1 Superinfection with intact HIV-1 results in conditional replication of defective proviruses

2 and nonsuppressible viremia in people living with HIV-1

3

4 **Authors**

5 Vivek Hariharan[1], Jennifer A. White[1], Filippo Dragoni[1], Emily J. Fray[1], Nicholas

6 Pathoulas[1], Milica Moskovljevic[1], Hao Zhang[2], Anushka Singhal[1], Jun Lai[1], Subul A. Beg[1],

7 Eileen P. Scully[1], Elizabeth A. Gilliams[1], David S. Block[1], Jeanne Keruly[1], Richard D.

8 Moore[1], Janet D. Siliciano[1], Francesco R. Simonetti[1], Robert F. Siliciano[1,3]

9

10 **Affiliations**

11 [1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD,

12 USA

13 [2]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg

14 School of Public Health, Baltimore, MD, USA

15 [3]Howard Hughes Medical Institute, Baltimore, MD, USA

16

17 Correspondence:

18 Francesco R. Simonetti, M.D, Ph.D.

19 Department of Medicine, Johns Hopkins University School of Medicine,

20 855 N Wolfe Street, 21205, Baltimore, MD, USA.

21 Phone: 401-955-7757

22 Email: fsimonetti@jhmi.edu

23

24 Robert F. Siliciano, M.D., Ph.D.

25 Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD,

26 USA

27 733 N Broadway, Baltimore, MD 21205

28 Howard Hughes Medical Institute, Baltimore, MD, USA

29 Phone: 410-955-2958

30 Email: rsiliciano@jhmi.edu

31

1

Abstract

During replication of some RNA viruses, defective particles can spontaneously arise and interfere with wild-type (WT) virus replication. Recently, engineered versions of these defective interfering particles (DIPs) have been proposed as an HIV-1 therapeutic. However, DIPs have yet to be reported in people with HIV-1 (PWH). Here, we find DIPs in PWH who have a rare, polyclonal form of non-suppressible viremia (NSV). While antiretroviral therapy (ART) rapidly reduces viremia to undetectable levels, some individuals experience sustained viremia due to virus production from cell clones harboring intact or defective proviruses. We characterized the source of NSV in two PWH who never reached undetectable viral load despite ART adherence. Remarkably, in each participant, we found a diverse set of defective viral genomes all sharing the same fatal deletions. We found that this paradoxical accumulation of mutations by viruses with fatal defects was driven by superinfection with intact viruses, resulting in mobilization of defective genomes and accumulation of additional mutations during untreated infection. We show that these defective proviruses interfere with WT virus replication, conditionally replicate, and, in one case, have an $R_0 > 1$, enabling *in vivo* spread. Despite this, clinical outcomes show no evidence of a beneficial effect of these DIPs.

2

Introduction

Antiretroviral therapy (ART) halts HIV-1 replication, reducing plasma HIV-1 RNA to below the limit of detection in most people with HIV-1 (PWH)[1–3]. However, ART is not curative due to the persistence of HIV-1 in a latent form in long-lived resting memory CD4$^+$ T cells[4–8]. This latent reservoir carrying infectious, replication-competent proviruses can cause viral rebound when ART is interrupted[9,10]. The reservoir decays very slowly (t$_{1/2}$ = 3.7 years)[11,12], necessitating lifelong ART. After many years on ART, the frequency of CD4$^+$ T cells with inducible, replication-competent virus may actually begin to increase[13], likely due to antigen-driven proliferation[14–16].

In most PWH on ART, more than 90% of proviruses are defective, principally due to large deletions and/or APOBEC3G/F-mediated G-to-A hypermutation[17–19]. Defective proviruses are generated throughout the course of untreated infection[20]. Missense mutations can arise due to the low fidelity of HIV-1 reverse transcriptase[21]. Some of these mutations are fatal, but others allow rapid evolution and escape from immune responses[22–26]. Proviruses with large internal deletions are common in PWH on ART. They result from the high propensity of HIV-1 to undergo recombination during reverse transcription through template switching events between the two viral genomes packaged in each virion[27]. Aberrant switching events commonly give rise to internal deletions[17,18,20]. Another major class of defective proviruses are those with G-to-A hypermutation introduced by APOBEC3G/F[17,18,28–30]. Hypermutation introduces multiple stop codons into most ORFs, causing loss of replication competence[18,30]. Proviruses with fatal defects are an evolutionary dead end; therefore, their sequences are unique and should not contribute to further evolution of intra-host viral lineages. However, they can persist and accumulate due to cell proliferation, resulting in cells with identical defective proviral sequences sharing the same integration site[31–36].

Some defective proviruses are unable to produce viral gene products due to internal deletions, stop codons, and APOBEC3-induced mutations which affect the transcription and translation of viral genes. However, a fraction of defective proviruses can express

HIV-1 RNA, and some can produce viral proteins and release virions[36–39]. We have recently shown that defective proviruses with deletions or point mutations in the 5'-leader region can cause non-suppressible viremia (NSV), a clinical phenomenon in which some PWH experience persistent low-level viremia[36]. NSV is not due to suboptimal adherence to ART or drug resistance leading to viral replication, but rather due to the release of virions from clonally expanded CD4+ T cells carrying infectious or defective proviruses[33,36,40–42]. The detection of low-level viremia on ART can also be an indication of virological failure, and thus understanding its underlying cause is both a research and clinical priority[43].

Engineered defective viral genomes have been proposed as an anti-viral strategy[44–46] because these viruses can impact the replication of wildtype (WT) virus through an interference phenomenon originally described by von Magnus for the influenza virus[47]. These interfering viruses, termed defective interfering particles (DIPs), are incapable of self-replication, but can specifically interfere with the replication of non-defective homologous virus in co-infected cells[48–50], thereby reducing cytotoxicity, viremia, and pathogenesis. Defective interfering genomes are defined by four biological properties: (i) they retain some genes encoding for structural proteins, (ii) they contain a portion of the viral genome, (iii) they can reproduce conditionally only with the WT virus and, (iv) they interfere with the replication of non-defective homologous virus[51]. DIPs have been observed to arise spontaneously for numerous RNA virus infections[52–55]. Despite the generation of defective proviruses during HIV-1 infection and extensive sequencing of HIV-1, HIV-1 DIPs have not been reported in PWH[46]. However, recent work from Pitchai, Tanner, and colleagues has shown that *in vitro* engineered versions of HIV-1 DIPs, therapeutic interfering particles (TIPs), can act as an antiviral therapeutic by reducing the HIV-1 viral load in humanized mice and non-human primate models[50].

Here, we demonstrate the presence of DIPs in PWH who have an unusual polyclonal form of NSV that prevents suppression of viremia to undetectable levels. We describe how DIPs contribute to viremia and consider the clinical implications.

4

Results

Persistent viremia despite long-term adherence to effective ART

The participants in this study were referred for evaluation due to sustained detectable viremia despite optimal adherence to ART.

Participant 1 (P1) is a Caucasian male diagnosed with HIV-1 in late 2019, presenting with plasma HIV-1 RNA of 1,240,000 copies/mL and a $CD4^+$ T cell count of 8 cells/µL. He was promptly treated with bictegravir, emtricitabine, and tenofovir alafenamide resulting in a 3-log reduction in HIV-1 RNA within 3 months (Fig. 1a). However, despite a significant increase in $CD4^+$ T cell count to 132 cells/µL, plasma HIV-1 RNA remained between 100 and 1,000 copies/mL for more than three years. Viral and proviral genotyping revealed only a G190A mutation in *pol*, which did not confer resistance to his treatment regimen. The treatment regimen was switched to emtricitabine, tenofovir alafenamide, darunavir (with cobicistat), plus dolutegravir (50 mg daily). Despite the change in treatment, HIV-1 RNA copies remained stable (100-500 copies/mL). Measurement of drug concentrations showed low trough levels for dolutegravir, prompting an increase to 50 mg dolutegravir twice daily. One month later, dolutegravir levels were within the normal range; however, the patient's viral load remained detectable (>50 copies HIV-1 RNA/mL plasma) for the duration of this study (Fig. 1a; 4.5 years).

Participant 2 (P2) is an African American female diagnosed with HIV-1 infection in 2003, with a viral load of 23,343 copies/mL and a $CD4^+$ T cell count of 301 cells/µL. P2 initiated ART with atazanavir, lamivudine, and tenofovir disoproxil fumarate one year after diagnosis, resulting in a temporary period of viral suppression (viral load below the limit of detection) (Fig. S1a). Despite multiple attempts at treatment optimization, P2 experienced intermittent adherence to therapy over the next several years, resulting initially in selection for drug resistance, followed by years of non-adherence with concomitant virological failure and an extended period of immunosuppression (Fig. S1a; CD4 nadir of 4 cells/µL). Following hospitalization due to opportunistic *Mycobacterium*

5

143   *intracellulare* infection with severe retroperitoneal lymphadenopathy, a new regimen of

144   darunavir, ritonavir, dolutegravir, tenofovir alafenamide, and emtricitabine was initiated

145   (Fig. 1b). Despite consistent adherence to the new regimen for more than nine years,

146   virologic suppression was not achieved during the duration of this study, a surprising

147   finding given that P2 achieved undetectable viremia with the first ART regimen of

148   atazanavir, lamivudine, and tenofovir disoproxil fumarate (Fig. S1a).

149

150   Both participants experienced recovery of their CD4 counts (peak CD4 count of 555 and

151   1,074 cells/μL, respectively) despite the low CD4 nadirs (Fig. 1a, b). Neither participant

152   has an HLA type known to be associated with elite control or rapid disease progression.

153   Since most PWH can achieve viral suppression (<50 copies HIV-1 RNA/mL plasma)

154   within 12 weeks of ART initiation[56–59], the persistence of detectable viremia after years of

155   treatment with optimal adherence represents a rare and troubling clinical outcome.

156

157   *Virus populations in plasma are diverse and show no evidence of evolution*

158

159   To investigate the cause of this non-suppressible viremia (NSV), we recovered

160   longitudinal single-genome sequences from plasma HIV-1 RNA during ART. Analysis of

161   the HIV-1 *pro-pol* region (1.2 kb amplicon, p6 of *gag* gene through the first 700 bp of

162   reverse transcriptase (RT) in *pol* gene) revealed that the plasma virus sequences were

163   highly heterogeneous (Fig. 1c, d). In P1, 135 out of 139 plasma p6-RT sequences were

164   unique; in P2, 35 out of 51 plasma p6-RT sequences were unique. Sequencing of the *env*

165   gene in P2 confirmed that the plasma virus was polyclonal (Fig. S1b). These results

166   contrast with previous reports of residual viremia following extended periods of

167   suppression which showed that the viremia is typically dominated by few predominant

168   plasma clones[33,36,40–42]. Plasma virus sequences from P1 and P2 lacked the temporal

169   structure characteristic of ongoing viral evolution[23]. Additional analyses showed no

170   accumulation of diversity, divergence, or shift in viral populations over time (Fig. S1c and

171   Methods). Notably, we found no drug-resistant mutations affecting the concurrent ART

172   regimen, a result confirmed through clinical genotyping. These results confirm that NSV

173   was not driven by ongoing replication or selection for drug resistance during ART.

6

Large pool of infected cells contributes to persistent viremia

Given the heterogeneity of viral populations in plasma, we hypothesized that consistently detectable viremia in the absence of ongoing replication might reflect the presence of an unusually high frequency of infected CD4+ T cells[60]. This hypothesis was supported by the intact proviral DNA assay (IPDA)[61], which showed that the frequency of proviruses classified as intact in P1 and P2 was markedly higher than previously reported in 81 PWH on ART (151 copies per 10^6 CD4+ T cells; Fig. 1e)[62]. The proportion of proviruses classified as intact by the IPDA at the final time point was 100% and 94% for P1 and P2, respectively – significantly higher than the 6.3% intact proviruses reported in previous studies of PWH on ART (Fig. 1e; P1: P = 2.39e-11; P2: P = 4.05e-10)[18,63]. We next performed quantitative viral outgrowth assays (QVOAs) to find the frequency of inducible, replication-competent proviruses in P1 and P2. Despite the high frequency of intact proviruses by IPDA (Fig. 1e), the frequencies of cells with inducible, replication-competent proviruses were comparable to previous studies: 3.18 infectious units per million (IUPM) CD4+ T cells for P1 and 1.44 IUPM for P2. (Fig. 1f)[11]. The discrepancy between the IPDA and QVOA values is highlighted by the ratio of IUPM to intact proviruses – a metric used to estimate reservoir inducibility[64] – which was much lower than previously reported in PWH on ART (Fig. 1g)[61]. These results suggest that a much smaller than normal fraction of the proviruses detected in the IPDA could be induced to produce infectious virus.

The vast majority of infected cells harbor proviruses with shared deletions in key HIV-1 genes

To explain the discrepancy between IPDA and QVOA reservoir measurements, we sequenced near-full length proviral genomes from P1 and P2. Proviruses belonged to subtype B, lacked drug resistance mutations to the concurrent ART regimen, and were CCR5-tropic based on co-receptor prediction algorithms[65]. However, in each participant, greater than 90% of the proviral sequences shared identical large deletions but, surprisingly, were divergent elsewhere in the genome (Fig. 2).

7

205

206   In P1, 50 out of 54 sequences contained the same set of two deletions of 1417 and 313

207   nucleotides in exactly the same positions. The 1417 nt and 313 nt deletions affected the

208   open reading frames of *vif, vpr, tat, rev, vpu,* and *env* (Fig. 2a, b). The 1417 nt deletion

209   eliminated *vpr* and *vpu* genes, the first exons of *tat* and *rev*, and the first 578 nucleotides

210   of the *env* gene, likely preventing expression of these proteins. Proviruses with these

211   deletions were classified as defective according to the NCI ProSeq-It tool[66]. The nature

212   of these deletions would almost certainly preclude further replication, yet the proviruses

213   harboring these exact two deletions were remarkably diverse elsewhere in the genome

214   (Fig. 2a). To assess the diversity within the proviral sequences, we employed two different

215   approaches. We first compared the number of nucleotide changes between each provirus

216   to a majority consensus sequence to estimate the extent of viral diversification. A

217   consensus sequence of all proviruses with the deletions provides a way to infer the

218   sequence of the provirus in the cell in which the deletions initially arose. The mean

219   distance of the proviral sequences to the majority consensus sequence was 81

220   nucleotides (Fig. 2c). Next, we measured the average pairwise distance to find the genetic

221   variation within the proviral population. The average pairwise distance of the unique near-

222   full length proviral sequences was 1.38% suggesting that, on average, there were 125

223   nucleotide differences between each unique sequence (Fig. 2d). Furthermore, some of

224   the proviruses with the recurring deletions also had signatures of APOBEC3G/F-mediated

225   hypermutation (Fig. 2a, dashed phylogenetic tree branches and green arrows), indicating

226   that viral RNA harboring these deletions entered a target cell where it underwent

227   hypermutation during reverse transcription. Only one proviral sequence matched the

228   plasma virus in the p6-RT regions, reflecting the extreme diversity and size of the infected

229   cells population.

230

231   In P2, we found a 270-nucleotide deletion impacting the reading frames of *vpr, tat, rev,*

232   and *vpu* in 64 out of 65 sequences (Fig. 2e, f). Given the critical roles of Tat and Rev in

233   the virus life cycle, it is expected that this deletion would preclude replication of the deleted

234   virus. Nevertheless, as with P1, proviruses with this deletion showed significant diversity

235   elsewhere in the viral genome. The mean distance from the majority consensus sequence

8

236 was 109 nucleotides (Fig. 2g). The average pairwise distance of the near full length
237 proviral sequences with this deletion was 1.81% (Fig. 2h), corresponding to an average
238 of 164 nucleotide differences among proviruses. The deletions found in proviruses from
239 P1 and P2 did not impact the binding of the IPDA primers and probes (Fig. 2a, e),
240 explaining the high frequency of proviruses classified as intact (Fig. 1e).

241

242 Identical sequences harboring fatal deletions typically indicate proliferation of CD4[+] T cell
243 clones carrying defective proviruses[31,33,40,67–69], since these deletions render the virus
244 unable to replicate[17,18,36]. However, analysis of the near-full length proviral sequences in
245 P1 and P2 suggests that the high diversity between each defective provirus cannot be
246 attributed to clonal expansion, errors made by human RNA polymerase II, or PCR and
247 sequencing errors. Instead, the finding of nucleotide differences in the viral genome that
248 carry exactly the same large fatal deletions implies that the defective viral RNA was
249 packaged into virions, entered a target cell, underwent reverse transcription leading to
250 additional HIV-1 RT-mediated mutations, and was then integrated into the host genome.

251

252 To specifically quantify proviruses harboring these deletions, we designed custom digital
253 PCR assays with probes spanning the deletion junction and primers bracketing the
254 deleted region (Fig. 3a). For P1, we designed a triplex assay that quantifies each deletion
255 separately in addition to the region upstream *gag* (LTR-gag). For P2, we designed a
256 competition assay that distinguishes proviruses with or without (WT-270) the 270 nt
257 deletion (Fig. 3a, b). Assay specificity was validated by testing each assay on genomic
258 DNA isolated from other PWH; only DNA from P1 and P2 yielded positive signals from
259 the deletion-specific PCRs (Fig. S2). We found that greater than 80% of proviruses
260 contained these deletions, and the proportion of proviruses containing the deletions
261 remained stable over time on ART (Fig. 3c). Taken together, these results show that most
262 infected cells contained these deletions, suggesting a unique situation that allowed their
263 accumulation during untreated infection and their persistence over time on ART.

264

265 *Proviruses with recurrent deletions can express viral RNA and contribute to persistent*
266 *viremia*

9

267

268   Given that some defective proviruses can express HIV-1 RNA and produce viral proteins

269   *in vivo*[36,38,39], we evaluated whether proviruses harboring these fatal deletions could be

270   induced to produce virions that package the defective viral RNA. We stimulated CD4[+] T

271   cells from P1 and P2 with anti-CD3/CD28 beads in the presence of antiretroviral drugs

272   (TDF, FTC, DTG) and quantified viral RNA in the supernatant using a digital RT-PCR

273   assay with the same primers and probes described above. In P1, we found that, on

274   average, 55% of virion-associated RNA contained either the 1417 nt or 313 nt deletion

275   (Fig. 3d). Due to shearing during the RNA isolation, it was not possible to quantify the

276   viral RNA that contained both deletions. Differences between the frequencies of the 1417

277   and 313 nt deletions can be explained by mutations surrounding the deletions which

278   prevent primer or probe binding or viral genomes which only had one of the deletions. In

279   P2, 13% of the virion-associated RNA in the supernatant contained the 270 nt deletion

280   (Fig. 3d).

281

282   Given that infected cells from both participants could be induced to make virions

283   packaging these defective RNAs, we determined if cells with these proviruses contributed

284   to the NSV. Using the same digital PCR assays, we found that in both participants, about

285   50% of virion-associated HIV-1 RNA in plasma harbored these deletions (Fig. 3e).

286

287   We also analyzed QVOA isolates to determine whether proviruses with these specific

288   deletions could be induced to produce infectious virus. Despite an abundance of cells

289   carrying defective proviruses seeded into each outgrowth well, the majority of isolates

290   that exponentially grew out in the culture were composed of virions packaging intact viral

291   genomes (Fig. S3). However, some wells with outgrowth also contained virions packaging

292   defective viral genomes in addition to virions with intact genomes, consistent with our

293   results from *ex vivo* stimulation of infected CD4[+] cells and analysis of plasma viremia.

294

295   Together, these results show that although the recurring deletions are incompatible with

296   replication, the viral RNA from the dominant deleted proviruses could still be packaged

297   into virus particles and contribute to persistent viremia.

298

299 *Molecular clones with these deletions show significantly reduced virion production*

300

301 To further investigate the impact of deletions on replicative fitness and virus production,

302 we introduced the recurring deletions into a reference proviral construct (NL4-3) (Fig.

303 4a)[70]. Next, we transfected HEK293T cells and measured virion production after 72 hours.

304 As expected, deletions in key HIV-1 genes resulted in an almost complete loss of virus

305 production (>100,000-fold reduction) compared to the wild type, as measured by p24

306 ELISA (Fig. 4b). We then stained the surface of the HEK293T cells with multiple broadly

307 neutralizing antibodies (bNAbs) to assess Envelope (Env) expression (Fig. 4C). Given

308 that the 1417 nt deletion impacts V1 and V2 of *env*, we tested bNAbs targeting V1/V2-

309 dependent and independent epitopes (Fig. S4). Consistent with the p24 ELISA results,

310 transfection with the defective molecular clone resulted in a dramatic reduction in the

311 percentage of HEK293T cells expressing Env, to levels that were not statistically different

312 from cells transfected with a construct lacking a functional *env* gene (NL4-3-ΔEnv, n.s. *p*

313 > 0.01; Fig. 4d).

314

315 These observations show that virion-associated RNA in the supernatant of activated

316 infected cells and in plasma contained the same deletions found in the proviral

317 sequencing. However, the deletions found *in vivo* resulted in abrogated virion production

318 when introduced into replication-competent molecular clones. To address this apparent

319 discrepancy, we investigated alternative mechanisms for the dissemination of these

320 proviruses.

321

322 *Dissemination of defective proviruses via superinfection*

323

324 We hypothesized that intact HIV-1 virions could superinfect cells carrying proviruses with

325 the recurrent deletions and drive the production of infectious viral particles carrying

326 defective genomic RNA. In superinfected cells, viral proteins can be produced from the

327 intact provirus, but genomic RNA from both defective and intact proviruses compete for

328 packaging into the virions[71]. This allows infectious particles that have packaged the

11

329 defective viral RNA to infect new cells, where the defective RNA genome undergoes
330 reverse transcription, accumulates additional mutations, and then integrates into the host
331 cell genome. In principle, the host cell could then be superinfected by another intact virus,
332 leading again to mobilization of the defective genome and accumulation of additional
333 mutations in the next round of infection. In this manner, defective proviruses carrying the
334 same fatal deletions could accumulate a diverse set of additional mutations. The
335 competition between replication-competent and defective viral genomes for packaging in
336 co-infected cells could result in viral interference, consistent with DIPs[44,50]. Therefore, we
337 hypothesized that viruses with the recurrent deletions could interfere with WT (intact) virus
338 replication. To assess this hypothesis, we generated NL4-3-based reporter construct
339 (NL4-3-ΔNef-BFP) which harbored the deletions found in each participant (P1:
340 Δ1417Δ313; P2: Δ270). We created model cell lines for each participant by transducing
341 SupT1 cells at low multiplicity of infection and bulk sorting BFP[+] cells for downstream
342 assays (Fig. S5a). To assess the ability of the deleted-variants to interfere with replication
343 of WT virus, we infected the model cell lines with an infectious WT reporter NL4-3 virus
344 carrying RFP in the *nef* ORF (NL4-3-ΔNef-RFP) and measured the frequency of cells
345 infected with the WT virus (RFP[+]) over time (Fig. 5a, b, Fig. S5b). The spread of WT virus
346 in cells carrying the defective proviruses was significantly reduced compared to WT virus
347 spreading in mock-transduced SupT1 cells, suggesting that these defective proviruses
348 interfere with WT virus replication (Fig. 5b). We quantified the infectivity of WT virus by
349 using the supernatant from day 3 of culture to infect new SupT1 cells in a single-round
350 infectivity assay (Fig. 5c)[72]. WT virus infectivity was significantly reduced when using the
351 supernatant from the deleted-variant cell lines compared with the mock-transduced cell
352 line, further supporting the conclusion that the defective proviruses interfere with WT virus
353 replication (Fig. 5c). WT viral titers were similarly reduced in both model cell lines carrying
354 the defects observed in P1 and P2 (Fig. 5c). To test whether the superinfected cells
355 produced infectious virions packaging either the intact or defective RNA, we quantified
356 the percentage of RFP[+] (WT virus) or BFP[+] (defective virus) SupT1 cells after a single
357 round of infection. When infected with supernatant from the superinfected cells, >40% of
358 target cells were infected with the defective virus (BFP[+]) (Δ1417Δ313: 43%; Δ270: 62%)
359 (Fig. 5d).

12

360

361　Next, to validate that the defective proviruses conditionally replicate in the presence of a

362　WT virus, we collected supernatants from either the deleted-variant cell lines alone or

363　from the deleted-variant cell lines which had been infected with WT virus for 3 days. These

364　supernatants were used to infect SupT1 cells. We then assessed the target SupT1 cells

365　for the presence of defective proviruses (BFP$^+$) (Fig. 5e). For the cell line representing

366　P1, the frequency of defective proviruses in the target cells increased by >170-fold when

367　using viral inoculum from superinfected cells, compared to inoculum from cells which

368　were not superinfected with WT virus. For the cell line representing P2, the frequency of

369　defective proviruses increased 3.69-fold when target cells were infected with viral

370　inoculum from cells that were superinfected, compared to inoculum from cells which were

371　not superinfected (Fig. 5e). These data show that these defective proviruses transmit their

372　genome to target cells when superinfected with WT virus. Together, these results

373　demonstrate that the recurrent proviral deletions found in P1 and P2 interfere with WT

374　virus replication and conditionally replicate only in the presence of WT virus, two

375　hallmarks of DIPs[73].

376

377　Given that the deletions found in P1 (Δ1417Δ313) strongly resembled the engineered

378　deletions which constitute the therapeutic interfering particle (TIP) described in Pitchai,

379　Tanner et al. (e.g., deletion of *tat*, *rev*, *vpu*, *env*, and retention of cPPT) (Fig. S5c), and

380　considering the ability for viral genomes harboring this deletion to integrate into target

381　cells (Fig. 5d), we hypothesized that the defective proviruses found in P1 functioned as a

382　TIP[50]. For a DIP to be considered a TIP, it must be mobilized with an $R_0 > 1$, meaning that

383　the defective provirus can give rise to virions that infect more than one target cell and

384　therefore spread *in vivo*. To directly measure the R_0 of the defective variants (R_0^{DEF}), we

385　implemented a three-color assay (Fig. 5h)[50]. First, we infected the Δ1417Δ313-BFP

386　reporter cell line with the WT-RFP virus as in Figure 5b. After 2 days, we cocultured

387　30,000 infected cells with 70,000 GFP$^+$ SupT1 target cells (see Methods for generation

388　of the GFP$^+$ target cell line). After 3 days of co-culture, we assessed the fraction of GFP$^+$

389　target cells with either the WT virus (RFP$^+$) or the defective virus (BFP$^+$), allowing us to

390　estimate the frequency of secondary infection events (Fig. S5d). The R_0 was found by

13

391   dividing the number of secondary infections (i.e., infected target cells) by the number of

392   primary infections (i.e., initial infection events) (see Methods). The R_0^{WT} following infection

393   of wildtype SupT1 cells with WT virus alone was 6.40, consistent with lower estimates in

394   previous reports[74–76]. As expected, the R_0^{WT} in the Δ1417Δ313 reporter cell line (R_0^{WT}:

395   0.84) was significantly less than the R_0^{WT} in WT SupT1 cells, consistent with viral

396   interference (Fig. 5b; $p = 0.0002$). The R_0^{DEF} represents the average number of target

397   cells infected with the defective virus resulting from WT virus superinfection of cells

398   carrying the deleted provirus. Notably, the R_0^{DEF} in the Δ1417Δ313 reporter cell line was

399   1.86 (Fig. 5h). This result is consistent with the spread of the defective viruses *in vivo*.

400   The low R_0 of the defective lineage suggests that although viral interference can be

401   detected *in vitro* (Fig. 5b), the interference does not affect WT viral replication and

402   pathogenesis *in vivo* to a clinically significant extent, as evidenced by P1's high initial viral

403   load and extremely low CD4 nadir at the time of ART initiation.

404

405   *Superinfected CD4$^+$ T cells carrying both intact and defective proviruses can be detected*

406   *ex vivo.*

407

408   To show that defective proviruses found in P1 and P2 were mobilized – before the

409   introduction of effective ART – by superinfection with intact proviruses, we looked for cells

410   carrying multiple proviruses in primary CD4$^+$ T cells from P1 and P2. We modified the

411   HIV-Flow assay[77] to single-cell sort p24$^+$ cells into individual wells and lysed the cells (Fig.

412   6a). Similar to previous work investigating the number of proviruses within individual

413   infected cells[78], we next distributed the cell lysate over 6 or 12 wells, resulting in a greater

414   than 80% probability that, if two proviruses were present, they would be separated into

415   different wells (Fig. 6a, Fig. S6a). Lastly, we amplified and sequenced the near-full length

416   proviral genomes[79]. Given the importance of sorting accuracy in this experiment, we first

417   sorted individual CD4$^+$ T cells from an HIV-1 negative male donor and confirmed that all

418   sorted cells were singlets by amplification of the *SRY* gene, which is found in a single

419   copy per cell on the Y chromosome (Fig. S6b; methods). We also validated this

420   experimental protocol by sorting p24$^+$ ACH-2 cells. In accordance with other reports, we

421   found that ACH-2's can harbor multiple proviruses per cell when cultured in the absence

14

422   of antiretroviral drugs (Fig. S6c)[80–82]. We further validated this protocol using cells from a

423   participant with low-level viremia: out of 5 sorted p24+ cells, none had evidence of

424   containing more than one provirus, consistent with previous findings[78,83,84] (Fig. S6d;

425   methods). In a recent study, Dufour and colleagues did not find any evidence of cells

426   infected with multiple proviruses from 305 single-sorted p24+ cells from 6 PWH on

427   suppressive ART[79] (Nicolas Chomont, Caroline Dufour, personal communication).

428

429   For P1, we analyzed ~13 million CD4+ T cells and recovered 7 sequences from the p24+

430   sorted cells (0.54 proviruses per 10^6 CD4+ T cells) (Fig. 6b). From two of the p24+ sorted

431   cells, we identified proviruses that had small deletions in the 5' region of the viral genome

432   – which are disproportionally found in p24+ cells[79,85,86] – yet did not have the 1417 nt and

433   313 nt deletions found in the bulk proviral population (Fig. 2a). From another two of the

434   p24+ sorted cells, we identified intact proviruses. Notably, we identified one cell which had

435   two integrated proviruses: one provirus with the 1417 nt and 313 nt deletions and one

436   nearly intact provirus with a small 5' deletion, demonstrating the presence of

437   superinfected cells *in vivo* (Fig. 6b, color boxed). Lastly, from another sorted cell, we

438   identified one provirus harboring both the 1417 nt and 313 nt deletions, a surprising

439   finding given that these two deletions decrease p24 production by several orders of

440   magnitude *in vitro* (Fig. 4b). Given the low efficiency of near-full length proviral

441   amplification[87], it is possible that we detected only one of the two proviruses in that cell.

442   For P2, we similarly analyzed ~13 million CD4+ T cells and recovered 4 sequences from

443   the sorted cells (0.30 proviruses per 10^6 CD4+ T cells) (Fig. 6c). From two cells, we

444   identified one intact provirus in each. Notably, in one p24+ sorted cell, we detected two

445   proviruses – one carrying the 270 nt deletion and one that was intact across the entire

446   viral genome (Fig. 6c, colored boxes).

447

448   Together, these results provide direct evidence of persistent infected cells carrying two

449   proviruses and support our *in vitro* studies showing that superinfection allowed the

450   mobilization of defective proviruses and the accumulation of mutations.

15

Discussion

Most PWH achieve suppression of plasma HIV-1 RNA to below the limit of detection (50 copies HIV-1 RNA/mL plasma) within 12 weeks of ART initiation[56–58]. However, HIV-1 persists in a latent reservoir in resting CD4+ T cells[5,6], and daily reactivation of a small fraction of these infected cells gives rise to residual viremia[40,88–90]. In rare cases, PWH experience detectable viremia after years of suppression despite adherence to ART[33,36,41,42]. In this situation, the viremia cannot be reduced by treatment intensification and is termed non-suppressible viremia (NSV)[91]. NSV is typically due to large clones of infected cells. In contrast, we describe a novel presentation of NSV in which treated PWH never achieved an undetectable viral load despite optimal adherence. Surprisingly, the NSV in these cases reflected a large and diverse reservoir of infected cells that was generated by mobilization of defective proviruses. Here, we provide direct evidence that the defective proviruses found in the study participants function as DIPs, a finding that has important implications for the use of TIPs as a novel therapeutic strategy.

We first interrogated the source of the viremia in participants who never achieved undetectable viral loads and found surprising sequence heterogeneity, suggesting that the NSV was caused by a novel mechanism distinct from previous reports[33,36,41,42]. The proviral frequency in both participants suggested that the reservoir was remarkably large compared to most PWH on ART, with almost every provirus being classified as intact by the IPDA. Near-full length sequencing revealed a very surprising result: almost all proviruses harbored exactly the same large deletions yet had numerous distinct mutations elsewhere in the genome. These deletions did not impact the binding of the IPDA primers and probes, explaining the high frequency of proviruses classified as intact by the IPDA. These defective proviruses could be induced to generate virions packaging defective RNA both *in vivo* and *ex vivo*. However, introducing these deletions into infectious molecular clones abolished virion production and prevented new infection events, suggesting that these deletions rendered the provirus fatally defective. Therefore, we hypothesized that before ART was initiated, cells harboring defective proviruses were superinfected with replication-competent viruses, leading to the incorporation of defective genomes into

16

infectious virus particles that could be released from superinfected cells, resulting in the propagation of the fatal deletions and introduction of new mutations through RT-generated errors in the newly infected cells. Subsequent infection of these cells by replication-competent virus could lead to further rounds of mobilization of the defective genome and accumulation of additional mutations. Overtime, this process could generate a diverse set of defective proviruses, all harboring the same fatal deletions. Given the HIV-1 RT error rate and the single round infectivity of the defective lineages, this process must have occurred for extended time intervals in order to result in the striking diversity we observed. We developed novel cell lines carrying the proviruses with the recurrent deletions and found that the defective proviruses could only propagate when superinfected with wildtype virus. Additionally, the recurrent deletions found in P1 replicated with an $R_0 > 1$ when introduced into the reference plasmid, NL4-3, suggesting that viruses with these deletions can propagate when superinfected. Lastly, we found direct evidence of superinfected cells from both participants, suggesting that superinfection was the mechanism responsible for the propagation of defective proviruses.

Although HIV-1 has evolved mechanisms to prevent superinfection (downregulation of CD4 by Nef, Env, and Vpr), superinfection has been documented in spleen tissue from PWH during untreated chronic infection[92–94]. Additionally, superinfection has been implicated in the loss of viral control in some post-treatment controllers[95]. Importantly, superinfection is the major mechanism for viral recombination, contributing to the formation of new circulating strains[96] and intra-host evolution[97], including the selection for drug resistance[98]. Although superinfection of cells *in vitro* has been shown to occur with greater than expected frequency[99–101], multiple studies have shown that >85% of CD4[+] T cells from the peripheral blood and lymph node tissues in PWH contain only a single viral genome[78,84]. Superinfection is also rare in other human retroviruses. During chronic HTLV-1 infection, the majority of infected cells contain a single provirus[102]. To our knowledge, previous to this work, there is no report of cells harboring multiple proviruses in PWH on ART.

17

513   Superinfection has also been investigated in the context of DIPs[48,73,103]. Although DIPs
514   have been shown to arise for most RNA viruses, they have not been found in PWH. This
515   study provides the first evidence for HIV-1 DIPs *in vivo*. HIV-1 DIPs are likely extremely
516   rare; the formation of an HIV-1 DIP requires a very specific type of error during reverse
517   transcription: the defective provirus must retain the packaging/encapsidation signal (Ψ)
518   and the Rev response element (RRE). These two regions are required for defective RNA
519   to be exported from the nucleus and to be packaged into viral particles. As a result, the
520   generation of proviruses which can serve as the "chassis" for DIPs is a rare event that
521   may occur only after years of untreated infection. Additionally, the accumulation of a large
522   number of distinct mutations on the deleted provirus "chassis" likely required a prolonged
523   period of untreated infection, as is documented in the extended clinical history of P2 (Fig.
524   S1a). Given that a pre-therapy clinical history was unavailable for P1, we estimated that
525   407 cycles of mobilization occurred, requiring a minimum of 2.23 years, based on the
526   error rate of the HIV-1 RT (3.4×10^{-5} mutations/bp/cycle) and a 2 day generation
527   time[21,104,105]. This estimate does not take into account the possibility of recombination.
528   Nevertheless, proviruses with the recurrent deletions were likely disseminated and
529   diversified during a prolonged period of high-level viral replication prior to ART.
530   Importantly, the defective proviruses identified in this study likely propagated at a time
531   with low CD4$^+$ T cell count (estimated range: 10^1 - 10^2 cells/µL), increasing the likelihood
532   of superinfection.

533

534   Interestingly, in both participants, the DIPs did not prevent CD4$^+$ T cell depletion.
535   Additionally, the putative DIPs did not result in a reduction in setpoint HIV-1 RNA levels,
536   although the clinical viral load measurements cannot discriminate between intact and
537   defective virus. Recent reports have suggested that molecular engineering of DIPs to
538   have an $R_0^{DEF} > 1$ can allow them to function as a therapeutic interfering particle (TIP),
539   reducing viremia and pathogenesis[50]. The defective proviruses from P1 appeared to have
540   an R_0^{DEF} of 1.86 in *in vitro* assays, which was markedly lower than engineered TIPs (~12)
541   despite their sequence similarity[50] (Fig. S5c). Despite promising results in pre-clinical
542   studies, the ability of TIPs to reduce viral set point and pathogenesis in PWH is unknown.
543   The very low CD4 nadirs and high viral loads described here raise concerns about the

18

544 effectiveness of the TIP strategy, but further investigation is clearly required. Our work
545 further shows that, in addition to 5' leader defects, other types of defective proviruses can
546 contribute to viremia[36,39,42]. These proviruses, if transcriptionally active, may complicate
547 the interpretation of viral load, leading to potential misinterpretations in clinical and cure-
548 focused studies.

549

550 Our study has limitations. Principally, we were restricted to peripheral blood samples
551 taken after years on ART and not during the period of untreated infection in which the
552 defective genomes propagated. Future work should aim to further interrogate the
553 prevalence of superinfected cells in PWH with high frequencies of infected cells who are
554 unable to reach an undetectable viral load.

555

556 In conclusion, we demonstrate that the failure to achieve an undetectable viral load
557 despite effective ART can be caused by a large, heterogeneous reservoir of infected cells
558 generated by a novel process involving the dissemination of defective proviruses through
559 DIPs. In this situation, the proviral landscape can be dominated by diverse proviruses that
560 share the same lethal deletions. Such a paradox led us to identify superinfection as a
561 novel mechanism for the mobilization of defective proviruses. Although we describe DIPs
562 in PWH for the first time, the DIPs found in both participants did not prevent CD4 T cell
563 depletion, raising concern about the effectiveness of DIPs as an HIV-1 therapeutic, unless
564 further engineered. Our work further demonstrates that defective proviruses, despite
565 being evolutionary dead ends, are not irrelevant, and prompt future studies to investigate
566 whether RNA and protein expression from defective proviruses contribute to chronic
567 immune activation, exhaustion of anti-HIV immunity, and the development of
568 comorbidities in PWH.

19

Materials and Methods

Study Participants. The study participants were referred by their HIV-1 care providers at the Bartlett Specialty Clinic, Johns Hopkins University. Peripheral blood samples (180mL) were collected at multiple time points. Historical samples were obtained through a longitudinal study at the Bartlett Specialty Clinic. HLA typing was performed by the Johns Hopkins University Immunogenetics Laboratory. Drug concentration levels for P1 were measured by University of Florida Infectious Disease Pharmacokinetics Laboratory.

Study approval. The Johns Hopkins Institutional Review Board approved this study. All study participants provided written informed consent before enrollment.

Study of HIV-1 sequences in plasma and CD4+ T cells. Whole blood samples were spun at 400g for 10 minutes at 4°C and plasma was removed and spun again for 400g at 10 minutes and frozen at -80°C. Upon thawing, HIV-1 virion-associated RNA was isolated as previously described[106]. RNA was used immediately for reverse transcription with Induro (NEB) according to the manufacturer's protocol. The cDNA was then used for single genome sequencing as previously described[36]. We initially recovered p6-RT sequences to exclude drug resistance and estimate plasma virus clonality. PCR products obtained with this approach were sequenced by Sanger sequencing (Azenta). CD4+ T cells were isolated by magnetic bead-based negative selection (Miltenyi Biotec). Genomic DNA was extracted (QIAamp DNA Mini Kit 51306, Qiagen) from each participant's CD4+ T cells and used for digital PCR (see below) or single genome sequencing. Near-full length HIV-1 proviral sequencing was conducted using a modified version of the FLIP-Seq assay[107].

Digital PCR. IPDA and RPP30 normalization was performed as previously described on total CD4+ T cells, using the QX200 Digital Droplet[61]. All IPDA experiments included a negative control (nuclease-free water) and a positive control (J-Lat DNA). All RPP30 experiments included a negative control (nuclease-free water) and a positive control (uninfected donor DNA). For P2, we designed custom primers and probes for the *env* region.

Quantification of the participant-specific deletions was conducted by designing custom primers and probes spanning either the 1417 nt or 313 nt (P1), or 270 nt (P2) deletion.

20

598 For P1, total proviruses was measured using a probe spanning LTR-gag[36]. For P2, we

599 quantified viruses without the 270 nt deletion by designing a custom primer and probe set

600 to bind to the region inside of the deletion (WT-270). To quantify the frequency of

601 proviruses with each respective deletion, we normalized the copies of the deletion targets

602 as a frequency of the total proviruses measured (P1: LTR-gag; P2: WT-270). We also

603 normalized the copies of targets of interests based on cell genome equivalents

604 (calculated by RPP30), as previously described[61]. All dPCR experiments included a

605 negative control (nuclease-free water) and a positive control (diluted amplicon containing

606 deletion of interest). Primer and probes are detailed in Supplemental Table 1. dPCR

607 reactions were run using the QIAcuity One Digital PCR System, with an initial

608 denaturation step of 95°C for 2 minutes followed by 45 cycles of 95°C for 5 seconds and

609 58°C for 45 seconds. RNA containing each deletion was also quantified (Fig. 3). Virion-

610 associated RNA was isolated as described above. RT-dPCR reactions were run using the

611 QIAcuity One Digital PCR system using the manufacturer's provided reverse

612 transcriptase and an additional incubation time of 50°C for 40 minutes for reverse

613 transcription before dPCR.

614 *QVOA.* QVOAs from total CD4$^+$ T cells were performed as previously described[13].

615 Supernatants from p24-postive wells were processed as previously described to

616 sequence virion-associated HIV-1 RNA[106]. Primers used to generate amplicons spanning

617 each deletion can be found in Supplementary Table 1.

618 *Analyses of HIV-1 sequences*

619 Multiple sequence alignments of longitudinal p6-RT or *env* plasma sequences were

620 performed using ClustalW. Root-to-tip distances to a majority consensus sequence for

621 each time point was determined using *p*-distances in MEGA (www.megasoftware.net).

622 Average pairwise distance was determined by averaging the *p*-distance between every

623 sequence pair. A test for panmixia was performed as previously described[108].

624 Near-full length amplicons were sequenced were performed by Plasmidsaurus using

625 Oxford Nanopore Technology. Raw FASTQ files were downloaded and either mapped to

626 HXB2 using minimap2[109] or *de novo* assembled using Flye[110]. Consensus sequences

627 were generated using >75 read coverage. Multiple sequence alignments were performed

21

628   using MUSCLE. Regions surrounding the recurrent deletions were aligned by hand.

629   Neighbor-joining trees were performed based on a *p*-distance and bootstrap analysis with

630   100 replicates. Phylogenetic trees were visualized using the *ape* package in R and

631   highlighter plots were generated using a custom script

632   (https://github.com/hariharanviv/highlighterplot).

633   *Analysis of HIV-1 expression upon T cell activation.* Total CD4$^+$ T cells were isolated from

634   PBMCs by negative selection. 4M CD4$^+$ T cells were plated into a 24-well plate at 2

635   million/mL in cR10 media (Roswell Park Memorial Institute 1640 Medium with GlutaMAX

636   (ThermoFisher), 10% heat-inactivated fetal bovine serum, 100 U/mL penicillin, and

637   100 µg/mL streptomycin) with 10 nM dolutegravir, 10 µM tenofovir disoproxil fumarate, 10

638   µM emtricitabine, and anti-CD3/CD28 antibody-coated magnetic beads (cell-to-bead

639   ratio, 1:1). Cells and culture supernatant were collected after 3 days (P1) or 4 days (P2).

640   Virion-associated RNA was isolated and quantified as described above.

641   *Testing the impact of deletion signatures on replicative fitness*. We used site-directed

642   mutagenesis (NEB) to introduce each major deletion signature into the NL4-3 plasmid,

643   obtained through the NIH HIV Reagent Program, Division of AIDS, NIAID, NIH: Human

644   Immunodeficiency Virus 1 (HIV-1), Strain NL4-3 Infectious Molecular Clone (pNL4-3),

645   ARP-114, contributed by Dr. M. Martin[70]. The primers for the site-directed mutagenesis

646   can be found in Supplementary Table 1. To generate infectious molecular clones, we first

647   seeded 16M HEK293T cells in a T150 flask with 20 mL of Dulbecco's Modified Eagle

648   Medium with 10% heat-inactivated fetal bovine serum. The next day we transfected the

649   HEK293T cells using lipofectamine 3000, 30 µg plasmid, and 5 µg pAdvantage

650   (Promega). 72 hours after transfection, we collected the supernatant, which was then

651   filtered and concentrated by ultracentrifugation with a 20% sucrose gradient. Virus

652   recovery was measured by p24 ELISA (PerkinElmer).

653   *Flow cytometry analysis of cells expressing HIV-1 Env*. HEK293T cells were transfected

654   with infectious molecular clones carrying the participant-specific deletions (as described

655   above), wildtype plasmid (NL4-3), or an infectious molecular clone with a truncated *env*

656   (NL4-3ΔEnv, a negative control; obtained through the NIH HIV Reagent Program, NIAID,

657   NIH: Human Immunodeficiency Virus Type 1 (HIV-1) Non-Infectious Molecular Clone,

22

658 pNL4-3ΔEnv, HRP-20281). Transfected HEK293T cells were washed with PBS and

659 dissociated using 2 mL trypsin-EDTA (0.25%) 2 days after transfection. The trypsin-EDTA

660 was quenched by the addition of 8 mL of cD10 followed by washing the cells with PBS.

661 Cells were incubated with 100 µL of yellow-fluorescent reactive dye (Invitrogen; 1:500

662 dilution in PBS) and incubated for 10 minutes at room temperature in the dark. Next, the

663 viability stain was quenched by the addition of FACS buffer (PBS + 10% fetal bovine

664 serum). Cells were stained with a cocktail of unlabeled primary broadly neutralizing

665 antibodies, 3BNC117 and VRC01 (15µg/mL each) for 1 hour at 37°C. Cells were also

666 stained using the same protocol with broadly neutralizing antibodies (bNAbs) targeting

667 different epitopes (10E8v4, PGT121, PGDM1400, PGT128) at the same concentration.

668 bNAbs were obtained through the NIH HIV Reagent Program, NIAID. After 2 washes, the

669 cells were then stained with APC-labeled secondary antibody against hu-IgG Fc for 30

670 minutes at 4°C. (100µL of 1:40 dilution; Rat IgG2a, κ; cat 410712 lot B3338999 clone

671 M1310G05). After 2 washes to remove excess antibodies, cells were analyzed using an

672 Intellicyt iQue cytometer. Nonspecific signal was assessed by staining cells with only the

673 secondary antibody.

674 *Design and production of stable cell lines containing reporter viral genomes with deletion*

675 *signatures.* The plasmid, NL4-3-ΔNef-mTagBFP2 (herein referred to as NL4-3-ΔNef-

676 BFP), was constructed using Gibson assembly to remove the Nef open reading frame

677 from the NL4-3 plasmid and insert mTagBFP2 (Addgene 54572[111]) in the same reading

678 frame. This design links BFP fluorescence with viral infectivity. Variants of this plasmid

679 containing the deletions of interest were generated via site-directed mutagenesis (as

680 described above), resulting in the plasmids NL4-3-ΔNef-BFP Δ1417Δ313 (representing

681 deletions identified in P1) and NL4-3-ΔNef-BFP Δ270 (representing the deletion identified

682 in P2). Due to the impact of these deletions on viral fitness (Fig. 4), a lentiviral system

683 was employed to package these plasmids into lentiviral particles, which were then used

684 to transduce target cells and create a cell line harboring proviruses with the respective

685 deletions. To generate the lentivirus, we transfected HEK293T cells using Lipofectamine

686 3000 (as described above) with 37.5 µg of the transfer plasmid (either NL4-3-ΔNef-BFP

687 Δ1417Δ313 or NL4-3-ΔNef-BFP Δ270), 10 µg of the packaging plasmid pHelp [112], 7.5 µg

688 of pWE, which encodes an HIV-1 CXCR4-tropic envelope, and 5 µg of pAdvantage

23

689 (Promega). 24 hours post-transfection, the media was removed, the cells were washed
690 with PBS, and fresh media was added (as described above). At 72 hours post-
691 transfection, lentivirus was filtered and concentrated using Lenti-X Concentrator (Takara)
692 according to the manufacturer's instructions. The lentivirus was resuspended in 1 mL
693 RPMI, aliquoted, and stored at -80°C until used. To generate polyclonal cell lines
694 harboring the defective proviruses, we transduced SupT1 cells (1.5 million cells / 2 mL
695 per well; SupT1 cells engineered to express CCR5; a gift from Dr. James Hoxie[113]) with
696 the previously described lentiviruses packaging defective proviruses. The transduction
697 was carried out by spinoculating SupT1s in the presence of 8 µg/mL polybrene (Sigma-
698 Aldrich), at 1200g for 2 hours at 37°C. The next day, the cells were spun and resuspended
699 in cR10 media containing 10 µM darunavir (MedChem Express) and 10µM enfuvirtide
700 (NIH HIV Reagent Program, NIAID, NIH, HRP-12732, contributed by DAIDS/NIAID) for
701 3-5 days to eliminate any replication-competent recombinants. Cells with less than 30%
702 transduction efficiency (to reduce the likelihood of multiple lentiviral integrations per cell)
703 as measured by BFP fluorescence were bulk sorted using the MolFlow Legacy and
704 cultured in cR10 medium supplemented with 10% conditioned media.

705 In vitro *superinfection experiments.* miRFP670 (Addgene 79987[114]) was similarly cloned
706 into the *nef* open reading frame to generate the construct NL4-3-ΔNef-miRFP670
707 (hereafter referred to as NL4-3-ΔNef-RFP). NL4-3-ΔNef-RFP was transfected following
708 the same protocol, with the exception that 55 µg of NL4-3-ΔNef-RFP and 5 µg of
709 pAdvantage were used during transfection. WT NL4-3-ΔNef-RFP virus was harvested 72
710 hours post transfection following the same protocol as above.

711 To test for interference, sorted BFP+ cell lines containing a provirus with either deletion
712 signature or mock-transduced (control) were infected with WT NL4-3-ΔNef-RFP virus.
713 100,000 cells were infected with NL4-3-ΔNef-RFP or mock-infected with cR10 by
714 spinoculation at 1200g for 2 hours at 37°C in a 96-well round-bottom plate in 200 µL cR10
715 containing 8 µg/mL polybrene. 24 hours after infection, cells were washed with PBS and
716 resuspended in fresh cR10 medium. Daily, 20-50 µL of the cells were analyzed for
717 fluorescent protein reporter expression by flow cytometry (Cytek Northern Lights), with
718 dead cells excluded using propidium iodide staining (BioLegend). On day 3, 50 µL

719   supernatant was removed, spun to eliminate cellular debris, and frozen at -80°C. The titer

720   of day 3 supernatant was determined by using 25 µL of the supernatant to spinoculate

721   100,000 WT SupT1 cells with the same infection protocol as above. To assess single-

722   round infectivity, 19hrs after infection the supernatant was removed, the cells were

723   washed, and the cells were resuspended in cR10 media containing 10 µM T20 and 5 mM

724   DRV. 48 hours after infection, the frequency of infected cells was measured by flow

725   cytometry (Cytek Northern Lights). The WT titer (Fig. 5c) was determined by normalizing

726   the percentage of cells infected with WT virus (% RFP$^+$) or defective virus (% BFP$^+$) with

727   condition where SupT1 cells were infected with WT virus only. The fraction of transmitted

728   proviruses (Fig. 5d) was determined by dividing the percentage of cells infected with WT

729   virus (% RFP$^+$) or defective virus (% BFP$^+$) by the total number of infected cells for each

730   condition (% RFP$^+$BFP$^-$ + % RFP$^-$BFP$^+$ + % RFP$^+$BFP$^+$).

731   Conditional replication was determined by using the defective cell lines and either mock-

732   infecting them or infecting them with the WT NL4-3-ΔNef-RFP virus. 3 days post infection,

733   the supernatant was removed and used to infect WT SupT1 cells in the same protocol as

734   described above. 24 hours after infection of WT SupT1 cells, the supernatant was

735   removed, and the cells were washed with PBS. Unlike the single-round infectivity

736   experiment, this experiment aimed to assess the replication competence of the viruses,

737   so we resuspended the cells in 200 µL of cR10 medium without antiretrovirals. Daily, 20-

738   50 µL of the cells were analyzed for fluorescent protein reporter expression by flow

739   cytometry (Cytek Northern Lights), with dead cells excluded using propidium iodide

740   staining (BioLegend).

741   R$_0$ quantification by three-color assay was conducted using a modified version of Pitchai,

742   Tanner et al.[50]. First, to generate a target cell line, SupT1-CCR5 were stably transduced

743   with a lentiviral reporter vector, pTY-eGFP (NIH HIV Reagent Program, Division of AIDS,

744   NIAID, NIH: Lentiviral Reporter Vector (pTY-EFeGFP), ARP-4828, contributed by Dr.

745   Lung-Ji Chang) using the same lentivirus protocol and transduction protocol as above in

746   *Design and production of model stable cell lines containing reporter viral genomes with*

747   *deletion signatures*. Cells with less than 30% transduction efficiency (increasing the

748   probability of only one transgene per cell) were single-cell sorted into cR10 media

25

supplemented with 10% conditioned media and expanded over 28 days. The cell population with the highest percentage of eGFP expression and the highest mean fluorescence intensity (MFI) for surface markers CXCR4 and CCR5 was selected for further experimentation (henceforth referred to as SupT1-eGFP cells). Either WT SupT1s or SupT1-Δ1417Δ313-BFP were infected with WT NL4-3-ΔNef-RFP using the same conditions as above. After 2 days of infection, 30,000 cells were cocultured with 70,000 SupT1-eGFP target cells in 200 μL cR10. Daily, the frequency of infected cells as analyzed as above. R_0 was calculated by taking the number of secondary infections and dividing it by the number of primary infections. The number of primary WT virus infections was calculated as the frequency of RFP$^+$ cells at the time of co-culture (% RFP$^+$ × GFP$^-$ cells in coculture); the number of secondary infections was calculated by multiplying the frequency of GFP$^+$RFP$^+$ cells 3 days after co-culture with all uninfected (target) cells (% GFP$^+$RFP$^+$ × (100,000 − HIV$^+$ infected cells at the time of co-culture)). R_0^{DEF} was calculated similarly, except calculating the frequency of BFP$^+$ cells as a measure of infection by the defective virus.

p24$^+$ single cell sorting. We modified the HIV-Flow assay[77] to single-cell sort p24$^+$ cells into individual wells and lysed them. To ensure that singlets were sorted, we first single sorted live, single cells from an uninfected male donor into 8 μL of lysis buffer (DirectPCR Lysis buffer + 400 μg/mL protease K) and then heated the mixture at 55°C for 60 mins and then 85°C for 15 minutes to inactive the protease K. We added PCR mastermix (Platinum SuperFi II Mastermix) directly to each sorted cell and distributed the mixture across 6 wells, with a final primer concentration of 0.4 μM per well. Primers for amplification of *SRY* are listed in Table S1. We then amplified the product in two rounds of PCR, using the manufacturer's recommended thermocycling conditions for both rounds and visualized the product by gel electrophoesis. For p24$^+$ sorting of ACH-2, P1, and P2 cells, we first activated CD4$^+$ cells using 162 nM PMA and 2 μg/mL ionomycin for 24 hours in the presence of 10 nM dolutegravir, 10 μM tenofovir disoproxil fumarate, 10 μM emtricitabine. The next day, extracellular staining was performed using CD3-BV785 (Clone UCHT1), CD8a-APC/Cy7 (Clone RPA-T8), CD4-BV421 (Clone OKT4), CD45RO-FITC (Clone UCHL1), and Live/Dead Fixable Near-IR viability dye (ThermoFisher Scientific cat. L34975). Cells were then fixed and permeabilized with the FOXP3 Buffer

26

780 Set (Biolegend cat. 421403), followed by intracellular staining of HIV-1 p24 with clone

781 28B7 APC (MediMabs cat. MM-0289-APC) and clone KC57 PE (Beckman Coulter cat.

782 6604667). p24$^+$ cells were single sorted into 8 µL of lysis buffer and digested as above.

783 Primers for the amplification of the near-full length genome (FLIP-Seq) and PCR

784 mastermix were added directly to each sorted cell and the mixture was distributed across

785 6-12 wells, with a final primer concentration of 0.4 µM per well. We then amplified the

786 product in two rounds of PCR, using the manufacturer's recommended thermocycling

787 conditions for both rounds of PCR. We determined the positive reactions were determined

788 by visualization of amplicons on an agarose gel. Amplicons were sequenced and

789 analyzed as above.

790 *Data Availability*. All HIV-1 sequences are available in NCBI's GenBank (accession

791 numbers pending).

792

793 **Acknowledgments**

802

803 **Author contribution**

804 V.H., F.R.S, and R.F.S. conceptualized the study. V.H., J.A.W, F.D., E.J.F., N.P., M.M., and

805 F.R.S. performed experiments. H.Z. conducted cell sorting experiments. A.S., J.L., and

806 S.A.B. enrolled study participants. E.P.S., E.A.G., D.S.B., J.K., and R.D.M. provided

27

807 clinical care and gathered clinical history. V.H. conducted analyses and generated figures.

808 V.H., J.D.S., R.F.S., and F.R.S. wrote the manuscript with feedback from all authors.

809

810 **Conflict of interest**

811 R.F.S. is an inventor on a patent application for the intact proviral DNA assay (IPDA)

812 (patent no. PCT/US16/28822) filed by Johns Hopkins University and licensed by

813 AccelevirDx. F.R.S. received payments from Gilead Sciences for participating at scientific

814 meetings.

28

References

1. Perelson, A. S. *et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191 (1997).

2. Hammer, S. M. *et al.* A Controlled Trial of Two Nucleoside Analogues plus Indinavir in Persons with Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less. *N Engl J Med* **337**, 725–733 (1997).

3. Gulick, R. M. *et al.* Treatment with Indinavir, Zidovudine, and Lamivudine in Adults with Human Immunodeficiency Virus Infection and Prior Antiretroviral Therapy. *New England Journal of Medicine* **337**, 734–739 (1997).

4. Chun, T.-W. *et al.* In vivo fate of HIV-1-infected T cells: Quantitative analysis of the transition to stable latency. *Nat Med* **1**, 1284–1290 (1995).

5. Chun, T.-W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).

6. Finzi, D. *et al.* Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science* **278**, 1295–1300 (1997).

7. Wong, J. K. *et al.* Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. *Science* **278**, 1291–1295 (1997).

8. Chun, T.-W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences* **94**, 13193–13197 (1997).

9. Davey, R. T. *et al.* HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proceedings of the National Academy of Sciences* **96**, 15109–15114 (1999).

10. Rothenberger, M. K. *et al.* Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proceedings of the National Academy of Sciences* **112**, E1126–E1134 (2015).

11. Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* **9**, 727–728 (2003).

12. Crooks, A. M. *et al.* Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies. *The Journal of Infectious Diseases* **212**, 1361–1365 (2015).

13. McMyn, N. F. *et al.* The latent reservoir of inducible, infectious HIV-1 does not decrease despite decades of antiretroviral therapy. *J Clin Invest* **133**, (2023).

14. Mendoza, P. *et al.* Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. *Journal of Experimental Medicine* **217**, e20200051 (2020).

15. Simonetti, F. R. *et al.* Antigen-driven clonal selection shapes the persistence of HIV-1–infected CD4+ T cells in vivo. *J Clin Invest* **131**, (2021).

16. Moskovljevic, M. *et al.* Cognate Antigen Engagement Induces HIV-1 Expression In CD4+ T Cells From People On Long-Term ART. 2024.02.27.582096 Preprint at https://doi.org/10.1101/2024.02.27.582096 (2024).

17. Ho, Y.-C. *et al.* Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell* **155**, 540–551 (2013).

18. Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043–1049 (2016).

19. Hiener, B. *et al.* Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants. *Cell Reports* **21**, 813–822 (2017).

20. Sanchez, G., Xu, X., Chermann, J. C. & Hirsch, I. Accumulation of defective viral genomes in peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected individuals. *Journal of Virology* **71**, 2233–2240 (1997).

21. Mansky, L. M. & Temin, H. M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *Journal of Virology* **69**, 5087–5094 (1995).

29

22. Koenig, S. *et al.* Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nat Med* **1**, 330–336 (1995).

23. Shankarappa, R. *et al.* Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology* **73**, 10489–10502 (1999).

24. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proceedings of the National Academy of Sciences* **100**, 4144–4149 (2003).

25. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).

26. Zanini, F. *et al.* Population genomics of intrapatient HIV-1 evolution. *eLife* **4**, e11282 (2015).

27. Onafuwa-Nuga, A. & Telesnitsky, A. The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination. *Microbiology and Molecular Biology Reviews* **73**, 451–480 (2009).

28. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).

29. Kieffer, T. L. *et al.* G→A Hypermutation in Protease and Reverse Transcriptase Regions of Human Immunodeficiency Virus Type 1 Residing in Resting CD4+ T Cells In Vivo. *Journal of Virology* **79**, 1975–1980 (2005).

30. Bender, A. M. *et al.* The Landscape of Persistent Viral Genomes in ART-Treated SIV, SHIV, and HIV-2 Infections. *Cell Host & Microbe* **26**, 73-85.e4 (2019).

31. Wagner, T. A. *et al.* Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573 (2014).

32. Cohn, L. B. *et al.* HIV-1 Integration Landscape during Latent and Active Infection. *Cell* **160**, 420–432 (2015).

33. Simonetti, F. R. *et al.* Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proceedings of the National Academy of Sciences* **113**, 1883–1888 (2016).

34. Lee, G. Q. *et al.* Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4$^+$ T cells. *J Clin Invest* **127**, 2689–2696 (2017).

35. Reeves, D. B. *et al.* A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nat Commun* **9**, 4811 (2018).

36. White, J. A. *et al.* Clonally expanded HIV-1 proviruses with 5′-leader defects can give rise to nonsuppressible residual viremia. *J Clin Invest* **133**, (2023).

37. Imamichi, H. *et al.* Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proceedings of the National Academy of Sciences* **113**, 8783–8788 (2016).

38. Pollack, R. A. *et al.* Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host & Microbe* **21**, 494-506.e4 (2017).

39. Imamichi, H. *et al.* Defective HIV-1 proviruses produce viral proteins. *Proceedings of the National Academy of Sciences* **117**, 3704–3710 (2020).

40. Bailey, J. R. *et al.* Residual Human Immunodeficiency Virus Type 1 Viremia in Some Patients on Antiretroviral Therapy Is Dominated by a Small Number of Invariant Clones Rarely Found in Circulating CD4+ T Cells. *Journal of Virology* **80**, 6441–6457 (2006).

41. Halvas, E. K. *et al.* HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing infectious virus. *J Clin Invest* **130**, 5847–5857 (2020).

42. Mohammadi, A. *et al.* Viral and host mediators of non-suppressible HIV-1 viremia. *Nat Med* **29**, 3212–3223 (2023).

30

43. Lanz, C. *et al.* Human Immunodeficiency Virus 1 Low-Level Viremia Predicts Viral Failure in Participants on Antiretroviral Therapy in the Swiss HIV Cohort Study. *Clinical Infectious Diseases* ciae569 (2024) doi:10.1093/cid/ciae569.

44. Weinberger, L. S., Schaffer, D. V. & Arkin, A. P. Theoretical Design of a Gene Therapy To Prevent AIDS but Not Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology* **77**, 10028–10036 (2003).

45. Metzger, V. T., Lloyd-Smith, J. O. & Weinberger, L. S. Autonomous Targeting of Infectious Superspreaders Using Engineered Transmissible Therapies. *PLOS Computational Biology* **7**, e1002015 (2011).

46. Rouzine, I. M. & Weinberger, L. S. Design Requirements for Interfering Particles To Maintain Coadaptive Stability with HIV-1. *Journal of Virology* **87**, 2081–2093 (2013).

47. von Magnus, P. Incomplete Forms of Influenza Virus. in *Advances in Virus Research* (eds. Smith, K. M. & Lauffer, M. A.) vol. 2 59–79 (Academic Press, 1954).

48. Huang, A. S. & Baltimore, D. Defective Viral Particles and Viral Disease Processes. *Nature* **226**, 325–327 (1970).

49. Klimatcheva, E. *et al.* Defective Lentiviral Vectors Are Efficiently Trafficked by HIV-1 and Inhibit Its Replication. *Molecular Therapy* **3**, 928–939 (2001).

50. Pitchai, F. N. N. *et al.* Engineered deletions of HIV replicate conditionally to reduce disease in nonhuman primates. *Science* **385**, eadn5866 (2024).

51. Marriott, A. C. & Dimmock, N. J. Defective interfering viruses and their potential as antiviral agents. *Reviews in Medical Virology* **20**, 51–62 (2010).

52. Lancaster, M. U., Hodgetts, S. I., Mackenzie, J. S. & Urosevic, N. Characterization of Defective Viral RNA Produced during Persistent Infection of Vero Cells with Murray Valley Encephalitis Virus. *Journal of Virology* **72**, 2474–2482 (1998).

53. Calain, P., Monroe, M. C. & Nichol, S. T. Ebola Virus Defective Interfering Particles and Persistent Infection. *Virology* **262**, 114–128 (1999).

54. Li, D. *et al.* Defective Interfering Viral Particles in Acute Dengue Infections. *PLOS ONE* **6**, e19447 (2011).

55. Frensing, T. *et al.* Continuous Influenza Virus Production in Cell Culture Shows a Periodic Accumulation of Defective Interfering Particles. *PLOS ONE* **8**, e72288 (2013).

56. Di Mascio, M. *et al.* In a Subset of Subjects on Highly Active Antiretroviral Therapy, Human Immunodeficiency Virus Type 1 RNA in Plasma Decays from 50 to <5 Copies per Milliliter, with a Half-Life of 6 Months. *Journal of Virology* **77**, 2271–2275 (2003).

57. Maldarelli, F. *et al.* ART Suppresses Plasma HIV-1 RNA to a Stable Set Point Predicted by Pretherapy Viremia. *PLOS Pathogens* **3**, e46 (2007).

58. Doyle, T. *et al.* Plasma HIV-1 RNA Detection Below 50 Copies/mL and Risk of Virologic Rebound in Patients Receiving Highly Active Antiretroviral Therapy. *Clinical Infectious Diseases* **54**, 724–732 (2012).

59. White, J. A. *et al.* Complex decay dynamics of HIV virions, intact and defective proviruses, and 2LTR circles following initiation of antiretroviral therapy. *Proceedings of the National Academy of Sciences* **119**, e2120326119 (2022).

60. Maldarelli, F. *et al.* HIV Populations Are Large and Accumulate High Genetic Diversity in a Nonlinear Fashion. *Journal of Virology* **87**, 10313–10323 (2013).

61. Bruner, K. M. *et al.* A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120–125 (2019).

62. Peluso, M. J. *et al.* Differential decay of intact and defective proviral DNA in HIV-1–infected individuals on suppressive antiretroviral therapy. *JCI Insight* **5**, (2020).

63. Simonetti, F. R. *et al.* Intact proviral DNA assay analysis of large cohorts of people with HIV provides a benchmark for the frequency and composition of persistent proviral DNA. *Proceedings of the National Academy of Sciences* **117**, 18692–18700 (2020).

31

64. Kwon, K. J. *et al.* Different human resting memory CD4+ T cell subsets show similar low inducibility of latent HIV-1 proviruses. *Science Translational Medicine* **12**, eaax6795 (2020).

65. Jensen, M. A., Coetzer, M., van 't Wout, A. B., Morris, L. & Mullins, J. I. A Reliable Phenotype Predictor for Human Immunodeficiency Virus Type 1 Subtype C Based on Envelope V3 Sequences. *Journal of Virology* **80**, 4698–4704 (2006).

66. Shao, W. *et al.* HIV Proviral Sequence Database: A New Public Database for Near Full-Length HIV Proviral Sequences and Their Meta-Analyses. *AIDS Research and Human Retroviruses* **36**, 1–3 (2020).

67. Maldarelli, F. *et al.* Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).

68. Bui, J. K. *et al.* Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLOS Pathogens* **13**, e1006283 (2017).

69. Josefsson, L. *et al.* The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences* **110**, E4987–E4996 (2013).

70. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *Journal of Virology* **59**, 284–291 (1986).

71. Rein, A. The heart of the HIV RNA packaging signal? *Proceedings of the National Academy of Sciences* **117**, 19621–19623 (2020).

72. Mascola, J. R. *et al.* Human Immunodeficiency Virus Type 1 Neutralization Measured by Flow Cytometric Quantitation of Single-Round Infection of Primary Human T Cells. *Journal of Virology* **76**, 4810–4821 (2002).

73. Bangham, C. M. & Kirkwood, T. B. L. Defective interfering particles: Effects in modulating virus growth and persistence. *Virology* **179**, 821–826 (1990).

74. Althaus, C. L., De Vos, A. S. & De Boer, R. J. Reassessing the Human Immunodeficiency Virus Type 1 Life Cycle through Age-Structured Modeling: Life Span of Infected Cells, Viral Generation Time, and Basic Reproductive Number, R0. *Journal of Virology* **83**, 7659–7667 (2009).

75. Ribeiro, R. M. *et al.* Estimation of the Initial Viral Growth Rate and Basic Reproductive Number during Acute HIV-1 Infection. *Journal of Virology* **84**, 6096–6102 (2010).

76. Eaton, J. W. & Hallett, T. B. Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence. *Proceedings of the National Academy of Sciences* **111**, 16202–16207 (2014).

77. Pardons, M. *et al.* Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. *PLOS Pathogens* **15**, e1007619 (2019).

78. Josefsson, L. *et al.* Majority of CD4+ T cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. *Proceedings of the National Academy of Sciences* **108**, 11199–11204 (2011).

79. Dufour, C. *et al.* Phenotypic characterization of single CD4+ T cells harboring genetically intact and inducible HIV genomes. *Nat Commun* **14**, 1115 (2023).

80. Sunshine, S. *et al.* HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay. *Journal of Virology* **90**, 4511–4519 (2016).

81. Symons, J. *et al.* HIV integration sites in latently infected cell lines: evidence of ongoing replication. *Retrovirology* **14**, 2 (2017).

82. Pinzone, M. R. & O'Doherty, U. Measuring integrated HIV DNA ex vivo and in vitro provides insights about how reservoirs are formed and maintained. *Retrovirology* **15**, 22 (2018).
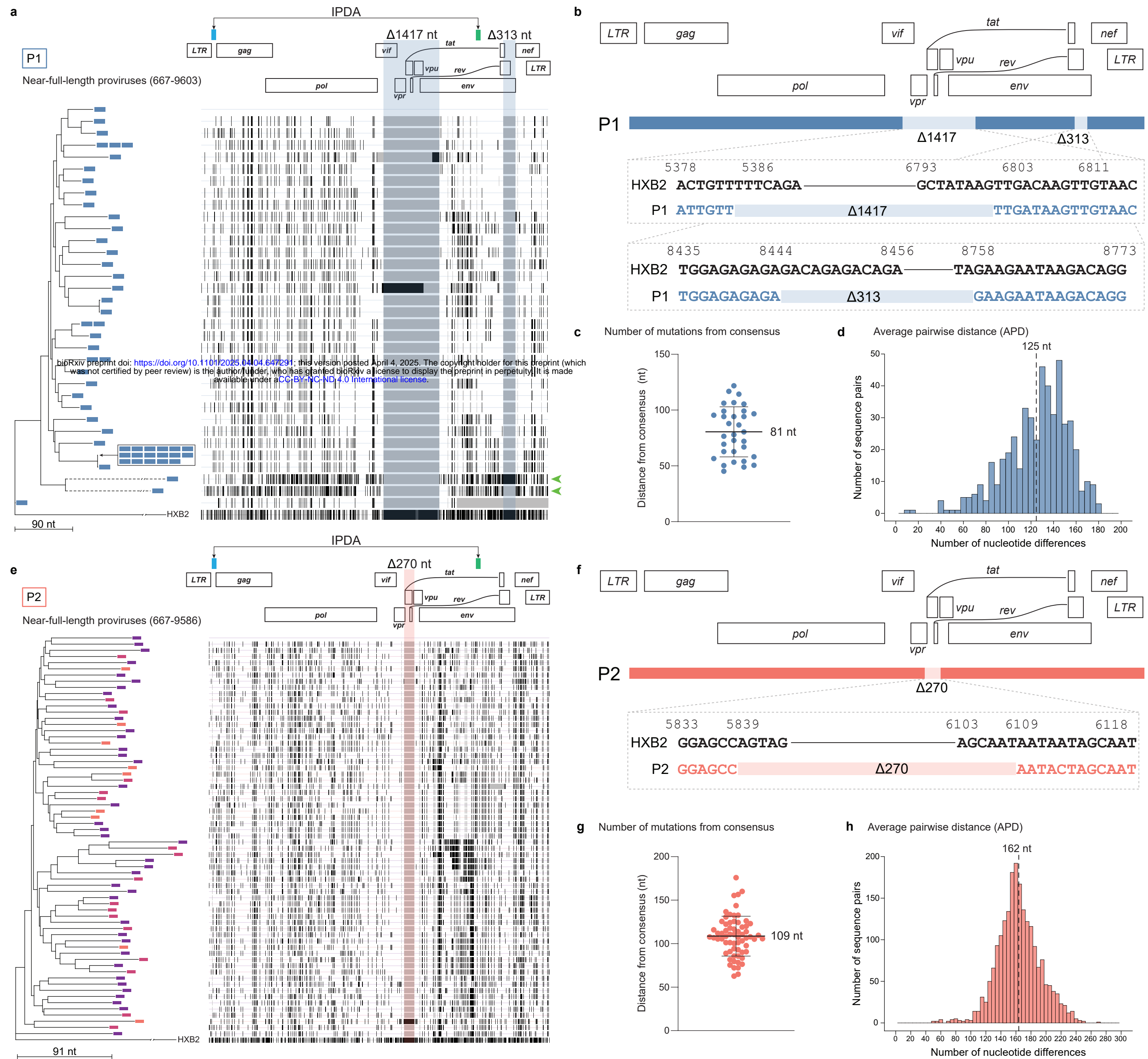
32

83. Simmonds, P. *et al.* Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of Virology* **64**, 864–872 (1990).

84. Josefsson, L. *et al.* Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals that the Majority of CD4+ T-cells Contain One HIV-1 DNA Molecule. *PLOS Pathogens* **9**, e1003432 (2013).

85. Cole, B. *et al.* In-depth single-cell analysis of translation-competent HIV-1 reservoirs identifies cellular sources of plasma viremia. *Nat Commun* **12**, 3727 (2021).

86. Pardons, M. *et al.* Potent latency reversal by Tat RNA-containing nanoparticle enables multi-omic analysis of the HIV-1 reservoir. *Nat Commun* **14**, 8397 (2023).

87. White, J. A. *et al.* Measuring the latent reservoir for HIV-1: Quantification bias in near full-length genome sequencing methods. *PLOS Pathogens* **18**, e1010845 (2022).

88. Dornadula, G. *et al.* Residual HIV-1 RNA in Blood Plasma of Patients Taking Suppressive Highly Active Antiretroviral Therapy. *JAMA* **282**, 1627–1632 (1999).

89. Palmer, S. *et al.* Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. *Proceedings of the National Academy of Sciences* **105**, 3879–3884 (2008).

90. Dinoso, J. B. *et al.* Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences* **106**, 9403–9408 (2009).

91. Wu, F. & Simonetti, F. R. Learning from Persistent Viremia: Mechanisms and Implications for Clinical Care and HIV-1 Cure. *Curr HIV/AIDS Rep* **20**, 428–439 (2023).

92. Jung, A. *et al.* Multiply infected spleen cells in HIV patients. *Nature* **418**, 144–144 (2002).

93. Gratton, S., Cheynier, R., Dumaurier, M.-J., Oksenhendler, E. & Wain-Hobson, S. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. *Proceedings of the National Academy of Sciences* **97**, 14566–14571 (2000).

94. Dixit, N. M. & Perelson, A. S. Multiplicity of Human Immunodeficiency Virus Infections in Lymphoid Tissue. *Journal of Virology* **78**, 8942–8945 (2004).

95. Altfeld, M. *et al.* HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature* **420**, 434–439 (2002).

96. Jost, S. *et al.* A Patient with HIV-1 Superinfection. *New England Journal of Medicine* **347**, 731–736 (2002).

97. Charpentier, C., Nora, T., Tenaillon, O., Clavel, F. & Hance, A. J. Extensive Recombination among Human Immunodeficiency Virus Type 1 Quasispecies Makes an Important Contribution to Viral Diversity in Individual Patients. *Journal of Virology* **80**, 2472–2482 (2006).

98. Nora, T. *et al.* Contribution of Recombination to the Evolution of Human Immunodeficiency Viruses Expressing Resistance to Antiretroviral Treatment. *Journal of Virology* **81**, 7620–7628 (2007).

99. Chen, J. *et al.* Mechanisms of Nonrandom Human Immunodeficiency Virus Type 1 Infection and Double Infection: Preference in Virus Entry Is Important but Is Not the Sole Factor. *Journal of Virology* **79**, 4140–4149 (2005).

100. Brégnard, C., Pacini, G., Danos, O. & Basmaciogullari, S. Suboptimal Provirus Expression Explains Apparent Nonrandom Cell Coinfection with HIV-1. *Journal of Virology* **86**, 8810–8820 (2012).

101. Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. Dynamics of HIV-1 recombination in its natural target cells. *Proceedings of the National Academy of Sciences* **101**, 4204–4209 (2004).

102. Cook, L. B., Rowan, A. G., Melamed, A., Taylor, G. P. & Bangham, C. R. M. HTLV-1–infected T cells contain a single integrated provirus in natural infection. *Blood* **120**, 3488–3490 (2012).
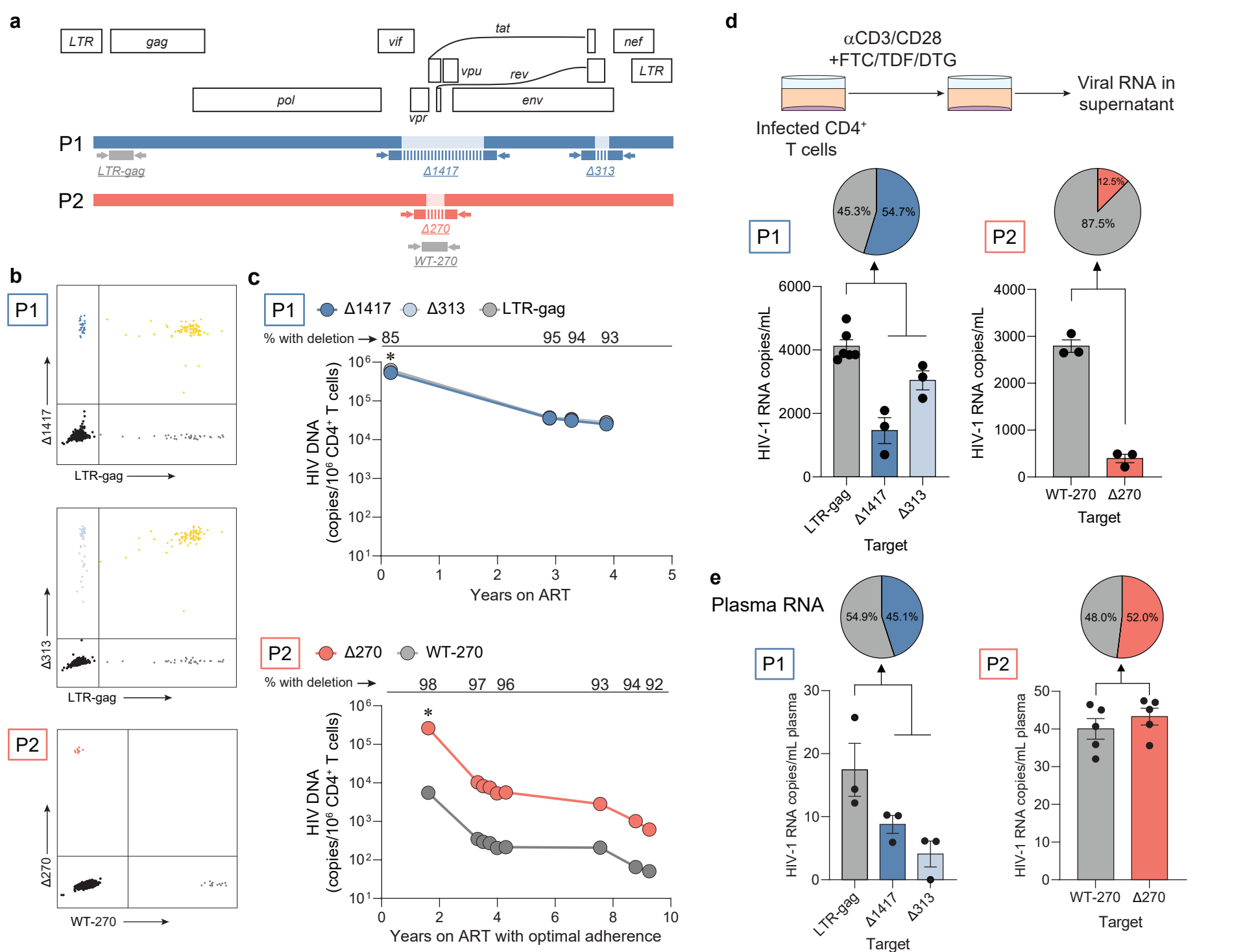
33

103. Huang, A. S. Defective Interfering Viruses. *Annual Review of Microbiology* **27**, 101–118 (1973).

104. Pathak, V. K. & Temin, H. M. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proceedings of the National Academy of Sciences* **87**, 6024–6028 (1990).

105. Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 Reverse Transcriptase. *Science* **242**, 1168–1171 (1988).

106. Tosiano, M. A., Jacobs, J. L., Shutt, K. A., Cyktor, J. C. & Mellors, J. W. A Simpler and More Sensitive Single-Copy HIV-1 RNA Assay for Quantification of Persistent HIV-1 Viremia in Individuals on Suppressive Antiretroviral Therapy. *Journal of Clinical Microbiology* **57**, 10.1128/jcm.01714-18 (2019).

107. Lee, G. Q. & Lichterfeld, M. Near-Full-Length Single-Genome HIV-1 DNA Sequencing. in *HIV Reservoirs: Methods and Protocols* (eds. Poli, G., Vicenzi, E. & Romerio, F.) 357–364 (Springer US, New York, NY, 2022). doi:10.1007/978-1-0716-1871-4_23.

108. Achaz, G. *et al.* A Robust Measure of HIV-1 Population Turnover Within Chronically Infected Individuals. *Molecular Biology and Evolution* **21**, 1902–1912 (2004).

109. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

110. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).

111. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLOS ONE* **6**, e28674 (2011).

112. Yang, H.-C. *et al.* Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. *J Clin Invest* **119**, 3473–3486 (2009).

113. Boyd, D. F. *et al.* Mutations in HIV-1 Envelope That Enhance Entry with the Macaque CD4 Receptor Alter Antibody Recognition by Disrupting Quaternary Interactions within the Trimer. *Journal of Virology* **89**, 894–907 (2014).

114. Shcherbakova, D. M. *et al.* Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat Commun* **7**, 12405 (2016).
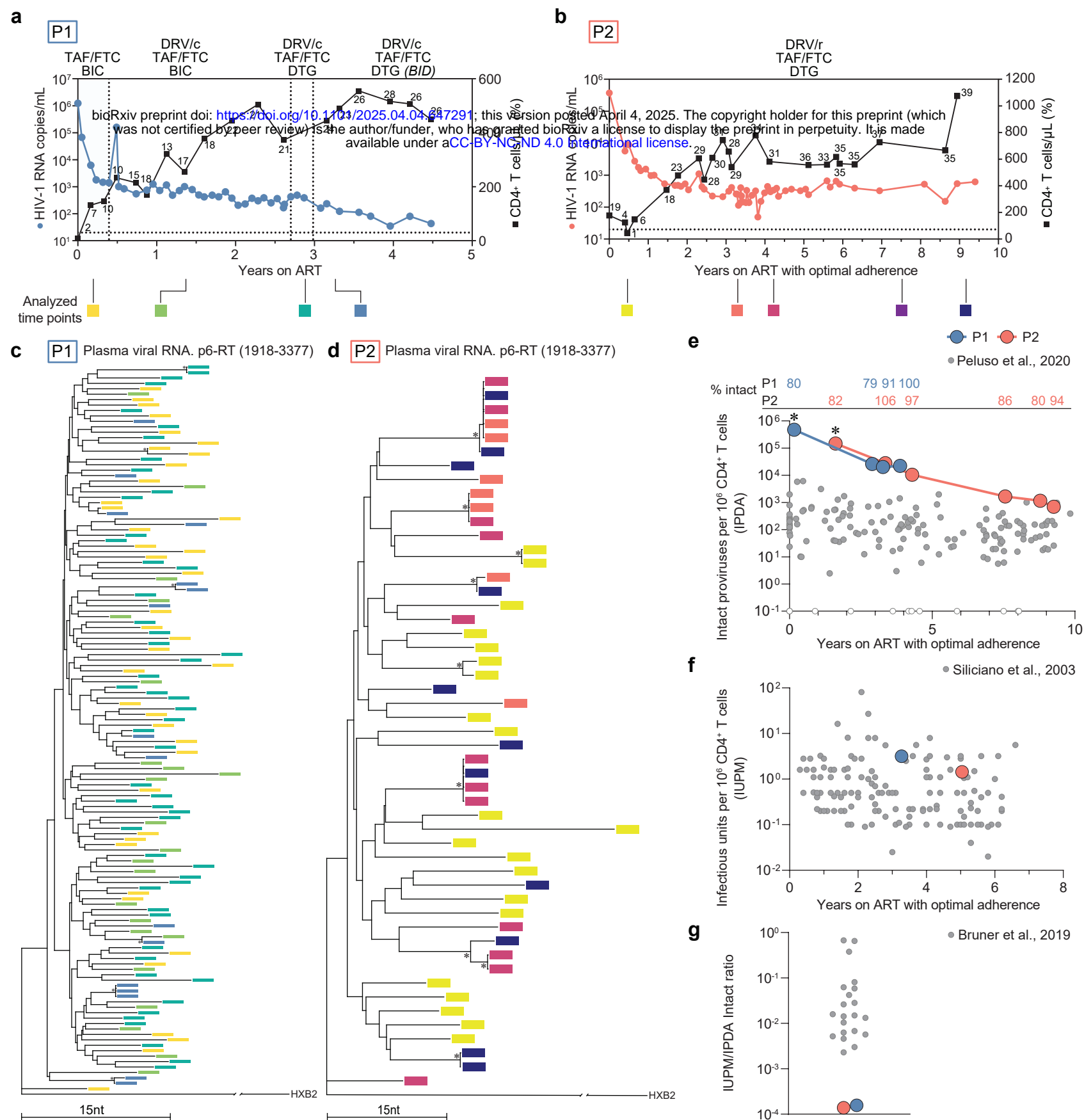
Figure 1. Failure to suppress viremia is characterized by diverse virus in plasma and high infected cell frequency. (**a** and **b**) Plasma HIV-1 RNA and CD4+ T cell counts over time for P1 and P2. Numbers above squares represent CD4+ T cell percentages. The dotted line at 20 copies/mL represents the current limit of detection for the clinical HIV-1 viral load assay. (**c** and **d**) Neighbor-joining phylogenetic trees of p6-RT single genome sequences obtained from plasma viral RNA for P1 and P2. Phylogenetic tree tip labels are color coded according to the plasma collection timepoint in **a**. Phylogenetic trees are rooted to HXB2 and HIV-1 coordinates refer to the HXB2 reference genome. Tree nodes with bootstrap values above 80 are marked with asterisks. (**e**) Intact proviral DNA frequences as measured by the IPDA. Open circles represent samples where no intact proviruses were detected. The percentage of proviruses classified as intact by the IPDA is shown on the top. Asterisks represent values from PBMC samples which were adjusted based on CD4+ T cell percentage. (**f**) Infectious unit per million (IUPM) CD4+ T cells as measured by the quantitative viral outgrowth assay. (**g**) IUPM to IPDA intact ratio as measured by dividing the IUPM by the closest IPDA timepoint value. TAF: tenofovir alafenamide; FTC: emtricitabine; BIC: bictegravir; DRV/c: darunavir/cobicistat; BID: *bis in die*; DRV/r: darunavir/ritonavir; DTG: dolutegravir

Figure 2. Near-full length sequencing reveals dominant proviral populations with unique deletion signatures and diverse mutations across the viral genome. (a and **e**) (Left) Neighbor-joining phylogenetic trees of near-full length proviral sequences obtained by single genome sequencing from primary CD4[+] T cells, rooted to HXB2. The color of each branch tip indicates sampling time as in Figure 1a and 1b. Dashed branches indicate sequences with hypermutation. (Right) Highlighter plot with black lines representing nucleotide changes compared to the top sequence. Gray vertical bars represent deletions compared to HXB2. Highlighted areas represent recurrent deletion patterns of 1417 nt, 313 nt (P1), and 270 nt (P2). IPDA primer probe regions are highlighted at the top. Green arrows point to sequences with significant G -> A hypermutation. (**b** and **f**) Mapped sequences of prominent deletion signatures found in majority of proviruses in P1 and P2 compared to HXB2. (**c** and **g**) Dot plots representing the number of mutations between each near-full length sequence and the majority consensus sequence. Horizontal line represents mean, error bars represent SD. (**d** and **h**) Histogram (bin width of 5) representing the nucleotide differences in unique proviruses from each participant. Dashed line represents the mean number of nucleotide differences between all the unique proviral sequences

Figure 3. Defective proviruses dominate the proviral landscape and contribute to NSV. (**a**) Location of primers (arrows) and probes (rectangles with vertical bars) to specifically quantify deletion signatures in viral RNA and DNA. Probes span the deletion region. (**b**) Representative two-dimension plots of dPCR showing duplex amplification of intact proviruses and proviruses of interest by deletion-specific assays. (**c**) Longitudinal quantification of proviruses with specific deletions of interest. (**d**) CD4⁺ T cells from P1 and P2 were cultured for 48 or 72 hours in the presence of emtricitabine (FTC), tenofovir disoproxil fumarate (TDF), dolutegravir (DTG) and anti-CD3/CD28 beads. The virion-associated RNA in the supernatant was measured by RT-dPCR. Error bars represent the SEM. Pie chart shows percentage of HIV-1 RNA copies with deletion normalized to the total number of copies. (**e**) Plasma virion-associated RNA was measured by RT-dPCR. Error bars represent SEM. Pie chart shows percentage of HIV-1 RNA copies with deletion normalized to the total number of copies. For P1, total proviruses were quantified by measuring the highly conserved region LTR-gag. For P2, proviruses without the 270 nt deletion were measured by using a primer-probe set inside of the deletion (WT-270).
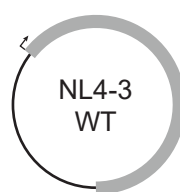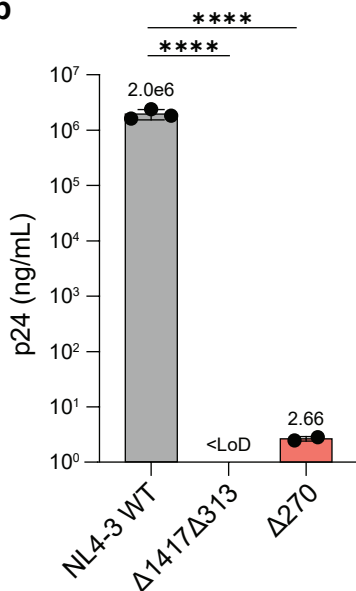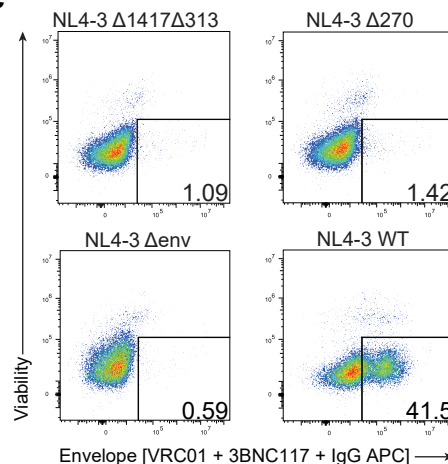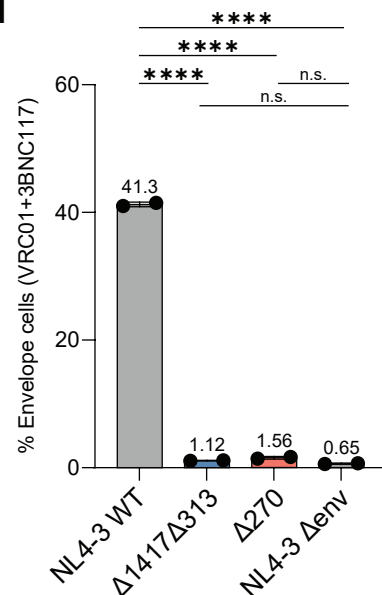
**Figure 4. Deletions found in proviruses abolish virus production *in vitro* (a)** Prominent deletions found in proviruses were introduced into an NL4-3 expression plasmid by site-directed mutagenesis. Arrows represent orientation of the provirus within the expression plasmid. (**b**) Virus produced upon HEK293T transfection was pelleted by ultracentrifugation, and p24 was measured by ELISA. Lower limit of detection (LoD) was 0.625 ng/mL. Horizontal bars represent mean, and error bars represent SD (**c**) Representative flow cytometry plots of the surface staining of transfected HEK293T cells with viability dye and bNAbs (VRC01 and 3BNC117). (**d**) Surface staining of HIV-1 Env with bNAbs (VRC01 and 3BNC117) on HEK293T cells 24 hours after transfection. Horizontal bars represent mean and error bars represent SEM. (**b** and **d**) Statistical significance between conditions was determined by 1-way ANOVA. **** $P < 0.0001$. n.s. $P > 0.01$
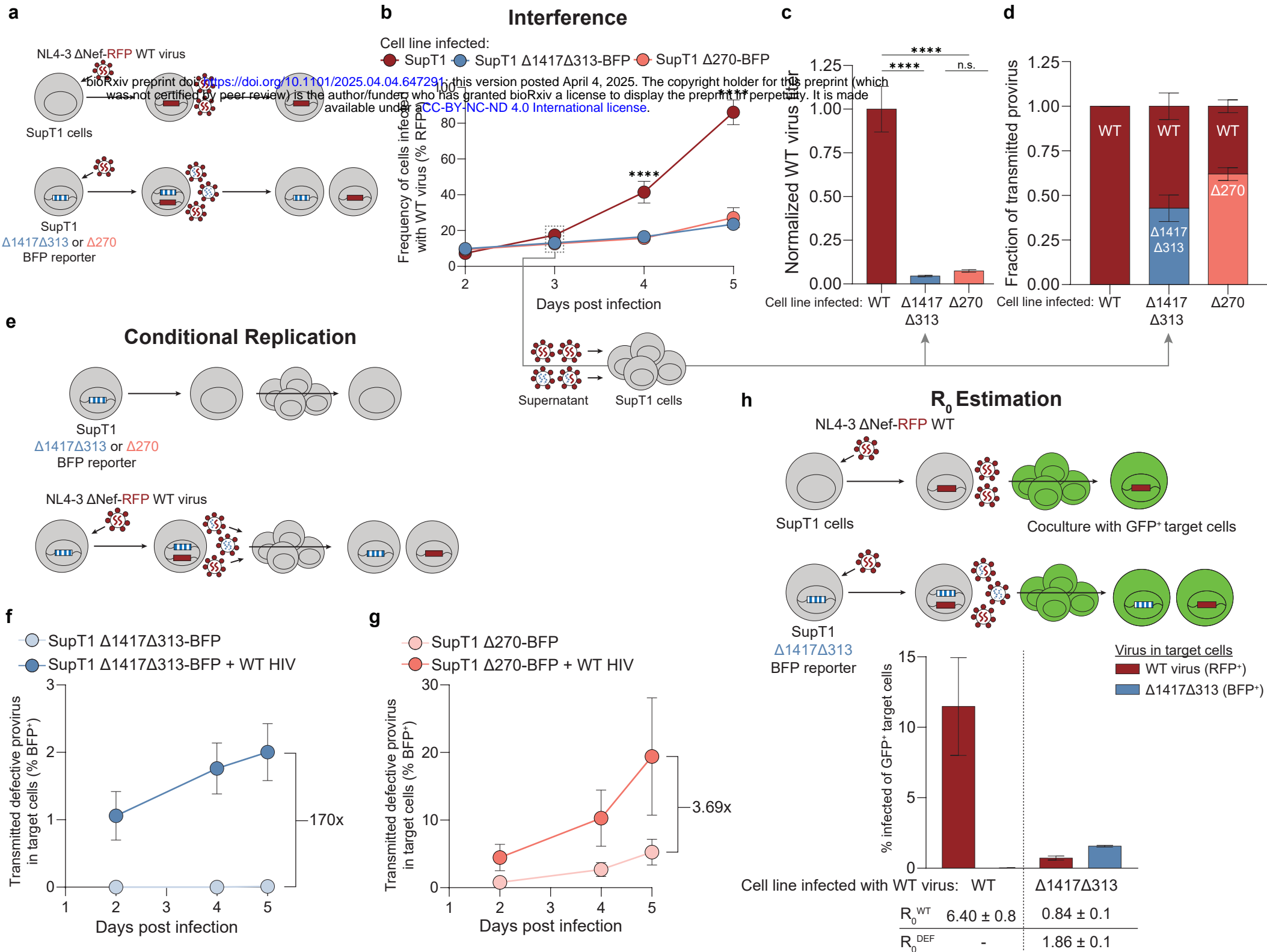
Figure 5. Cell culture model to assess superinfection *in vitro*. (**a**) Schematic for assessment of superinfection *in vitro*. (Top) SupT1 cells are infected with NL4-3-ΔNef-RFP WT virus. (Bottom) Model cell lines transduced with reporter defective viruses representing P1 (Δ1417Δ313) or P2 (Δ270) are infected with NL4-3-ΔNef-RFP WT virus. Virions produced by these superinfected cells may package RNA from the defective (blue) or WT (red) viral genome. (**b**) Frequency of cells infected with the WT virus (% RFP+) measured over time. Points represent mean and error bars represent SD. (**c**) Viral supernatant from day 3 of infection of each cell line was used to infect target SupT1 cells to determine the viral titer of WT virus after single round of infection. WT viral titer was normalized to virus from the WT SupT1 cell line. Horizontal bars represent mean and error bars represent SEM. (**d**) Viral supernatant from day 3 of infection of each cell line was used to infect SupT1 cells to determine the fraction of transmitted provirus after single round of infection as measured by flow cytometry. (**e**) Schema for assessment of conditional replication *in vitro*. Model cell harboring reporter defective viruses are either mock-infected or infected with WT virus. After 3 days of culture, the resulting supernatant is used to infected target SupT1 cells. (**f** and **g**) Frequency of target SupT1 cells infected with defective provirus measured over time. Points represent mean and error bars represent SD. (**h**) (Top) Schema for three-color experiment to estimate R_0 *in vitro*. WT SupT1 or SupT1-Δ1417Δ313-BFP are infected with WT virus. After 2 days, the cells were cocultured with excess GFP+ SupT1 target cells. (Bottom) Frequency of target GFP+ SupT1 target cells infected with either WT virus (red) or defective virus (blue). R_0 values and standard error of mean for both the WT virus (R_0^{WT}) and defective virus (R_0^{DEF}) are listed below. Horizontal bars represent mean and error bars represent SD. (**b** - **h**) n=6 for all conditions assessed. (**b** - **c**) Statistical significance between cell lines was determined by 1-way ANOVA. **** P <0.0001. n.s. P > 0.05.
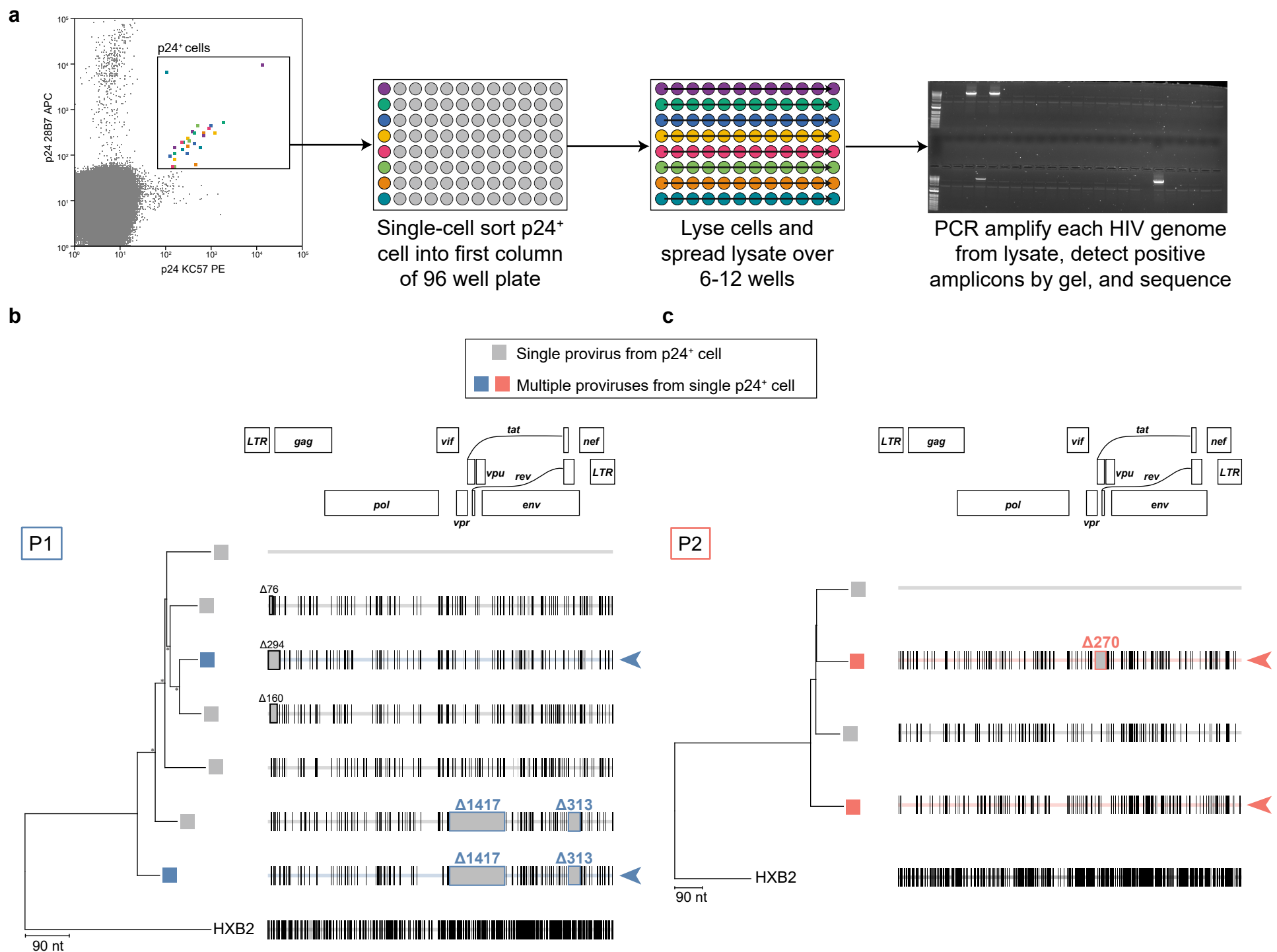
Figure 6. Single p24⁺ cell sequencing identifies dually infected cells *in vivo* (a) Experimental design to characterize p24⁺ cells and detect multiple integrated proviruses. Sorted p24⁺ cells are lysed and spread over multiple wells. Near-full length proviral amplification, agarose gel electrophoresis, and sequencing are conducted on the cell lysate. Example agarose gel electrophoresis shown here. (**b** and **c**) Neighbor-joining phylogenetic tree of near-full length proviral genomes from p24⁺ cells. Highlighter plot with black lines representing nucleotide changes compared to the top sequence. Gray boxes represent deletions compared to the top sequence of each tree. Tree nodes with bootstrap values above 80 are marked by asterisks. 5' internal deletions are boxed in black and annotated. Common deletions found in P1 and P2 are boxed and annotated. Phylogenetic tree tips with gray boxes represent single provirus found from a sorted cell. Colored boxes represent multiple proviruses found from a single-sorted cell, likely representing multiple integrated proviruses within one cell. Colored arrows point to multiple proviruses found in a single-sorted cell.