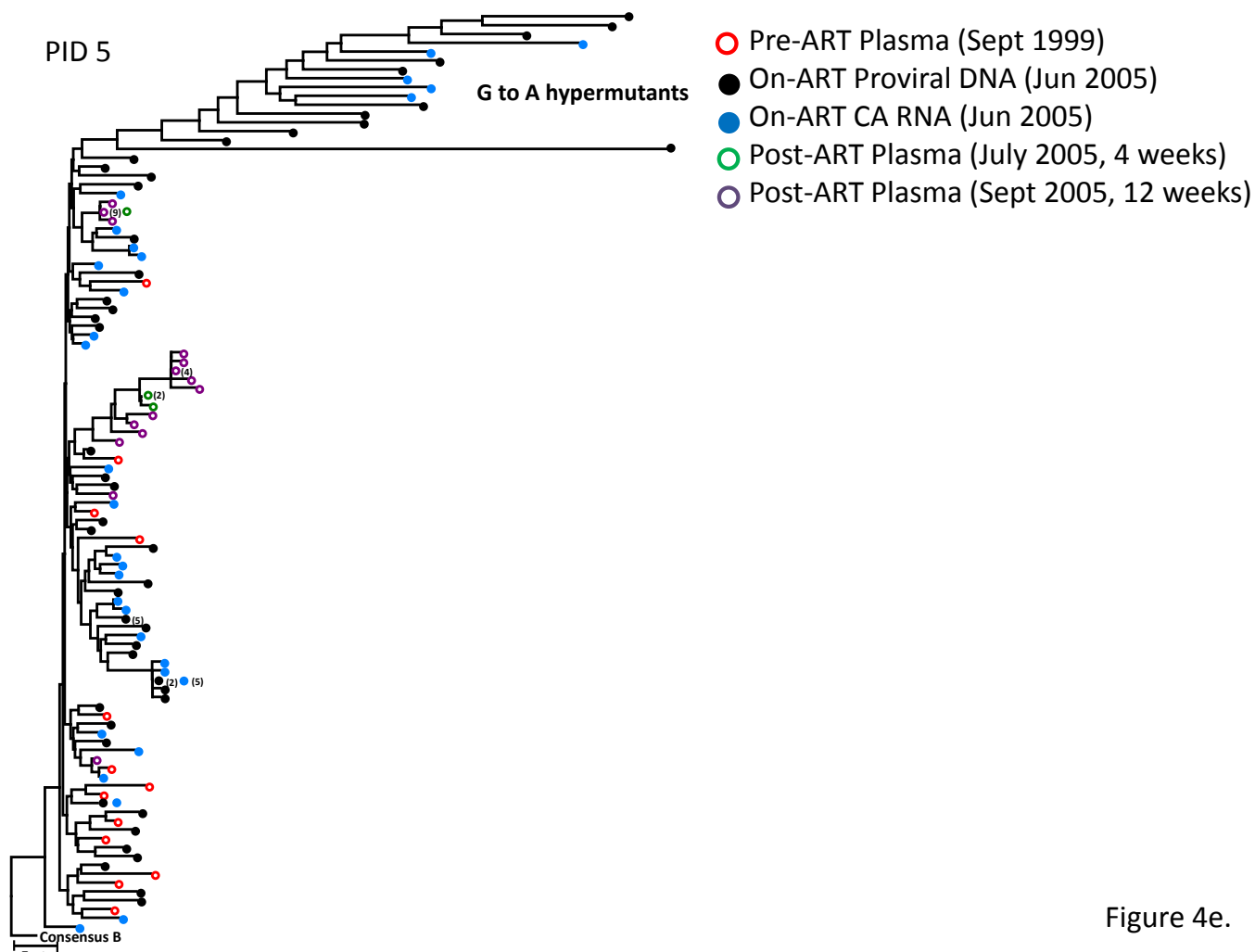


Figure 4e.

PID 6

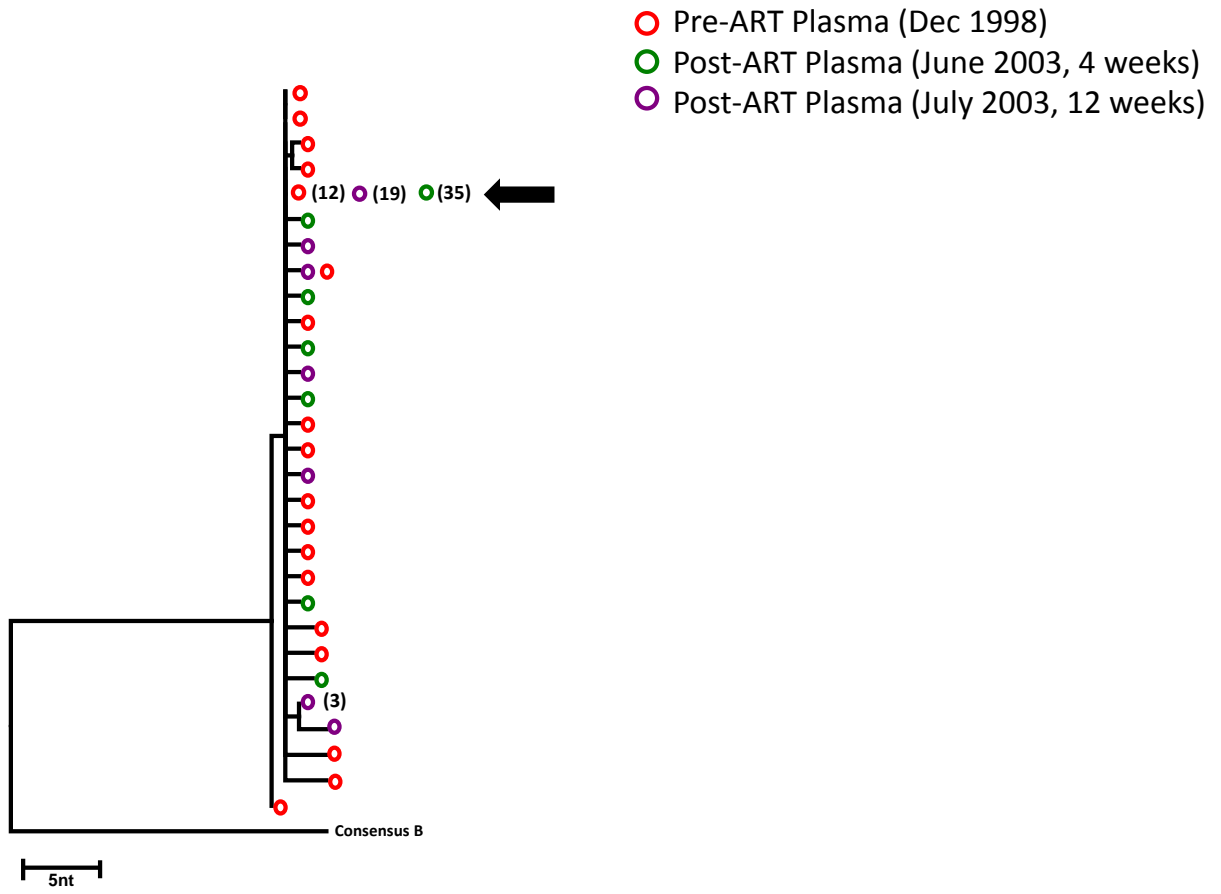
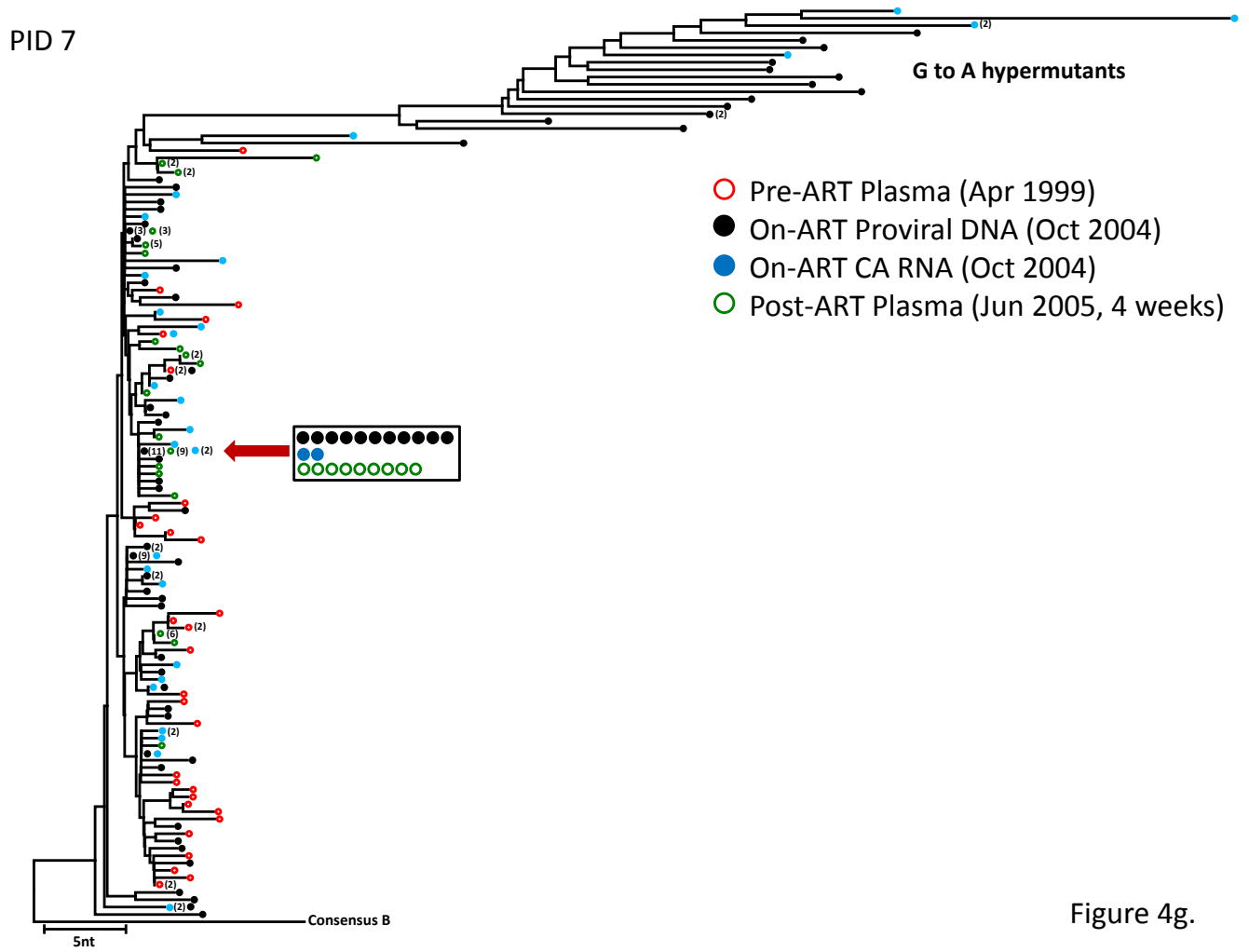
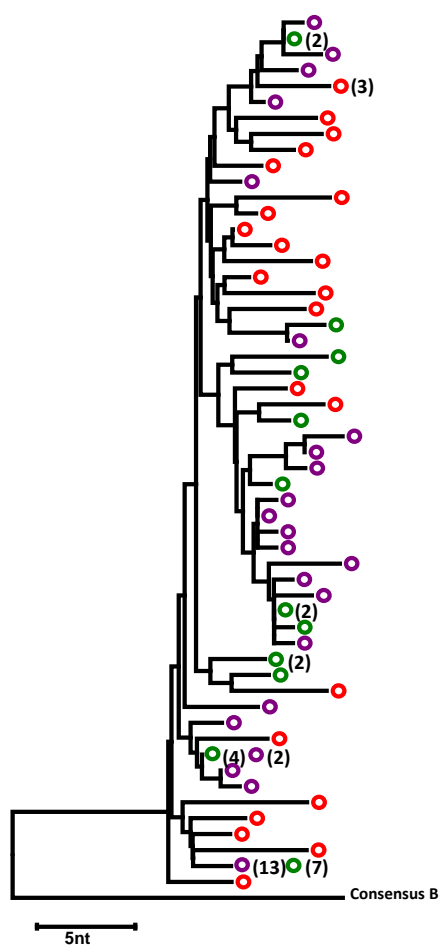


Figure 4f.



PID 8



- Pre-ART Plasma (Aug 1999)
- Post-ART Plasma (March 2003, 4 weeks)
- Post-ART Plasma (April 2003, 11 weeks)

Figure 4h.

PID 9

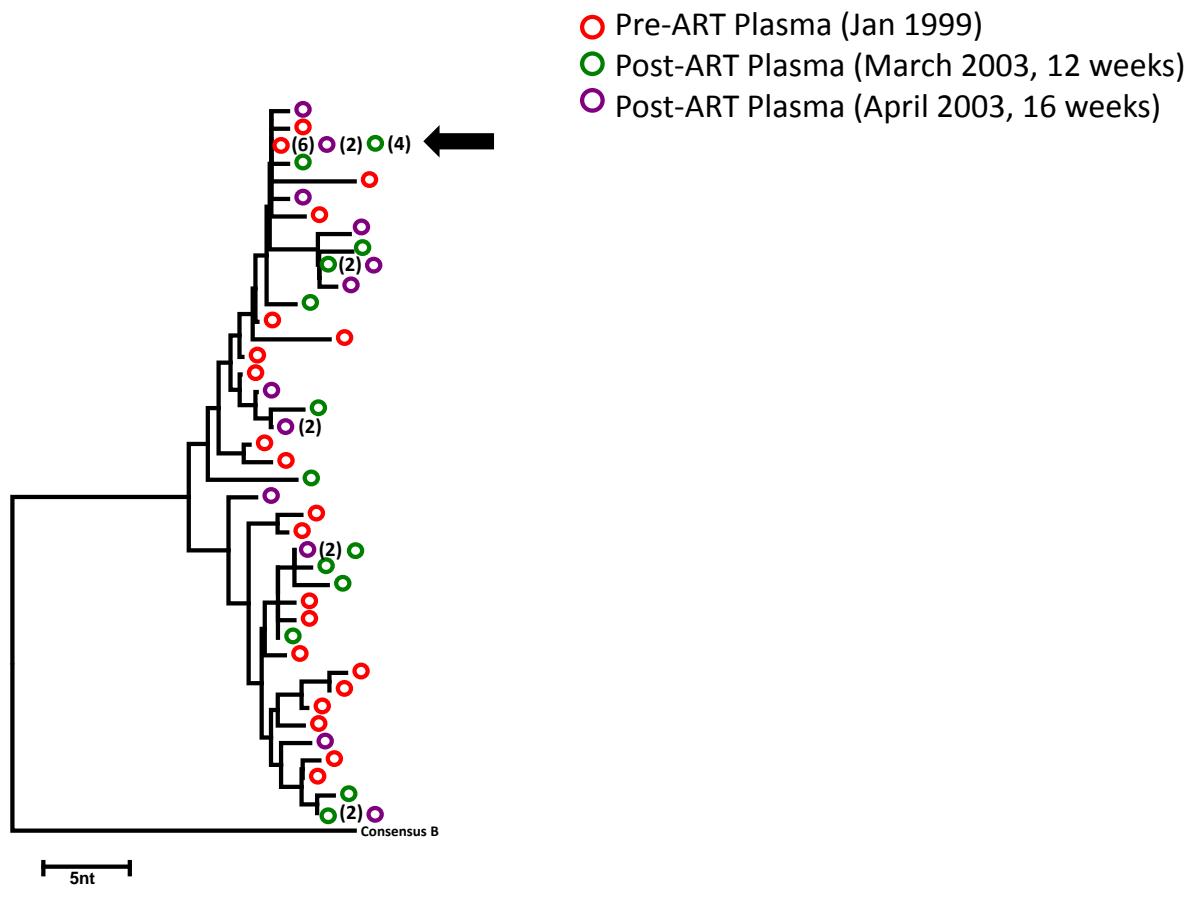


Figure 4i.

