




HIV-1 diversity considerations in the application of the Intact Proviral DNA Assay (IPDA)

Natalie N. Kinloch^{1,2,11}, Yanqin Ren^{3,11}, Winiffer D. Conce Alberto³, Winnie Dong², Pragya Khadka³, Szu Han Huang³, Talia M. Mota³, Andrew Wilson⁴, Aniqah Shahid^{1,2}, Don Kirkby², Marianne Harris^{2,5}, Colin Kovacs⁶, Erika Benko⁶, Mario A. Ostrowski⁷, Perla M. Del Rio Estrada⁸, Avery Wimpelberg⁹, Christopher Cannon⁹, W. David Hardy⁹, Lysay MacLaren⁹, Harris Goldstein ¹⁰, Chanson J. Brumme^{2,5}, Guinevere Q. Lee³, Rebecca M. Lynch⁴, Zabrina L. Brumme ^{1,2,12}✉ & R. Brad Jones ^{3,4,12}✉

The Intact Proviral DNA Assay (IPDA) was developed to address the critical need for a scalable method for intact HIV-1 reservoir quantification. This droplet digital PCR-based assay simultaneously targets two HIV-1 regions to distinguish genomically intact proviruses against a large background of defective ones, and its application has yielded insights into HIV-1 persistence. Reports of assay failures however, attributed to HIV-1 polymorphism, have recently emerged. Here, we describe a diverse North American cohort of people with HIV-1 subtype B, where the IPDA yielded a failure rate of 28% due to viral polymorphism. We further demonstrate that within-host HIV-1 diversity can lead the IPDA to underestimate intact reservoir size, and provide examples of how this phenomenon could lead to erroneous interpretation of clinical trial data. While the IPDA represents a major methodological advance, HIV-1 diversity should be addressed before its widespread adoption as a principal readout in HIV-1 remission trials.

¹Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada. ²British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada. ³Infectious Diseases Division, Department of Medicine, Weill Cornell Medical College, New York, NY, USA. ⁴Department of Microbiology, Immunology and Tropical Medicine, George Washington University, Washington, DC, USA. ⁵Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada. ⁶Maple Leaf Medical Clinic, Toronto, ON, Canada. ⁷Department of Medicine, University of Toronto, Toronto, ON, Canada. ⁸Center for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico. ⁹Whitman Walker Health, Washington, DC, USA. ¹⁰Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA. ¹¹These authors contributed equally: Natalie N. Kinloch, Yanqin Ren. ¹²These authors jointly supervised this work: Zabrina L. Brumme, R. Brad Jones. ✉email: zbrumme@sfu.ca; rbjones@med.cornell.edu

The Intact Proviral DNA Assay (IPDA) was developed to address the critical need for a precise and scalable method to quantify intact HIV-1 proviruses, which represent the main barrier to achieving HIV-1 remission or cure¹. This duplexed droplet digital PCR (ddPCR) assay simultaneously targets two HIV-1 regions, the Packaging Signal (Ψ) near the 5' end of the viral genome and the Rev Responsive Element (RRE) within Envelope (*env*), to distinguish genomically intact proviruses against a large background of defective ones. The IPDA requires less time, resources and biological material than the current gold standard for replication-competent HIV-1 reservoir measurement, the Quantitative Viral Outgrowth Assay (QVOA)², while also quantifying the total proviral burden. Application of the assay to increasing numbers of individuals living with HIV-1 has yielded insights into reservoir composition and dynamics^{3–6}, and the assay is being recommended for use in clinical trials evaluating HIV-1 remission strategies^{4,5,7}.

Recently however, instances of assay failure, attributable to natural HIV-1 polymorphism in primer and/or probe binding regions, have been described^{5–7}. While detection failures are expected for any molecular assay targeting a genetically variable pathogen, this rate has varied markedly across studies, ranging from 0% in the initial two reports (0/62¹ and 0/81⁴) to 12% (6/50) in the most recent publication⁷, where the assay was performed by the same group. A recent multi-cohort study comprising 400 individuals reports an overall 6.3% failure rate⁵; however, because this study includes the above⁴ and other published cohorts with 0% failure rates^{8–10} it can be inferred that the rates for the unpublished datasets are higher. Thus, while the IPDA is undeniably a major methodological advance, further study is needed to delineate the impact of HIV-1 sequence diversity. This is particularly relevant to clinical trials, which would either require an accurate anticipated rate of failure for power calculations, or strategies to address diversity directly, which might include pre-enrollment screening to determine IPDA-detectability of an individual's virus and/or development of secondary primers and probes.

Here, we aimed to further develop the IPDA by assessing its performance within a diverse cohort of individuals with HIV-1 subtype B from across North America. Our study is unique in a number of ways. First, with the exception of one study that investigated a subset (20%) of assay failures by HIV-1 sequencing⁵, no studies to date have combined IPDA, QVOA, and HIV sequencing to examine the frequency and potential implications of assay failures in detail. We describe an IPDA failure rate of 28% in our cohort, where we used QVOA to confirm that intact proviruses were indeed present in these cases, and where we present the underlying sequence polymorphisms for each case of failure. Second, we illustrate how within-host HIV diversity—specifically, cases where an individual harbors viral variants that are detectable and others that are undetectable by the assay—could lead the IPDA to underestimate intact HIV-1 reservoir size. We further provide examples of how this type of error could negatively impact clinical trial results interpretation. Third, our study provides an assessment of the IPDA by a group other than its original developers (or the associated company Accelevir Diagnostics). This could benefit the field, given that a major advantage of the IPDA is its potential for broad accessibility. Our failure rate of 28% occurred in the context of our efforts to precisely match the reported experimental conditions, with one exception which could have affected of a minority of these cases. Whereas Peluso et al.⁴ state that gates to define positive droplets can be drawn tightly against double-negative droplets as defined in samples from individuals without HIV-1, we demonstrate that fluorescence signal from Ψ -positive templates

(labeled with FAM) can bleed into the *env* channel, creating a false double-positive population, if gated in this manner. Thus, we required separation between negative and positive droplet populations in order to consider a result valid. This represents a previously unreported source of error in the IPDA. Taken together, we present an independent assessment of the scope of the challenge of addressing HIV-1 diversity in the IPDA, and—through sequence characterization—contribute to understanding the catalogue of polymorphisms that will need to be addressed with secondary primers and probes, towards improved iterations of this valuable assay.

Results

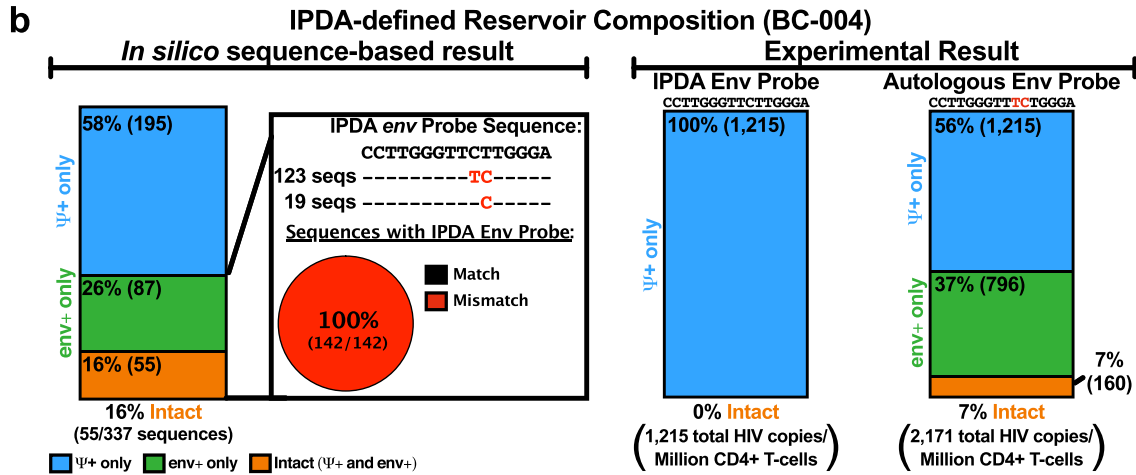
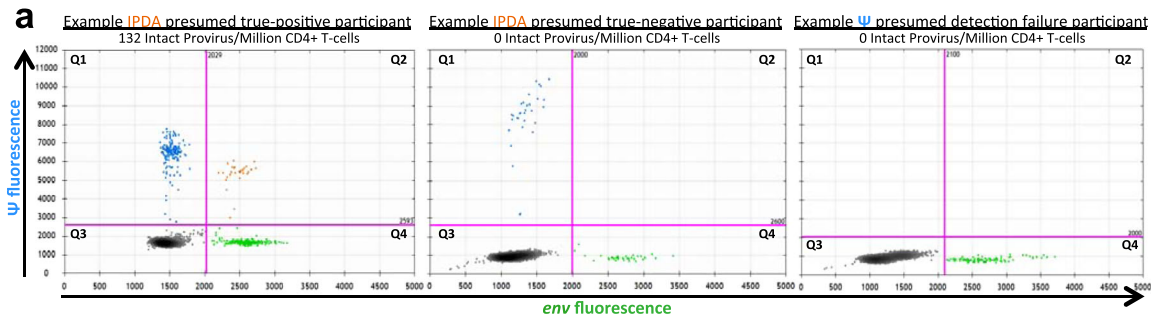
IPDA measurements in a diverse North American cohort. We applied the IPDA to 46 individuals with HIV-1 subtype B who had achieved viral suppression from cohorts across Canada, the USA and Mexico. Initial assay results yielded a median of 29 (interquartile range [IQR] 0–93) intact proviruses/million CD4⁺ T-cells (Supplementary Fig. 1); however, among these were 17 participants (37%) for whom the IPDA did not detect any intact (i.e., Ψ and *env* double-positive) proviruses. In four of these individuals, both Ψ - and *env*- single-positive proviruses were detected, suggesting a true-negative result (see example in Fig. 1a and Supplementary Fig. 2). In the remaining 13 individuals, however, the IPDA did not detect Ψ - and/or *env*- single-positive proviruses above background levels and instead yielded ddPCR plots that typify assay failure^{5,7} (Fig. 1a, Supplementary Fig. 2). Specifically, in eight of these participants only Ψ -positive proviruses were detected, in four only *env*-positive proviruses were detected, and in one participant no proviruses were detected. Furthermore, replication-competent HIV-1 was recovered in 11/11 (100%) cases where sufficient biological material was available to perform QVOA, consistent with the failure of IPDA to detect autologous intact proviruses.

HIV sequence polymorphism can cause IPDA detection failure

We next investigated the presumed cases of IPDA failure using HIV-1 sequencing. Near-full-length single-genome proviral sequencing performed on IPDA *env*-negative participant BC-004 revealed mismatches to the IPDA *env* probe, where in silico predicted reservoir distributions that took these polymorphisms into account (Fig. 1b, left) differed markedly from the original experimentally-obtained result (Fig. 1b, center). Substituting an autologous *env* probe rescued detection to in silico predicted levels (Fig. 1b, right), confirming that HIV-1 polymorphism can cause the IPDA to fail. Mismatches in the probe and/or at the 3' end of a primer were confirmed in all 13 cases of presumed assay detection failure (Fig. 1c). Together, this yielded an overall failure rate of 13/46 (28%). Indeed, exclusion of these 13 datapoints markedly improved the correlation between IPDA and QVOA results among those for whom sufficient biological material was available to perform the latter assay, from $\rho = 0.03$, $p = 0.83$ to $\rho = 0.35$, $p = 0.07$ (Fig. 1d), bringing this correlation more in line with that reported by the original authors¹ and what is expected biologically. By contrast, the correlation between QVOA and total HIV-1 *gag* DNA, measured using different HIV-1 primers and probe located in a conserved region¹¹, was significant without exclusion of these datapoints (Fig. 1e).

Within-host HIV diversity can cause reservoir underestimation

Although laborious to correct, complete detection failures due to HIV-1 polymorphism are nevertheless easy to flag due to their characteristic ddPCR plot presentation (Fig. 1a, Supplementary Fig. 2). In contrast, within-host HIV-1 diversity in IPDA primer or probe regions (see Fig. 1b for an example of



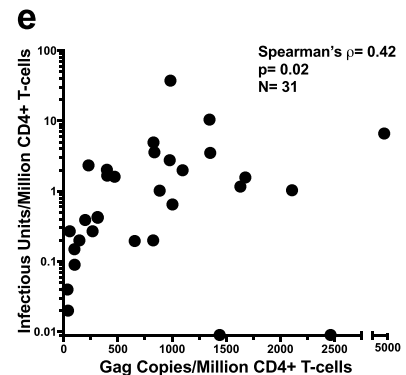
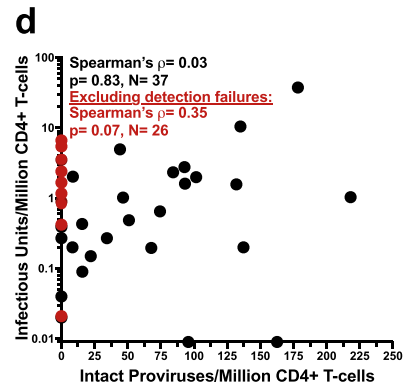
c Ψ sequence polymorphisms in IPDA Ψ detection failure samples

ID	Ψ Forward Primer	Ψ Probe	Ψ Reverse Primer
HGLK001	-----GC	ACTGGTGAAGTACGCCAAA	-----T
OM5365	-----GC	R-----TT	-----TT
WWH-012	-----GC	-----TT	-----TT
WWH-031	-----A	-ACA-A-----TT	-----A-----A
BC-014	-----AC	-----TT	-----TT

env sequence polymorphisms in IPDA *env* detection failure samples

ID	<i>env</i> Forward Primer	<i>env</i> Probe	<i>env</i> Reverse Primer
HGLK002	-----G-----	---C---ATC-	-----R-
HGLK005	-----*R-K---G-----	---C---	-----R
CIENI223	-----G-----	---C---	-----R
OM5334	-----G-----	---A---	-----R
CIRC0333	-----G-----	---T---	-----A
WWH-011	-----G-----	---T---	-----A
WWH-026	-----G-----	---T---	-----A
WWH-031	-----G-----	---TC---	-----A
BC-004	-----A-----	---YC---	-----A

* = ATGCWR insertion prior to R



diversity within a probe region) could lead the IPDA to underestimate intact reservoir size if the within-host variants were differentially detectable by the assay. Such partial detection failures would not be easy to identify. Moreover, if IPDA-detectable and nondetectable reservoir subpopulations were differentially susceptible to HIV-1 cure interventions, this could lead to erroneous conclusions regarding intervention efficacy.

Although such a scenario remains hypothetical, we nevertheless illustrate our point using HIV-1-specific broadly neutralizing antibodies (bNABs) as an example. bNABs can facilitate elimination of HIV-1-infected cells¹² in part by targeting them for antibody-dependent cellular cytotoxicity (ADCC)^{12,13}. As such, they are being evaluated in clinical trials, some of which use the IPDA as a readout (e.g., ACTG A5386¹⁴). Participant 91C33 from a published trial¹⁵

Fig. 1 Interindividual HIV diversity can lead to detection failure by the IPDA. **a** IPDA 2D ddPCR plots. Ψ -single-positive (Q1, blue), Ψ - and *env*- double-positive (Q2, orange), double-negative (Q3, gray) and *env* single-positive (Q4, green) events for an IPDA true-positive individual, a presumed IPDA true-negative individual and a presumed case of Ψ detection failure. Plots show merge of four replicate wells. **b** IPDA detection failure due to interindividual HIV diversity. In silico analysis of 337 linked Ψ and *env* probe sequences from proviral DNA of participant BC-004, who originally failed detection in *env*, revealed single nucleotide mismatches (red, inset) to the IPDA *env* probe. These sequences yielded in silico Ψ +(blue), *env*+(green) and intact (Ψ +*env*+; orange) provirus frequencies as shown on left. Experimental results using the published IPDA and autologous *env* probe are shown on right. Source data, including the 337 Ψ and *env* probe sequences and their classifications, are provided in the Source Data File. **c** HIV polymorphisms in IPDA primer and probe region sequences for participants with Ψ or *env* detection failure. Hyphens (-) indicate matches to the IPDA primer or probe; red letters indicate mismatches; asterisk indicates insertion location. **d** Correlation between reservoir size as measured by IPDA (Intact Proviruses/Million CD4 + T-cells) and QVOA (Infectious Units/Million CD4 + T-cells). When data from all 37 virally-suppressed participants are included, Spearman's $\rho = 0.03$ with a two-tailed $p = 0.83$. After excluding 11 presumed IPDA detection failures (red datapoints), Spearman's $\rho = 0.35$ with a two-tailed $p = 0.07$. Two individuals for whom no replication-competent viruses were detected (IUPM = 0) are plotted on the X-axis. QVOA replicates are detailed the methods; IPDA data represent the point estimate from four merged technical replicates per sample. Source data are provided in the Source Data file. **e** Correlation between QVOA and HIV *gag* copies/million CD4 + T-cells. Analysis comprises 31 unique participants, yielding Spearman's $\rho = 0.42$, two-tailed $p = 0.02$. Two individuals for whom no replication-competent viruses were detected (IUPM = 0) are plotted on the X-axis. HIV *gag* copies report the point estimate from at least four merged technical replicates for each sample. Source data are provided in the Source Data File.

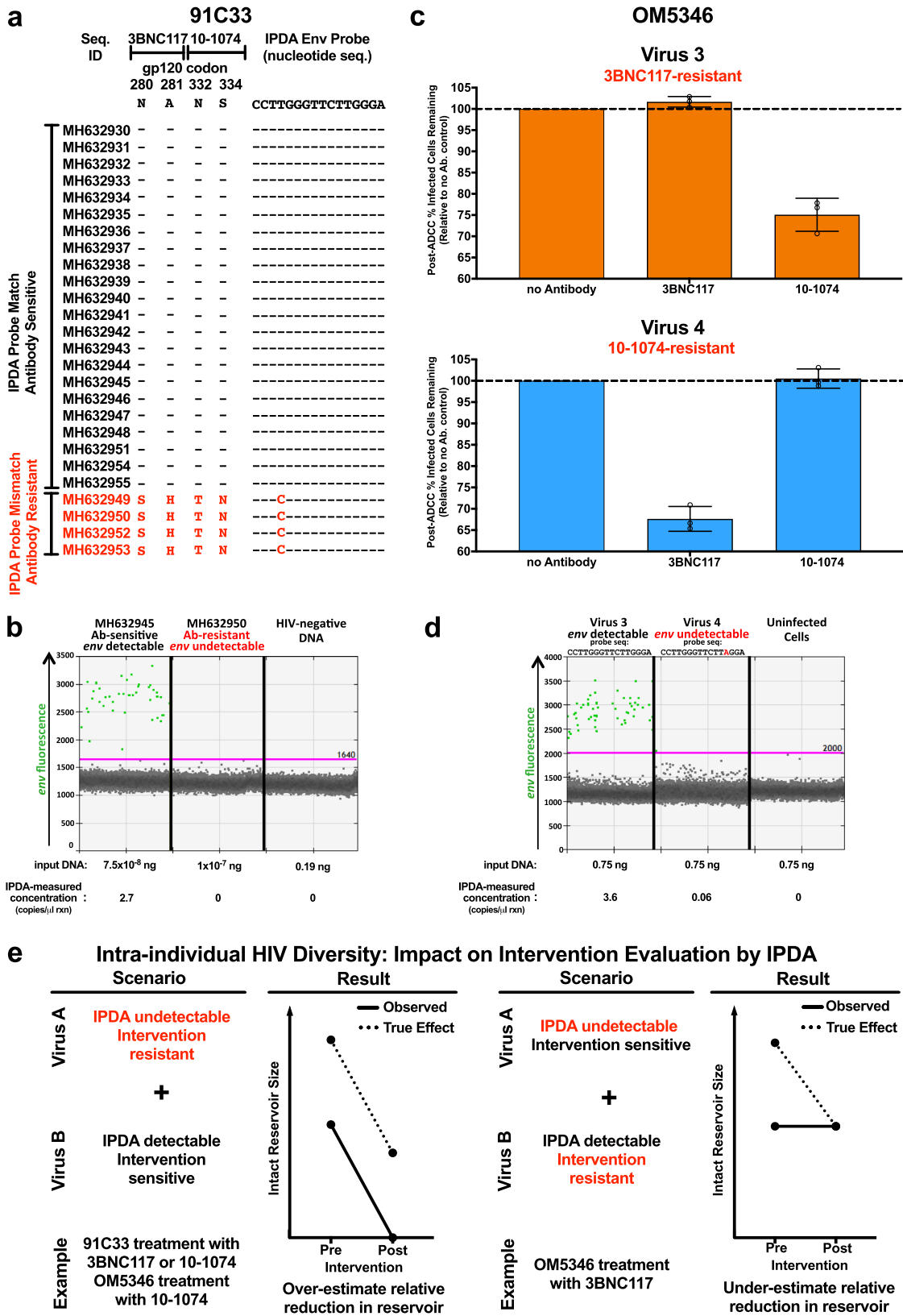
provides a hypothetical example. This individual did not respond to (off-ART) infusions of the bNAbs 3BNC117 and 10-1074 because they harbored a plasma HIV-1 subpopulation that was resistant to both bNAbs¹⁵ (Fig. 2a). Of note, this plasma HIV-1 subpopulation also harbored a mismatch to the IPDA *env* probe that we predicted would cause signal failure. We experimentally confirmed that, using the published conditions, the IPDA could not detect templates harboring this polymorphism (Fig. 2b), though signal could be detected if the annealing/extension temperature was lowered to 53 °C to account for the mismatch (Supplementary Fig. 3). In contrast, templates representing the bNAb-sensitive strains could be detected readily using the published conditions (Fig. 2b). Theoretically, if a person harboring such diversity in their reservoir were to be successfully treated with one or both bNAbs in a trial that used the IPDA as the sole readout, the intervention's effect on reducing the reservoir would be overestimated because the assay was only capable of detecting the susceptible reservoir fraction (Fig. 2e, left).

ART-suppressed participant OM5346 provides another example. Pre-ART HIV-1 drug resistance genotyping identified subtype B infection; however, single-genome sequencing of pre-ART plasma and proviruses sampled during long-term ART revealed co-infection with a non-B strain (Supplementary Fig. 4). While the IPDA is only designed for subtype B, the co-infecting strain harbored no critical mismatches within IPDA target regions except for a G-to-A mismatch at position 13 (G13A) of the *env* probe. Importantly, G13A is amongst the most frequent *env* probe polymorphisms in subtype B (and was observed in 2/10 study participants for whom *env* was sequenced), and the original report indicated that the IPDA was capable of detecting *env* sequences harboring it as the sole variant, at least when present on a plasmid template¹. As such, the assay should be able to detect this strain despite its non-B status. Using QVOA, we successfully isolated replication-competent subtype B (virus 3) and co-infecting (virus 4) strains from OM5346's reservoir, confirming both of these as true intended targets of the IPDA (Fig. 2c). We further observed that these viruses were differentially susceptible to 3BNC117- and 10-1074-mediated ADCC: while virus 3-infected cells could be eliminated by 10-1074-, but not 3BNC117-, mediated ADCC (25% and 0% relative reduction, respectively), the opposite was true for virus 4 (0% and 32% relative reduction, respectively) (Fig. 2c, Supplementary Fig. 5). We further confirmed that, while the IPDA could readily detect provirus 3 at concentrations approximating biological samples, the assay could not distinguish provirus 4 at these concentrations (Fig. 2d, Supplementary Fig. 6a). Similarly, at biological concentrations, the IPDA could not detect a synthetic DNA template encoding the virus 4 sequence, though noticeable signal

elevation occurred at ~100-fold higher template concentrations (Supplementary Fig. 6b). Importantly, a secondary *env* primer/probe set (see below and methods) readily detected both viruses at all concentrations, regardless of template type (Supplementary Fig. 6). In the hypothetical case where this individual were to be successfully treated with a bNAb in a trial that used IPDA as a sole readout, different erroneous interpretations could occur depending on the intervention used. Specifically, if this individual were to be successfully treated with 10-1074, the IPDA would overestimate the intervention's effect on the reservoir because the assay was only capable of detecting the susceptible reservoir fraction. In contrast, if the individual were to be successfully treated with 3BNC117, the IPDA would erroneously conclude that the intervention had no effect because the assay was only capable of detecting the resistant reservoir fraction (Fig. 2e, right).

Optical bleed-through as a source of error in the IPDA.

Although HIV-1 polymorphism can clearly cause the IPDA to fail in some instances, certain mismatches can be tolerated^{4,5}. This is due to the assay's use of endpoint (rather than real-time) PCR detection, such that polymorphisms that reduce signal can yield lower-amplitude droplets that can nevertheless still be scored as positive, provided that they are sufficiently separated from the negative population. In fact, participant OM5346's two viral populations illustrate this phenomenon: virus 3, which harbors a single nucleotide deletion as well as a mismatch in the Ψ probe, produces modest yet clearly discernible signal in the Ψ channel, whereas virus 4, which also harbors minor primer/probe mismatches, produces higher-amplitude Ψ signal (Fig. 3). Detailed investigation of such instances of HIV-1 sequence-driven amplitude modulations led us to identify another source of potential error, in the form of optical bleed-through, a challenge in multiplexed digital PCR assays¹⁶. This occurs when fluorescence from the Ψ channel, which features the FAM fluorophore, bleeds into the *env* channel, which features the VIC fluorophore, to varying extents. This in turn has implications for the appropriate placement of thresholds defining negative and positive populations. Critically, because the extent of bleed-through is determined by the amplitude of the Ψ signal, which is in turn determined by the extent of match to the assay probe, the extent of fluorescence spillover is HIV-1 sequence-specific. We demonstrate this by applying the IPDA to synthetic templates encoding the Ψ regions of OM5346 viruses 3 and 4, without a corresponding *env* template present (Fig. 3). Virus 3's Ψ signal, which is modest, does not bleed into the *env* channel. By contrast, virus 4's high amplitude Ψ signal bleeds noticeably into the *env* channel. Thus, whereas a tight threshold based on the negative



population can be drawn for virus 3, doing so for virus 4 creates *env* signal (and in fact intact provirus signal) even when no *env* template is present. This Ψ channel bleed-through for virus 4 thus likely contributed to the small amount of *env* signal in Fig. 2d (events adjoining the negative population). While ddPCR instruments are factory-tuned to mitigate spectral overlap, this basic property of

fluorophores underscores the instrument manufacturer's instruction that appropriate thresholds can only be drawn if negative and positive populations clearly separate from one another¹⁷. Similarly, it cautions against the setting of gates solely based on data from uninfected donors, as suggested by Peluso et al.⁴ as a strategy for suboptimal amplifications.

Fig. 2 Intra-individual HIV diversity can lead to reservoir underestimation by IPDA. **a** *Env gp120* amino acid and IPDA *env* probe nucleotide sequences from participant 91C33 identified by GenBank accession numbers. 91C33 harbored one HIV subpopulation that matched the IPDA *env* probe and was sensitive to bNAbs 3BNC117 and 10-1074, and another that harbored a mismatch at position 4 of the IPDA *env* probe and was resistant to both bNAbs. Hyphens (-) indicate matches to the reference; red letters denote mismatches (IPDA *env* probe) or amino acid substitutions (bNAbs). Sequence is abbreviated seq. **b** ddPCR *env* plots from one of 2 independent experiments assessing the ability of the IPDA *env* reaction to detect representative within-host sequences MH632945 (*env* probe match) and MH632950 (*env* probe mismatch), present as purified *env* PCR products of equal length and comparable quantities. Positive droplets are green and negative droplets are grey. See Supplementary Fig. 3 for additional experiments. **c** Virus 3, isolated from participant OM5346's reservoir, was sensitive to 10-1074-mediated ADCC but resistant to 3BNC117; vice-versa for virus 4. Box and error bars indicate mean and standard deviation of three technical replicates, from one of two independent experiments. Open circles denote individual datapoints. Source data are provided in the Source Data File. **d** 1D IPDA *env* ddPCR plots from one of 2 independent experiments for cells infected in vitro with virus 3 (IPDA *env* probe match) or virus 4 (G13A mismatch). Both viruses were detectable using an alternative primer/probe set (Supplementary Fig. 6). **e** Possible impacts of intra-individual HIV diversity on IPDA-measured reservoir changes following a hypothetical intervention. The solid line indicates the observed effect by IPDA; dashed line indicates the true effect. An intervention's impact on the reservoir could be overestimated when a subset of within-host HIV sequences are undetectable by IPDA and resistant to treatment (e.g., 91C33 during treatment with 3BNC117 or 10-1074, or OM5346 during treatment with 10-1074). Conversely, an intervention's impact on the reservoir could be underestimated when a subset of within-host HIV sequences are undetectable by IPDA and sensitive to treatment (e.g., OM5346 during treatment with 3BNC117). In a worst-case scenario, one could erroneously conclude that the treatment had no effect.

Addressing interhost HIV diversity in the IPDA. The IPDA offers major scalability and accessibility advantages over existing molecular¹⁸ or culture-based approaches², and a recent study that employed a quadruplexed qPCR assay confirmed that, of all probes evaluated, those used in the IPDA offered the greatest selectivity for intact proviruses¹⁸. Unfortunately, the requirement to discriminate proviral defects in the Ψ region restricts the options for placement and sequence of these oligonucleotides¹⁹, and the limited data currently available for Ψ region polymorphisms in IPDA failure cases (Fig. 1c and ref. 5) reveal few common polymorphisms that could be used to design secondary primer/probe sets. By contrast, placement of the 3' target is less constrained¹. This allowed us to develop a secondary primer/probe set in the intact-discriminating RRE region, approximately 50 bases downstream of the original location, as a first step towards addressing HIV-1 diversity with a strategy other than autologous oligonucleotide design^{3,5} or data exclusion⁷. This primer/probe set rescued detection of *env*-positive proviruses in 9/9 participants with IPDA *env* detection failure (Fig. 4a). When applied to the 36 participants for whom the IPDA detected *env*-positive proviruses (excluding OM5346 who harbored within-host diversity), it failed to detect *env*-positive proviruses in 3 (8%) individuals, indicating that it is not a universal solution. For the remaining 33 (92%), however, it yielded measurements that correlated strongly with those from the IPDA (Fig. 4b). The secondary primer/probe set could therefore be used to confirm instances of detection failure without confirmatory viral sequencing as well as to estimate intact provirus levels, though users must note that, unlike the IPDA *env* reaction, it cannot discriminate hypermutated sequences.

Discussion

The IPDA is undeniably a major advance in HIV-1 persistence research, but recent reports of variable and increasing failure rates due to HIV-1 polymorphism are concerning. While the original and early follow-up studies acknowledged HIV-1 diversity as a potential limitation that would require custom primer/probe sets, neither study reported any failures due to viral polymorphism (out of 62 and 81 participants, respectively)^{1,4}. More recent studies of 61 and 50 participants reported failure rates of 9.8% and 12%, respectively^{6,7}. The largest study undertaken to date using the IPDA, a multi-cohort study published by the assay's developers, reports an overall failure rate of 6.3% of 400 participants⁵, although it is important to note that this study includes a number of published cohorts including four with 0% failure rates^{4,8–10}. As such, the inferred failure rate for some of the unpublished

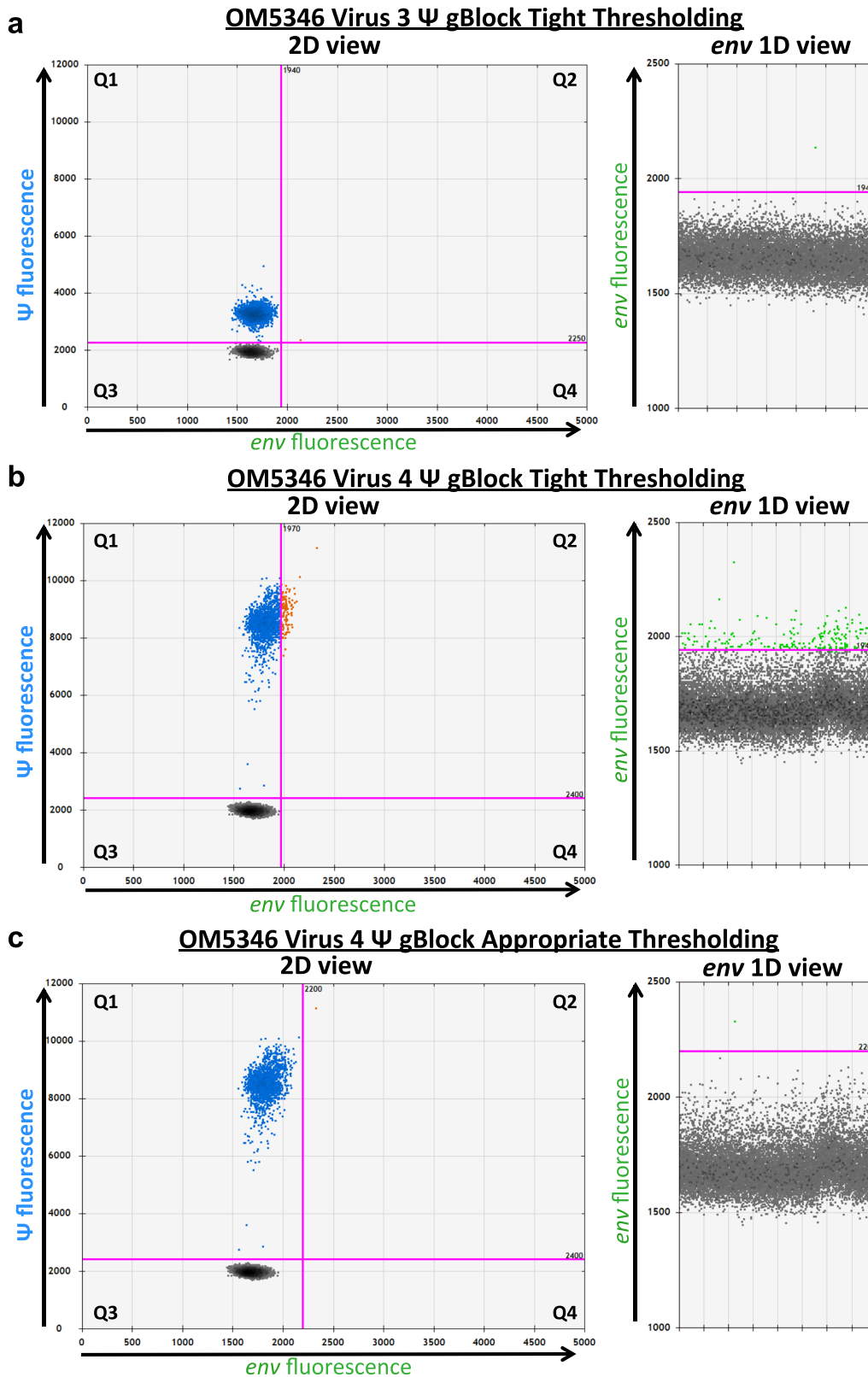
datasets could be as high as 18%, a value that approaches the 28% rate reported in the present study.

Taken together, our findings indicate that as the IPDA is applied to diverse cohorts, detection failures will occur, and not infrequently. The data also together suggest that failure rates may vary markedly by cohort, presumably due to local HIV-1 diversity in regions where HIV-1 subtype B circulates. Indeed, failure rates differed markedly between regions in our own study: whereas the failure rates in the Vancouver and Toronto cohorts were 2/15 (13%) and 3/19 (16%) respectively, those in the Washington, D.C. and Bronx, N.Y. cohorts were 4/7 (57%) and 3/4 (75%) respectively. Moreover, these failures were not attributable to shared HIV-1 polymorphisms nor to epidemiologic linkage (Fig. 1c), and all infections were HIV-1 subtype B.

The first step towards mitigating this issue is to raise awareness: samples that yield no Ψ or *env* single-positive proviruses should be flagged as unreportable until HIV-1 polymorphism has been addressed (e.g., using autologous primers/probes^{1,4}). Nevertheless, elevated failure rates limit the assay's scalability and utility: autologous probe design is time-consuming and costly, and exclusion of assay failures could bias study results if failure rates are substantial. Our findings therefore underscore the importance of better understanding the frequency and nature of HIV-1 polymorphisms within IPDA target regions and their impact on intact proviral quantification. Analysis of HIV-1 subtype B sequences from unique individuals in the Los Alamos HIV database revealed that 23% of 9360 *env* sequences harbored at least one *env* probe mismatch (which is similar to our *env* detection failure rate of 9/46 or 20%), while 50% of 1489 sequences harbored at least one Ψ probe mismatch (which is substantially greater than our Ψ detection failure rate of 5/46 or 11%). This suggests that the *env* reaction may be more sensitive to polymorphism than the Ψ reaction, but larger studies linking proviral sequence to IPDA readouts are needed to confirm this.

Similarly, due to a paucity of studies that pair detailed within-host sequence characterization with IPDA measurements, it is not currently possible to estimate the extent and impact of within-host HIV-1 diversity on intact reservoir quantification by the IPDA. Nevertheless, when taken together with Gaebler et al.'s¹⁸ observation that four of nine (44%) studied reservoirs were heterogeneous in an IPDA probe region, our combined observations suggest that the impact of within-host HIV-1 diversity on IPDA accuracy may be non-negligible.

In conclusion, we provide an independent assessment of HIV-1 sequence diversity as a source of error in the IPDA by integrating IPDA, QVOA, and viral sequence data to characterize detection



failures. We report a substantially higher rate of complete detection failures than previous studies, and additionally identify partial detection failures, attributable to within-host viral variation, as a source of error in the IPDA. The latter issue is particularly challenging given that it is essentially impossible to detect without HIV-1 sequencing, but given the reservoir's highly dynamic nature²⁰ and our identification of cases where within-host diversity is linked to

HIV-1 cure intervention susceptibility, it should not be ignored. Given the clear value of the IPDA, iterative efforts to refine this assay are a priority. In the meantime, we recommend that proviral sequencing, at minimum within the IPDA target regions, should be performed for participants of HIV-1 remission clinical trials in which the IPDA is employed, and secondary and/or autologous primers/probes used to mitigate HIV-1 diversity.

Fig. 3 HIV sequence-specific optical bleed-through prohibits tight thresholding on negative populations. **a** 2D and *env* 1D plots of OM5346 virus 3 Ψ region sequence when tested as a synthetic DNA gene fragment (Virus 3 Ψ gBlock) without corresponding *env* template. Minimal Ψ- to *env*- channel spillover occurs and the positive droplet threshold can be drawn tightly to the double-negative population without consequence (note the presence of one false double-positive droplet). **b** 2D and *env* 1D plots of OM5346 virus 4 Ψ region sequence, tested as a synthetic DNA gene fragment (Virus 4 Ψ gBlock), without corresponding *env* template. Drawing a tight threshold causes optical bleed-through of Ψ (FAM) fluorescence into the *env* (VIC) channel to yield a false-positive *env* (and by extension, false-positive intact) signal. **c** 2D and *env* 1D plots of OM5346 Virus 4 Ψ region sequence, tested as a synthetic DNA gene fragment (Virus 4 Ψ gBlock) without corresponding *env* template, with a threshold drawn at an appropriate distance from the double-negative population. This threshold accommodates the fluorescence shift and thus avoids the creation of false-positive intact or *env*-positive droplet population (note the presence of a single false double-positive droplet). Representative plots from one of three technical replicates from one experiment are shown.

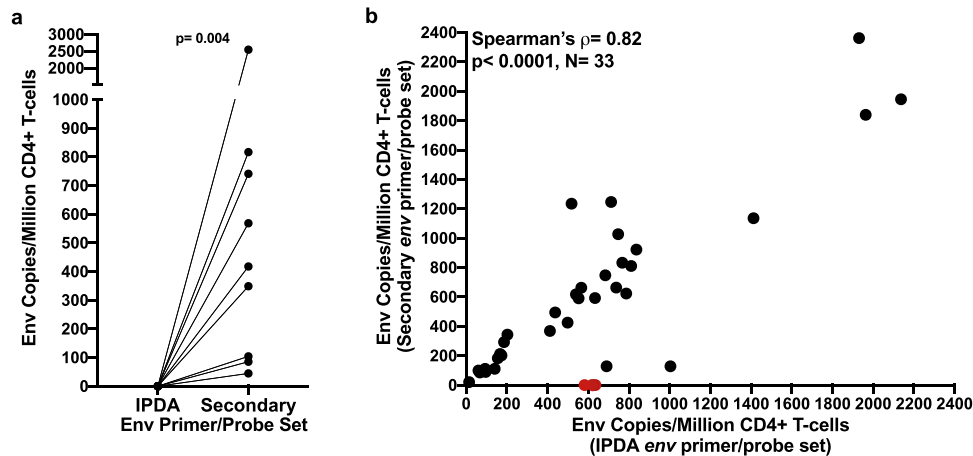


Fig. 4 Performance of the secondary *env* primer/probe set. **a** A secondary *env* primer/probe set rescued detection in all 9 cases of IPDA *env* detection failure ($p = 0.004$, two-sided Wilcoxon signed-rank test). **b** A secondary *env* primer/probe set performs comparably to the IPDA in samples detectable by the latter assay. Black datapoints denote 33 participants whose reservoir was detectable by both IPDA and secondary *env* primer/probe sets; red datapoints denote three participants whose reservoir was detectable by IPDA but not using the secondary *env* primer/probe set. Excluding the latter three points yielded a Spearman's $\rho = 0.82$, $p < 0.0001$ (shown on figure). Including these three red datapoints yielded a Spearman's $\rho = 0.68$, $p < 0.0001$. Datapoints show point estimate from four merged technical replicates for each participant sample. Source data are provided in the Source Data File.

Methods

Participants and ethics statement. We studied 46 individuals with HIV-1 subtype B who were recruited to cohorts in Toronto ($N = 19$), Vancouver ($N = 15$), New York City ($N = 4$), Washington, D.C. ($N = 7$), and Mexico City ($N = 1$). For Vancouver participants, peripheral blood mononuclear cells (PBMCs) were isolated by density gradient separation and cryopreserved (-150°C , 90% Fetal Bovine Serum + 10% DMSO). Participants recruited at other sites provided a leukapheresis sample from which PBMC were isolated and cryopreserved as above. All participants were on long-term, virally suppressive combination antiretroviral therapy (cART) at time of sampling, with the exception of one elite controller. Participants' duration of untreated HIV-1 infection ranged from 1 month to >10 years, though the exact duration was unknown for most participants. Ethical approval to conduct this study was obtained from the Institutional Review Boards of Simon Fraser University, Providence Health Care/University of British Columbia, Weill Cornell Medicine, and the George Washington University. All participants provided written informed consent.

Quantitative Viral Outgrowth Assay (QVOA). For participants for whom sufficient biological material was available, the Quantitative Viral Outgrowth Assay (QVOA) was performed². CD4 + T-cells were isolated from PBMCs by negative selection and plated in serial dilution at either 4 or 6 concentrations (12 replicate wells/concentration, 24-well plates). CD4 + T-cells were stimulated with phytohemagglutinin (PHA, $2\ \mu\text{g}/\text{mL}$) and irradiated allogeneic HIV-1-negative PBMCs were added to further induce viral reactivation. MOLT-4/CCR5 cells were added at 24-h post-stimulation as targets for viral infection. Culture media (RPMI 1640 + 10% FBS + 1% Pen/Strep + 50 U/mL IL-2 + 10 ng/mL IL-15) was changed every 3 days and p24 enzyme-linked immunosorbent assay (ELISA) was run on day 14 to identify virus-positive wells. Infectious Units per Million CD4 + T-cells (IUPM) was determined using the Extreme Limiting Dilution Analysis (ELDA) software (version 1.0, <http://bioinf.wehi.edu.au/software/elda/>)²¹. Culture supernatants from virus-positive wells were frozen (-80°C) for future use.

Intact Proviral DNA Assay (IPDA). Genomic DNA was isolated from a median 4.5 (interquartile range [IQR] 4–5) million CD4 + T-cells using the QIAamp DNA Mini Kit (Qiagen) with precautions to minimize DNA shearing. Intact HIV-1 copies/million CD4 + T-cells were determined by droplet digital PCR (ddPCR)

using the Intact Proviral DNA Assay (IPDA)¹, where HIV-1 and human RPP30 reactions were conducted independently in parallel and copies were normalized to the quantity of input DNA. In each ddPCR reaction, a median 7.5 ng (IQR 7–7.5 ng) (RPP30) or a median 750 ng (IQR 700–750 ng) (HIV-1) of genomic DNA was combined with ddPCR Supermix for Probes (no dUTPs, BioRad), primers (final concentration 900 nM, Integrated DNA Technologies), probe(s) (final concentration 250 nM, ThermoFisher Scientific) and nuclease free water. Primer and probe sequences ($5' \rightarrow 3'$) were: RPP30 Forward Primer- GATTTGGACCTGCGA GCG, RPP30 Probe- VIC-CTGACCTGAAGGCTCT- MGBNFQ, RPP30 Reverse Primer- GCGGCTGTCTCCACAAGT; RPP30-Shear Forward Primer- CCATTTGCTGCTCCTTGGG, RPP30-Shear Probe- FAM- AAGGAGCA AGTTTCTATTGTAG- MGBNFQ, RPP30-Shear Reverse Primer- CATGCA AAGGAGGAAGCCG; HIV-1 Ψ Forward Primer- CAGGACTCGGCTTGCTGA AG, HIV-1 Ψ Probe- FAM- TTTTGGCGTACTACCAGT- MGBNFQ, HIV-1 Ψ Reverse Primer- GACCCATCTCTCTCCTTCTAGC; HIV-1 *env* Forward Primer- AGTGGTGCAGAGAGAAAAAGAGC, HIV-1 *env* Probe- VIC-CCTTGGGTTCTTGGGA- MGBNFQ, anti-Hypermutant *env* Probe- CCTTAGGTTCTTAGGAGC- MGBNFQ, HIV-1 *env* Reverse Primer- GTCTGGCCTGTACCCTGAGC. For participant BC-004, an autologous *env* probe was designed by modifying the published IPDA *env* probe to match the participant's sequence (VIC-CCTTGGGTTTCTGGGA- MGBNFQ). Droplets were prepared using either the Automated or QX200 Droplet Generator (BioRad) and cycled at 95°C for 10 min; 45 cycles of (94°C for 30 sec, 59°C for 1 min) and 98°C for 10 min¹. Droplets were analyzed on a QX200 Droplet Reader (BioRad) using QuantaSoft software (BioRad, version 1.7.4), where replicate wells were merged prior to analysis. Four technical replicates were performed for each participant sample, where a median (IQR) 243,100 (106,200–265,950) cells were assayed in total. Intact HIV-1 copies (Ψ and *env* double-positive droplets) were corrected for DNA shearing based on the frequency of RPP30 and RPP30-Shear double-positive droplets. The median (IQR) DNA shearing index (DSI), measuring the proportion of sheared DNA in a sample, was 0.31 (0.28–0.35), highly comparable to that reported by the original authors¹. For the experiments that evaluated the IPDA's ability to detect the specific *env* sequences harbored by participants 91C33 (Fig. 2b, Supplementary Fig. 3) and OM5346 (Supplementary Fig. 6b), synthetic templates (purified PCR amplicons for 91C33 and commercially-synthesized gBlocks [IDT] for OM5346) were used as targets. Total HIV-1 *gag* copies/million CD4 + T-cells¹¹ (Fig. 1e) were measured in reactions prepared using RPP30, as described above, and the following HIV-1-specific primers

and probe (5'→3'): HIV-1 *gag* Forward—TCTCGACGCGAGGACTCG; HIV-1 *gag* Reverse—TACTGACGCTCTCGCACG; HIV-1 *gag* Probe—FAM—CTCTCTCTCT/ZEN/TCTAGCCTC-3IABkFQ (IDT). Reactions were cycled at: 95 °C for 10 min; 40 cycles of (94 °C for 30 sec, 53 °C for 1 min) and 98 °C for 10 min. A minimum of four (maximum eight) technical replicates were performed for each sample, where data from replicate wells was merged prior to analysis.

HIV-1 proviral amplification and sequencing. Single-template, near-full-length proviral amplification was performed on DNA extracted from CD4+ T-cells by nested PCR using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) such that ~25% of the resulting PCR reactions yielded an amplicon. First round primers were: Forward—AAATCTCTAGCAGTGGCGCCGAACAG, Reverse—TGAGGG ATCTCTAGTTACCAGAGTC. Second round primers were: Forward—GC GCCCGAACAGGGACYTGAARCGAAAG, Reverse—GCACTCAAGGCA AGCTTTATTGAGGCTTA. Reactions were cycled as follows: 92 °C for 2 min; 10 cycles of (92 °C for 10 sec, 60 °C for 30 sec, and 68 °C for 10 min); 20 cycles of (92 °C for 10 sec, 55 °C for 30 sec, and 68 °C for 10 min); 68 °C for 10 min^{22,23}. Amplicons were sequenced using Illumina MiSeq technology and de novo assembled using the custom software MiCall (<https://github.com/cfe-lab/MiCall>) which features an in-house modification of the Iterative Virus Assembler (IVA)²⁴, or through collaboration with the Massachusetts General Hospital CCIB Core.

In silico predicted IPDA results. Linked Ψ and *env* probe region sequences were excised from BC-004 proviral sequences for in silico prediction of IPDA results under the assumption that intact probe regions would be detected by the assay. A Ψ or *env* probe region sequence was considered defective if it contained mutations associated with defective proviruses (e.g., those consistent with hypermutation¹ or common splice donor site mutations¹⁹) or if it was completely absent. All other probe region sequences, regardless of whether they matched the published IPDA probe sequence, were considered intact.

Targeted HIV-1 sequencing. For participant OM5346, HIV-1 RNA was extracted from QVOA outgrowth viruses and pre-ART plasma, after which gp160 (QVOA) and gp120/Pol (plasma) were amplified from endpoint-diluted templates by nested RT-PCR using HIV-1-specific primers and high fidelity enzymes. For participants with suspected IPDA detection failure, and for whom HIV-1 sequences were not already available, the IPDA Ψ and/or *env* amplicon regions were bulk-amplified from extracted proviral DNA using HIV-1-specific primers. Amplicons were sequenced using either Sanger (3730xl, Applied Biosystems) or Next Generation (Illumina MiSeq) technologies. Sanger chromatograms were analyzed using Sequencher (version 5.0.1, Gene Codes), while Illumina MiSeq reads were de novo assembled as described above.

Antibody-dependent cellular cytotoxicity (ADCC) assays. Total CD4+ T-cells were isolated, as described above, from HIV-1-negative donors and activated with anti-CD3 (Biolegend, catalogue number: 317326, clone: OKT3, 1:1000 dilution) and anti-CD28 antibodies (Biolegend, catalogue number: 302934, clone: CD28.2, 1:1000 dilution). Cells were then infected with the participant virus of interest collected from QVOA supernatant, and monitored by flow cytometry by intracellular staining for HIV-1-Gag (Beckman Coulter, catalogue number: 6604665, clone: KC57-RD1, 1:100 dilution) until cells were >5% Gag+. CD4+ T-cells were washed and incubated with the broadly neutralizing antibody of interest (either 3BNC117 or 10-1074, NIH AIDS Reagent Program, catalogue numbers: 12474 and 12477, respectively, 10 µg/mL) for 2 h. Natural Killer (NK) cells were negatively selected (EasySep, StemCell Technologies) from PBMCs of allogeneic, HIV-1-negative donors and activated using interleukin-15 (IL-15) provided by the National Cancer Institute Biological Research Branch. Activated NK cells were then co-cultured with the infected, antibody-treated CD4+ T-cells for 16 h at an effector-to-target (E:T) ratio of 1:1. Following co-culture, cells were stained with fluorophore-conjugated antibodies against human IgG (Southern BioTech, Catalogue number: 2040-31, 1:200 dilution), CD3 (Biolegend, catalogue number: 344842, clone SK7, 1:200 dilution), CD56 (Biolegend, catalogue number: 318334, clone: HCD56, 1:200 dilution), and CD4 (Biolegend, catalogue number: 200521, clone: RPA-T4, 1:200 dilution), as well as intracellular Gag (as above) and LIVE/DEAD Fixable Aqua Stain amine-reactive dye (Invitrogen). The percentages (%) of infected cells remaining post-ADCC relative to no antibody control were then determined using the following formula: $(1 - ((\%Gag+ \text{ cells among viable CD3} + \text{ cells in no antibody condition}) - (\%Gag+ \text{ cells among viable CD3} + \text{ cells in test condition})) / (\%Gag+ \text{ cells amongst viable CD3} + \text{ cells under no antibody condition})) \times 100$. Two independent experiments were conducted, each with three technical replicates. Data were collected using Attune™ NxT software (ThermoFisher Scientific, version 2.7) and analyzed using FlowJo (Becton Dickinson, version 10.6).

Within-host HIV-1 phylogenetic analysis. For participant OM5346, single-genome HIV-1 amplification and sequencing was performed from pre-ART plasma (sampled in 2012), proviruses from purified CD4+ T-cells isolated during suppressive cART (sampled in 2017 and 2019), and replication-competent HIV-1 strains isolated by QVOA (from the 2017 sample) as described above. The 2012

and 2017/2019 samples were shipped directly to, and amplified in, two separate laboratories. A plasma HIV-1 *pol* sequence derived from clinical drug resistance testing in 2012 was also incorporated into the analysis. *pol* and *gp120* sequences were specifically amplified or retrieved from near-full genome proviral sequences using GeneCutter (version 1, https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). Sequences were multiply-aligned using MAFFT (version 7.427)²⁵ in a codon-aware manner and inspected in AliView (version 1.19)²⁶. Maximum-likelihood phylogenies were inferred using PhyML (version 3.0)²⁷ under a General Time Reversible nucleotide substitution model and visualized using FigTree (version 1.3.1).

Design of secondary *env* ddPCR primer/probe set. HIV-1 Subtype B *env* envelope sequences, limited to one per individual, were retrieved from the HIV LANL database ($N = 4670$, <https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) and aligned. Sequence conservation was assessed across 16–25 base-pair intervals, and three intervals that maximized sequence conservation while meeting ddPCR assay specifications (i.e., amplicon length, T_m) were chosen as secondary primer and probe locations. These were: Secondary *env* Forward Primer ACTATGGGCGCAGCGTC (representing nucleotides 7809–7825 in the HIV-1 genomic reference HXB2; predicted sequence conservation 79%), Secondary *env* Probe VIC-CTGGCCTGTACCGTCAG- MGBNFQ (HXB2 nucleotides 7833–7849, 83% conservation), Secondary *env* Reverse Primer CCCCAGA CTGTGAGTTGCA (HXB2 nucleotides 7939–7921, 88% conservation). Reaction composition and cycling conditions were same as those used for the IPDA as described above. Accurate quantification using the Secondary *env* primer/probe set was verified using DNA extracted from the J-Lat 9.2 cell line (obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, contributed by Dr. Eric Verdin)²⁸ (Supplementary Fig. 7), where three independent experiments, each with four technical replicates, were performed and replicate wells were merged prior to data analysis.

Statistical analyses. All statistical analyses were performed using GraphPad Prism (version 8). All statistical tests were two-sided.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data that support the findings of this study are available in the manuscript, Figures and Supplementary Figures. HIV sequences used for phylogenetic inference are available in GenBank under accession numbers MT792083–MT792233; HIV sequences for participant 91C33 were previously published under GenBank accession numbers MH632930–MH632955¹⁵. HIV-1 subtype B sequences used to identify the location for the secondary *env* primer/probe set are available through the Los Alamos National Laboratory HIV sequence database (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>). Additional information is available from the corresponding authors upon request. Source data are provided with this paper.

Code availability

Code for the custom software MiCall is available at <https://github.com/cfe-lab/MiCall>.

Received: 1 September 2020; Accepted: 2 December 2020;

Published online: 08 January 2021

References

- Bruner, K. M. et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120–125 (2019).
- Laird, G. M. et al. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398 (2013).
- Antar, A. A. et al. Longitudinal study reveals HIV-1-infected CD4+ T cell dynamics during long-term antiretroviral therapy. *J. Clin. Invest.* **130**, 3543–3559 (2020).
- Peluso, M. J. et al. Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI insight* <https://doi.org/10.1172/jci.insight.132997> (2020).
- 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. *Proc. Natl Acad. Sci. USA* **117**, 18692–18700 (2020).
- Falcinelli, S. D. et al. Impact of biological sex on immune activation and frequency of the latent HIV reservoir during suppressive antiretroviral therapy. *J. Infect. Dis.* <https://doi.org/10.1093/infdis/jiaa298> (2020).
- Gandhi, R. T. et al. Selective decay of intact HIV-1 proviral DNA on antiretroviral therapy. *J. Infect. Dis.* <https://doi.org/10.1093/infdis/jiaa532> (2020).

8. Papasavvas, E. et al. Intact HIV reservoir estimated by the intact proviral DNA assay correlates with levels of total and integrated DNA in the blood during suppressive antiretroviral therapy. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciaa809> (2020).
9. Garliss, C. C., Kwaa, A. K. & Blankson, J. N. A comparison of different immune activation strategies to reverse HIV-1 latency. *Open Forum Infect. Dis.* **7**, ofaa082 (2020).
10. Kwaa, A. K., Garliss, C. C., Ritter, K. D., Laird, G. M. & Blankson, J. N. Elite suppressors have low frequencies of intact HIV-1 proviral DNA. *AIDS* **34**, 641–643 (2020).
11. Huang, S. H. et al. Latent HIV reservoirs exhibit inherent resistance to elimination by CD8⁺ T cells. *J. Clin. Invest.* **128**, 876–889 (2018).
12. Bruel, T. et al. Elimination of HIV-1-infected cells by broadly neutralizing antibodies. *Nat. Commun.* **7**, 10844 (2016).
13. Mendoza, P. et al. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* **561**, 479–484 (2018).
14. ClinicalTrials.gov. *ACTG A5386 (NCT04340596) Study Details.* <https://clinicaltrials.gov/ct2/show/study/NCT04340596> (2020).
15. Bar-On, Y. et al. Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in viremic individuals. *Nat. Med.* **24**, 1701–1707 (2018).
16. Whale, A. S., Huggett, J. F. & Tzovev, S. Fundamentals of multiplexing with digital PCR. *Biomol. Detect. Quantif.* **10**, 15–23 (2016).
17. BioRad-Laboratories-Inc. *Droplet Digital PCR Applications Guide*, https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf (2019).
18. Gaebler, C. et al. Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. *J. Exp. Med.* **216**, 2253–2264 (2019).
19. Ho, Y. C. et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
20. Wang, Z. et al. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proc. Natl Acad. Sci. USA* **115**, E2575–E2584 (2018).
21. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
22. Li, B. et al. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J. Virol.* **81**, 193–201 (2007).
23. 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).
24. Hunt, M. et al. IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* **31**, 2374–2376 (2015).
25. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
26. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
27. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
28. Jordan, A., Bisgrove, D. & Verdin, E. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**, 1868–1877 (2003).

Acknowledgements

This work was supported by the Martin Delaney 'BELIEVE' Collaboratory (NIH grant 1UM1AI26617), which is supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, and OAR. It was also supported in part by the NIH funded R01 grants AI31798 and AI147845 (to R.B.J.) and Canadian Institutes of Health Research (CIHR) project grant

PJT-159625 (to Z.L.B.) and HB1-164063 (to Z.L.B. and M.A.O.). N.N.K. is supported by a CIHR Vanier Doctoral Award. A.S. is supported by a CIHR Doctoral Award. Z.L.B. is supported by a Scholar Award from the Michael Smith Foundation for Health Research. We thank the National Cancer Institute (NCI) Biological Research Branch for provision of IL-15 and IL-2, the NCI's AIDS and Cancer Virus Program for provision of HIV-1 p24 ELISA reagents, and the NIH AIDS Research and Reference Reagent Program for provision of the CCR5⁺ MOLT-4 cells. We thank Marina Caskey for providing de-identified samples from ART-naïve donors used in this study. We thank the BC Centre for Excellence in HIV/AIDS for support. We thank the Massachusetts General Hospital Center for Computational & Integrative Biology DNA Core, specifically Nicole Stange-Thomann, Amy Avery, Kristina Belanger, and Huajun Wang for providing us with the Illumina MiSeq deep sequencing service used in this manuscript. We thank Bruce Ganase, Jeff Knaggs, Daniel MacMillan, Hanwei Sudderuddin and James Nakagawa for research and bioinformatics support and Gursev Anmole, Mark Brockman and Robert Holt for helpful discussions. We thank Michel Nussenzeig, Christian Gaebler and Yotam Bar-On for provision of materials and helpful discussions. We gratefully acknowledge the contributions of the study participants, without whom this work would not be possible.

Author contributions

N.N.K., Y.R., Z.L.B., and R.B.J. designed the study. N.N.K., Y.R., W.D.C.A., W.D., P.K., S.H.H., T.M.M., A.S., and A.W. performed experiments. N.N.K., Y.R., W.D.C.A., W.D., P.K., S.H.H., T.M.M., A.S., A.W., D.K., C.J.B., G.Q.L., and R.M.L. analyzed data. M.H., C.K., E.B., M.A.O., P.M.D.R.E., A.W., C.C., W.D.H., L.M., and H.G. provided participant samples. N.N.K., Y.R., Z.L.B., and R.B.J. wrote the manuscript. All authors contributed to the critical revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-20442-3>.

Correspondence and requests for materials should be addressed to Z.L.B. or R.B.J.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021