RESEARCH ARTICLE

A spatio-temporal assessment of simian/ human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host

Alison F. Feder[1], Christopher Kline[2], Patricia Polacino[3], Mackenzie Cottrell[4], Angela D. M. Kashuba[4], Brandon F. Keele[5], Shiu-Lok Hu[3], Dmitri A. Petrov[1], Pleuni S. Pennings[6]*, Zandrea Ambrose[2]*

1 Department of Biology, Stanford University, Stanford, CA, United States, 2 Division of Infectious Diseases, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States, 3 Washington National Primate Research Center, University of Washington, Seattle, WA, United States, 4 Division of Pharmacotherapy and Experimental Therapeutics, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, 5 AIDS and Cancer Virus Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, United States, 6 Department of Biology, San Francisco State University, San Francisco, CA, United States

* pennings@sfsu.edu (PSP); zaa4@pitt.edu (ZA)

Abstract

The process by which drug-resistant HIV-1 arises and spreads spatially within an infected individual is poorly understood. Studies have found variable results relating how HIV-1 in the blood differs from virus sampled in tissues, offering conflicting findings about whether HIV-1 throughout the body is homogeneously distributed. However, most of these studies sample only two compartments and few have data from multiple time points. To directly measure how drug resistance spreads within a host and to assess how spatial structure impacts its emergence, we examined serial sequences from four macaques infected with RT-SHIV$_{mne027}$, a simian immunodeficiency virus encoding HIV-1 reverse transcriptase (RT), and treated with RT inhibitors. Both viral DNA and RNA (vDNA and vRNA) were isolated from the blood (including plasma and peripheral blood mononuclear cells), lymph nodes, gut, and vagina at a median of four time points and RT was characterized via single-genome sequencing. The resulting sequences reveal a dynamic system in which vRNA rapidly acquires drug resistance concomitantly across compartments through multiple independent mutations. Fast migration results in the same viral genotypes present across compartments, but not so fast as to equilibrate their frequencies immediately. The blood and lymph nodes were found to be compartmentalized rarely, while both the blood and lymph node were more frequently different from mucosal tissues. This study suggests that even oft-sampled blood does not fully capture the viral dynamics in other parts of the body, especially the gut where vRNA turnover was faster than the plasma and vDNA retained fewer wild-type viruses than other sampled compartments. Our findings of transient compartmentalization across multiple tissues may help explain the varied results of previous compartmentalization studies in HIV-1.

Author summary

HIV-1 is difficult to treat because the virus can evolve to become drug-resistant within the body, but we have an incomplete understanding of where drug-resistant viruses originate and how they spread within a person. In this study, four macaques were infected with RT-SHIV, a simian immunodeficiency virus with an HIV-1 reverse transcriptase coding region, which can be targeted with standard HIV drugs. We sampled virus from the macaques before, during and after their viral population became resistant to administered drugs and determined the genetic viral sequences in several parts of the body: blood, lymph nodes, gut, and vagina. We found that drug resistance emerged across compartments nearly simultaneously, and drug resistance evolved multiple independent times within each macaque. Although migration of RT-SHIV between compartments is fast, compartments do not have the same distribution of viral genotypes. This is important because although studies typically sample virus from the blood to study how HIV-1 evolution in humans, our study suggests that it is not fully representative of other parts of the body, particularly the gut.

Introduction

The development of drug resistance in human immunodeficiency virus type 1 (HIV-1) to anti-retroviral therapies represents a major public health obstacle. The prevalence of drug resistance in many regions of the world is now above 5% and increasing, driven in part by drug resistance transmitted from one infected individual to another [1]. Transmitted drug resistance has been observed at frequencies between 2.8% and 11.5% in different regions of the world, rendering standard first line therapies ineffective and necessitating costlier therapies for treatment-naïve patients. Despite the increasing prevalence of drug resistance, we still have an incomplete understanding of how drug resistance emerges and spreads spatially within an individual.

A comprehensive understanding of the drug resistance spatial landscape, such as when and where drug resistance arises and how it spreads within a patient, could inform administration practices of current therapy and the design of new antiretroviral therapies. Previous studies have suggested that HIV-1 drug resistance originates primarily in tissues with low or absent drug concentration (i.e., drug sanctuaries) [2–4]. If this holds true, then efforts should focus on improving drug penetration in these tissues.

Understanding how drug resistance spreads within the body could improve understanding of the ecology of the virus within a patient, which, in turn, could help determine the best way to sample the viral population, monitor transmission chains, and design eradication strategies. Although nearly all sampling of HIV-1 is done in the blood plasma, it is unclear how well the plasma represents the broader HIV-1 population. Differences between viruses from plasma and tissues could drive misdiagnoses of pre-existing drug-resistant variants, resulting in incorrectly applied therapies and selection for further drug resistance. If virus in the genital tract forms a different population than virus circulating in the blood, this should be taken into account when sampling in order to allow better reconstruction of transmission chains and epidemiological tracking of transmitted drug resistance. Finally, understanding HIV-1 genetic variation across the body might contribute to our understanding of viral reservoirs and the potential for virus eradication.

Table 1. Summary of results from a subset of HIV-1 compartmentalization studies.

|	Evidence mostly against compartmentalization	Evidence both for and against compartmentalization	Evidence mostly for compartmentalization
Plasma or PBMCs v. Lymph Node	Josefsson et al., 2013 [19]; Kearney, et al., 2015 [18]; Boritz et al., 2016 [20]	Vanderford et al., 2011 [12]	Haddad et al., 2000 [21]
Plasma or PBMCs v. Female Genital Tract		Overbaugh, et al, 1996 [22]; Si-Mohamed, et al., 2000 [23]; Bull, et al., 2009 [24]; Kelley, et al., 2010 [6]; Chaudhary et al., 2011 [25]; Bull, et al., 2013 [5]	Poss et al., 1995 [26]; Poss, et al., 1998 [27]; Kemal, et al., 2003 [28]; Tirado, et al., 2004A [29]; Tirado, et al., 2004B [30]; Philpott, et al., 2005 [31]
Plasma or PBMCs v. Gut	Imamichi, et al., 2011 [16]; Lerner, et al., 2011 [PBMCs v. gut] [32]; Vanderford et al., 2011 [PBMCs v. gut] [12]; Evering, et al., 2012 [33]; Kearney, et al., 2015 [18]	Barnett et al., 1991 [34]; Zhang, et al., 2002 [35]; Van Marle, et al., 2010 [10]; Lewis, et al., 2013 [15]; Rozera, et al 2014 [11]; Vanderford et al., 2011 [Plasma v. gut] [12]	van Marle, et al., 2007 [17]; Lerner, et al., 2011 [Plasma v. gut] [32]
Among different T cell populations and/or plasma	Heeregrave et al 2003 [13]	Potter, et al., 2003 [36]; Potter, et al., 2004 [37]; Fulcher, et al., 2004 [7]; Delobel, et al., 2005 [9]; Josefsson et al., 2013 [19]; Boritz et al., 2016 [20]	Wang, et al., 2000 [14]; Potter, et al., 2006 [8]; Rozera, et al., 2014 [11]

It has long been recognized that it is important to understand the relationship between virus in the blood and different tissues and many studies have examined this. However, the results are equivocal and some studies find evidence for compartmentalization, whereas others do not (see Table 1 for a summary of some of the major studies which examine the same tissues as this study). Nearly all studies that compared compartments across time found that compartments were only different transiently [5–12]. Many of these studies with multiple time points examined the relationship among different T cell subsets in the blood [7–9, 13, 14], but there is recent interest in examining temporal data in tissues as well [5, 6, 10, 12, 15]. Many studies that have been conducted focused on the comparison between blood and only one type of tissue (though see [10, 12, 16–18]), and thus cannot give us information on the relationships between several compartments. Taken together, previous studies suggest the presence of a complicated spatial landscape on which viral evolution occurs, and much still needs to be done to connect not only how other compartments relate to the blood plasma, but also how they relate to each other.

To increase our understanding of how drug resistance spreads through space and how different compartments in a host are connected to each other, we obtained comprehensive single-genome sequence (SGS) data from four different compartments and the blood, at four (median) time points, in four macaques infected with a simian immunodeficiency virus containing HIV-1 reverse transcriptase (RT-SHIV) [38, 39]. The macaques were treated with monotherapy to induce the emergence of two of the most prevalent drug resistance mutations (DRMs) in HIV-1 reverse transcriptase (RT): K103N and M184V [1]. We sampled viral RNA (vRNA) from plasma and four different compartments (peripheral blood mononuclear cells, lymph node, vagina, and gut) before, during, and after the detection of drug-resistant variants in the blood. Like HIV-1, RT-SHIV is a lentivirus that produces a double-stranded DNA genome after infection of host cells, and we also sampled reverse transcribed viral DNA (vDNA) from the same four compartments.

Results

Experimental design

Four macaques were infected with RT-SHIV. To intentionally select drug-resistant virus, macaques were treated with daily monotherapy beginning at 12 weeks post-infection
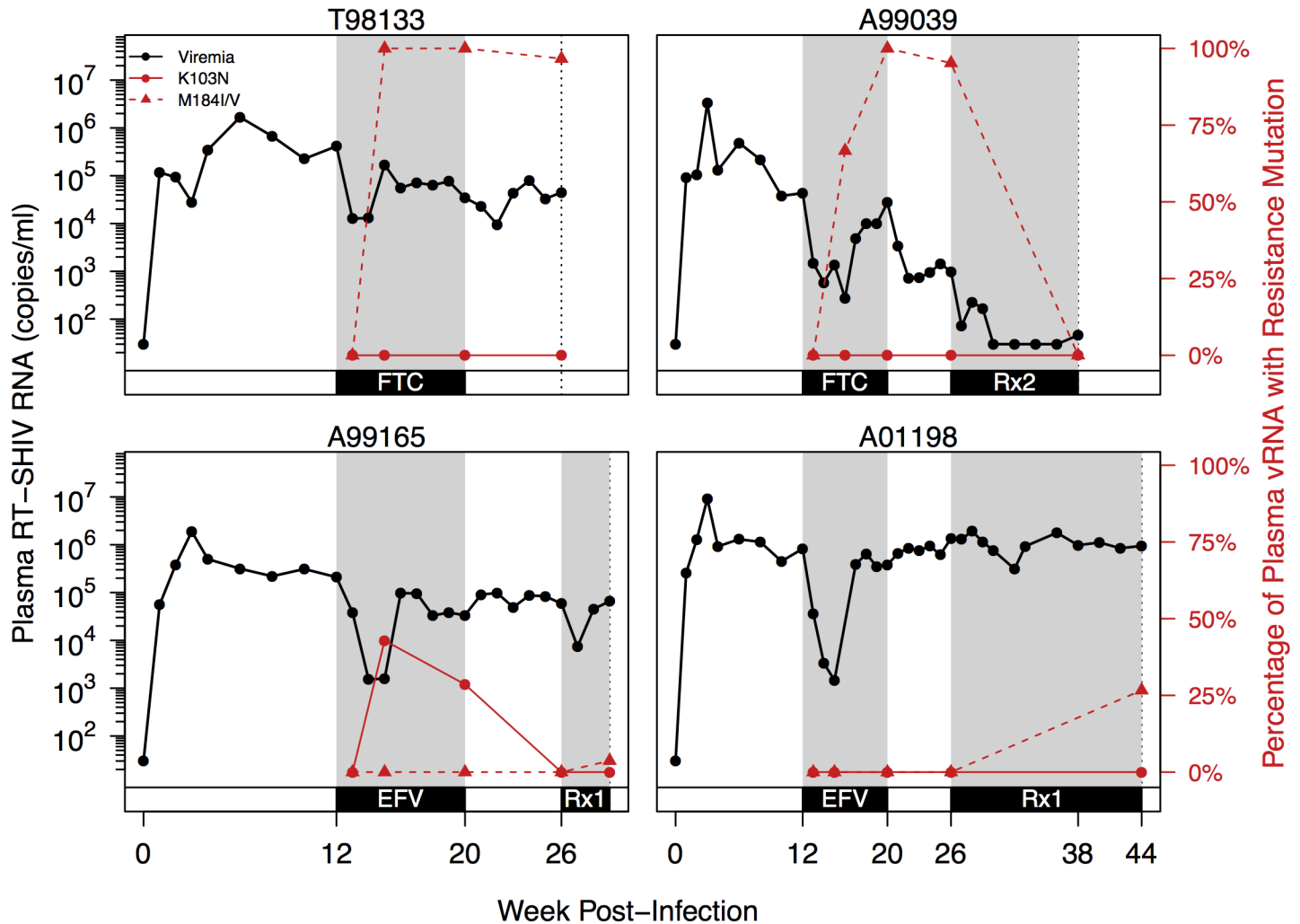
Fig 1. Plasma viremia and drug resistance in four macaques over the course of infection. Plasma RT-SHIV (black) and frequency of plasma vRNA with DRMs (red) are plotted over time for each animal, as determined by SGS. Red lines show the frequency of K103N (solid) or M184I/V (dashed) in the plasma vRNA. Gray shading denotes administration of drug(s) between weeks 12–20 and weeks 26–44. Black dotted line indicates the final time point for each animal. Rx1 is treatment FTC+TFV+EFV and Rx2 is the treatment TFV+L870812+DRV/r.

(approximately 84 generations). Two macaques were treated with the nucleoside analog (NRTI) emtricitabine (FTC; macaques T98133 and A99039) and two received the non-nucleoside RT inhibitor (NNRTI) efavirenz (EFV; macaques A99165 and A01198). Treatment was administered for 8 weeks, followed by a 6 week treatment interruption, and then combination therapy: 1) FTC, tenofovir (TFV), and EFV (macaques A99165 and A01198; which includes a recycled drug to which virus may be resistant) or tenofovir, L-870812, and DRV/r (macaques T98133 and A99039; to ensure complete virus suppression by all 3 drugs). RT-SHIV populations were sampled three times during monotherapy for each macaque (weeks 13, 15/16 and 20), once after treatment interruption (week 26), and once after 12–18 weeks of combination therapy (between weeks 38–44). See the black boxes on the bottom of Fig 1 for a summary of sampling and treatments.

Samples were taken at the different time points from the blood, LN, gut, and vagina. The first 770 nucleotides of reverse transcriptase (RT) were sequenced from vRNA and vDNA using SGS (S1 Table). We did not find that diversity changed significantly over time except in

the vRNA of A01198 and the vDNA of A99165 (S1 Fig, S2 Table), and values of average pair-wise diversity (π) ranged between 0.2 and 0.6%, similar to what is seen in HIV-1 at a comparable amount of time after infection [40].

Drug resistance emergence and persistence during and after monotherapy

Overview of drug resistance emergence in plasma viral RNA over time. In all animals, after challenge with RT-SHIV, but prior to the start of monotherapy, plasma viremia increased and a set point was reached at approximately 4 weeks post-infection (10^5–10^6 viral copies/mL plasma, **Fig 1**). After initiation of monotherapy, plasma viremia transiently decreased in all animals (ranging from 33- to 566-fold) before rebounding to near pre-therapy levels. In both FTC-treated macaques (T98133 and A99039), the expected DRMs M184V/I emerged between weeks 13–15 (1–3 weeks post-treatment), concomitantly with viral rebound and FTC concentrations in the plasma and tissues (S2A **Fig**). The rate at which viruses carrying DRMs displaced wild-type (WT) viruses suggests a selection coefficient for M184V/I of 0.74 in T98133 and 0.41 in A99039 (i.e., relative fitness benefits of 74% and 41%, respectively). In the EFV treated macaques however, the expected DRM (K103N) emerged but did not reach fixation in one macaque (A99165) and did not arise at all in the other macaque (A01198). In the animal in which K103N arose (A99165), plasma viremia initially declined but rebounded around week 15, coinciding with the spread of resistance. We suspect that the expected resistance mutation did not fix in A99165 and A01198 because selection pressure was weak due to the rapid metabolism of NNRTIs in macaques (S2B **Fig**), which we have observed previously [41, 42]. Because EFV was not present at consistently high concentrations, we did not estimate a selection coefficient *s* for K103N.

After monotherapy treatment was interrupted, drug-resistant plasma vRNA persisted between weeks 20–26 in both FTC-treated macaques (T98133 and A99039) (**Fig 1**), which suggests that the cost of drug resistance in the absence of the drug is insufficient for it to notably decrease in frequency over 6 weeks. In the EFV-treated macaque A99165, drug resistance was not observed in the plasma at the end of the treatment interruption, but in this case, the decline in frequency had already started at week 20, likely due to a decrease in EFV concentrations in plasma and tissues. Drug resistance remained absent in A01198.

Combination therapy suppressed RT-SHIV in A99039, and no further drug resistance emerged. In this macaque M184V/I were undetectable in plasma samples taken during virologic suppression (week 38), which suggests that the cost of FTC resistance in the absence of drug is significant on this timescale or that the viruses that were sampled when viremia was suppressed stem from a reservoir that had always remained WT. A99165 and A01198 both began combination therapy that included FTC, but viral replication was not suppressed in either animal. Drug resistance to FTC emerged (in the form of M184I and M184V, respectively), but did not fix in either macaque, reaching frequencies of 4% and 27% in the plasma, respectively.

Drug resistance emerges simultaneously in viral RNA from multiple tissues. In a clinical setting, drug resistance is assessed using samples taken from the blood plasma, but how representative these samples are to levels of drug resistance in vRNA in PBMC and tissues is unclear. Here, we compared the kinetics and composition of drug-resistant variants across multiple anatomic compartments (blood, lymph node, gut, and vagina). Across the three monkeys in which drug resistance to monotherapy was observed, DRMs occurred at similar time points in the plasma and other compartments (**Fig 2**). In T98133, all compartments had drug-resistant vRNA detected at approximately the same time point (week 15/16). A99039 had
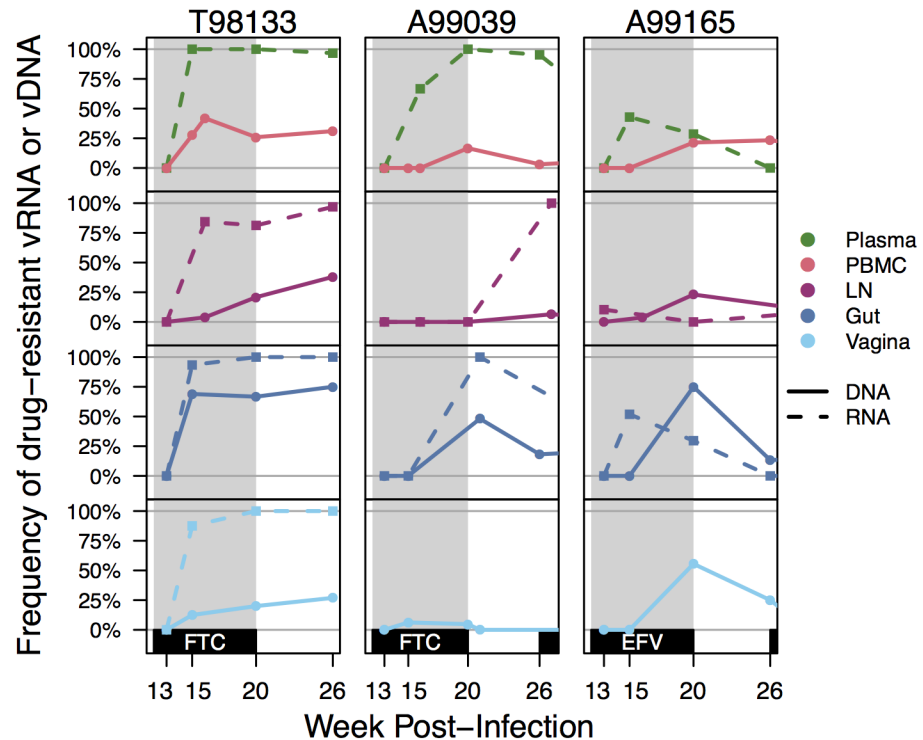
Fig 2. Drug resistance increase over time across all compartments. Drug resistance was defined as the proportion of vRNA (dashed lines) or vDNA (solid lines) sequences having K103N and/or M184V/I within each compartment for animals T98133, A99039, and A99165. Gray shading indicates the FTC or EFV treatment periods for the macaques.

https://doi.org/10.1371/journal.ppat.1006358.g002

drug-resistant vRNA first in the plasma at week 15, and then only later in the gut (week 20) and the lymph node (week 27). Although we have no vRNA samples from the vagina in this animal, the vDNA from the vagina contained drug resistance at week 15, suggesting that M184I/V emerged early in that compartment. In A99165, drug resistance was observed first at low frequency in the lymph node at week 13 and then simultaneously in gut and plasma vRNA two weeks later.

Although drug resistance appeared at similar times across compartments, there was some evidence the drug resistance was maintained at different frequencies in different compartments. Among all paired vRNA samples from the same macaque at the same time point (±1 week) where DRMs were neither at 0% or 100% frequency in both compartments, drug resistance was at a lower frequency in the lymph node vRNA than the gut vRNA (5/6 comparisons), PBMC vRNA (5/6 comparisons), the plasma vRNA (5/8 comparisons) and vagina vRNA (4/5 comparisons). None of these comparisons were significant, but they suggest that drug-resistant viruses may not have replicated as rapidly in the lymph node as in other sampled compartments. This could be due to weaker selective pressure consistent with previous reports of poor drug penetrance to the lymph nodes [43, 44]. We also observed that a greater proportion of vDNA from the gut contained DRMs than vDNA from other compartments (**Fig 2**). Among paired vDNA samples from the same macaque at the same time point (±1 week) where DRMs were neither at 0% or 100% frequency in both compartments, drug resistance was at a higher frequency in the gut vDNA than the LN vDNA (6/6 comparisons), PBMC vDNA (7/7 comparisons), and vagina vDNA (7/7 comparisons).

Not only did drug resistance emerge at similar times across different compartments, but identical viral genotypes were observed at similar times in different compartments. Figs 3, 4 and 5 show the frequencies of different viral genotypes throughout time in different compartments in each of the three monkeys that had drug resistance (T98133, A99165, and A99039, respectively). Each color represents a different genotype and the WT is depicted in grey. Genotypes with DRMs are hatched. This visualization clearly shows that genotypes are usually not restricted to single compartments, but they are observed in several compartments.

For each monkey, the most frequent drug-resistant genotypes are often observed across vRNA from all compartments. For example, in T98133, at week 15/16 after initial infection, vRNA sampled from the plasma, lymph node, gut, and vagina all had two different genotypes with DRM M184V (one genotype with only M184V in orange and one genotype with an additional synonymous substitution at the 255th amino acid of RT, N255N, in light blue) (**Fig 3A**). Three of the four compartments also contained vRNA with DRM M184I. The similarity of vRNA sequences found in each compartment, particularly vRNA with synonymous mutations (i.e., M184V linked to N255N), suggests that migration was responsible for spreading viral genotypes from one compartment to another, as opposed to independent origins of each genotype in each compartment. The frequencies of genotypes remain similar between the compartments over time (**Fig 3A**), suggesting that substantial migration was ongoing. Similarly, in A99165 and A99039 (**Figs 4A and 5A**, respectively), the same drug-resistant genotypes (including linked mutations) are observed at similar time points in the vRNA in different tissue compartments.

DRMs develop on multiple viral genomes. If mutation generates DRMs fast enough (mutation rate multiplied by population size larger than 1), we would expect multiple independent origins of DRMs on multiple backgrounds that spread simultaneously (i.e., both present at intermediate frequencies before either reaches 100% frequency) in what is known as a soft sweep. Alternatively, if DRMs arise only rarely, a single DRM may increase to 100% frequency in the population due to selection before a second DRM arises on a different background in the population. This is called a hard sweep [45].

There are two methods we can use to determine how many times drug resistance developed on an individual viral genome in the macaques. First, we can determine whether different DRMs (distinguishable at either the amino acid or nucleotide level) are observed simultaneously in the same population. Second, by SGS we can determine whether the same DRM occurred on different genetic backgrounds. If drug resistance increases in frequency due to a hard sweep, we should see only one DRM on one genetic background spreading (although the mutational background can later diversify due to development of new mutations). If drug resistance increases in frequency due to a soft sweep, we may see several different DRMs or we may see only one DRM that is present on different genetic backgrounds.

We first observe that drug resistance occurs multiple times through mutations distinguishable from their encodings (i.e., the DRMs themselves have different sequences). In FTC-treated macaques, RT-SHIV developed both M184V and M184I, similar to what has been observed in HIV-infected individuals [46, 47]. In both cases, M184V was more common than M184I: 91.5% and 97.0% of drug-resistant vRNA variants had M184V in T98133 and A99039, respectively (**Figs 3 and 5**), which is similar to what has been shown in HIV-infected individuals due to a fitness advantage of M184V over M184I [47]. The macaque treated with EFV (A99165) only acquired one DRM (K103N). However, K103N can be caused by mutations either from A to U or from A to C at the third residue of amino acid 103 and these can both be present in HIV-infected individuals [48]. Both of these mutations were observed within A99165 (**Fig 4**). In both of the above cases, the encoding of two different amino acid mutations (M184V/I) or the two different nucleotide-level encodings of the same amino acid (K103N)
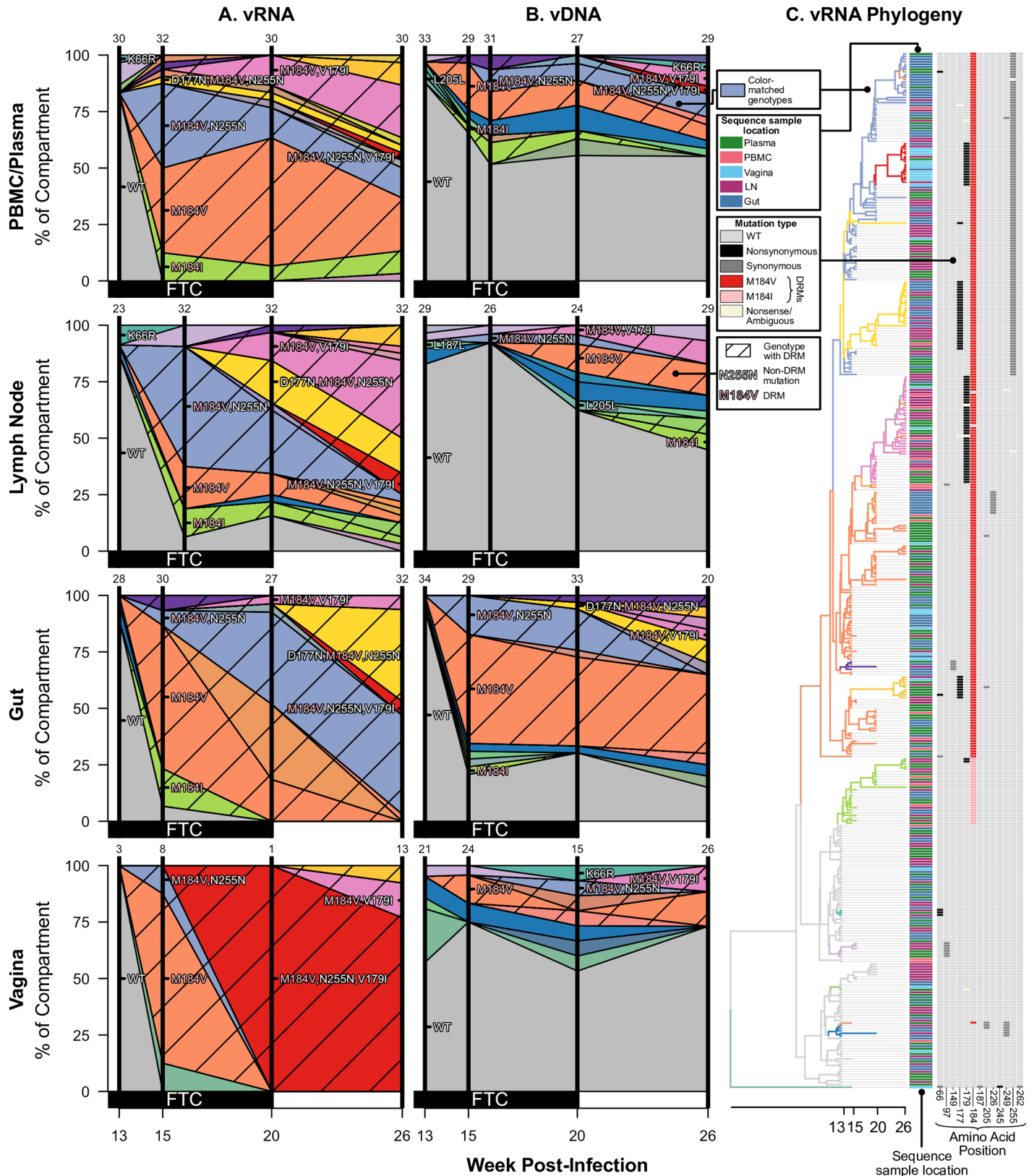
Fig 3. Drug resistance spreads dynamically across several compartments in macaque T98133 over the course of treatment. Colored polygons show the prevalence of different genotypes of RT-SHIV RNA (A) or DNA (B) over time in different compartments. Vertical distance at black vertical lines

indicates the frequency of the sample with an exact set of mutations (i.e., genotype) at the time of sampling. Grey indicates WT virus, and striations on the background of a color indicate a drug-resistant genotype. Common mutations are labeled, and known DRMs are labeled in pink. For plasma/PBMC, plasma vRNA and PBMC vDNA are shown. Two sequences are considered the same genotype when all mutations found above frequency 1% in the macaque are shared. Sample sizes for at each sampling point are given above each vertical line. (C) Time-structured phylogenies reveal relationships between vRNA sequences (shown as leaf nodes). Branch lengths indicate sequence sampling time, and branches are colored to match (A) and (B). Sampling location is indicated for each sequence, and the identity of all mutations at frequency >1% is shown to the right. Colors indicating mutation type (synonymous, nonsynonymous, DRM, stop or nonsense/missense) are shown in the legend.

demonstrates that at least two independent mutations must have occurred within each of the three macaques with drug resistance. This shows that soft sweeps are occurring and shows that the mutation rate multiplied by the effective population size is larger than 1 in all three monkeys.

However, SGS allowed further resolution to determine the number of times each mutation arose. In T98133, DRM M184V was seen at high frequency both with and without a synonymous N255N mutation in the sampled vRNA (**Fig 3**). This suggests a second origin of drug resistance in which the M184V mutation occurred on a background that already had the synonymous N255N mutation. In principle, the N255N mutation may have arisen shortly after drug resistance began to spread, but given that the M184V+N255N lineage reaches a high frequency, it is more likely that M184V occurred on a genotype that already had the N255N mutation. Combined with the observation of M184I, at least three origins of resistance were likely in T98133. One genotype containing a DRM is observed in the vDNA but was not detected in the sampled vRNA (S3A **Fig**).

In A99039, the DRM M184V occurred on three distinct backgrounds in sampled vRNA (**Fig 5**): 1) with L187L+M245T, 2) with L205L+R277K, and 3) with K223K+L109L+V75L. As M184I was also observed at low frequency, at least four origins of resistance were likely in A99039. As in T98133, one additional background is also observed in the vDNA but not in the vRNA (S3C **Fig**).

In A99165, the DRM K103N emerged on three distinguishable backgrounds in sampled vRNA (**Fig 4**): 1) with no other high frequency mutations, 2) with a synonymous L187L mutation, and 3) with four non-DRMs (K249K, L187L, L205L, Q174R). All three of those backgrounds were observed before detection of K103N. Further, both the nucleotide level encodings of K103N (A→U and A→C) were observed on the L187L background, suggesting at least four origins of resistance were likely in A99165. The vDNA displayed both nucleotide level encodings of K103N on three backgrounds each (S3B **Fig**).

Even A01198, the macaque in which drug resistance to monotherapy did not arise, had DRMs appearing on multiple backgrounds at the final time point when examining the vRNA and vDNA phylogenies, with six unique backgrounds for DRMs visible in the vDNA (S4 **Fig**).

We used the occurrence of multiple DRMs on different backgrounds to estimate the effective population size (following [49]) at the point of initial treatment. The idea behind this approach is that the number of times a DRM occurs in a population depends on the size of the population. In a large population, we expect DRMs to occur more often than in smaller populations. We differentiated independent origins by examining drug-resistant viral sequences with different codons at a particular DRM (e.g., K103N encoded by AAU or AAC) and different identities by genotype (e.g., a DRM linked to another high frequency mutation), and calling the resulting estimates $N_{e,DRM}$ and $N_{e,geno}$, respectively. We performed both these estimates for just vRNA and combined vRNA and vDNA (see Materials and Methods). We find estimates of N_e among the vRNA between $1.17 * 10^4$ and $5.71 * 10^5$ depending on the macaque and which technique was used (**Table 2**). vRNA and vDNA combined estimates are similar. These estimates were between two and three orders of magnitude higher than
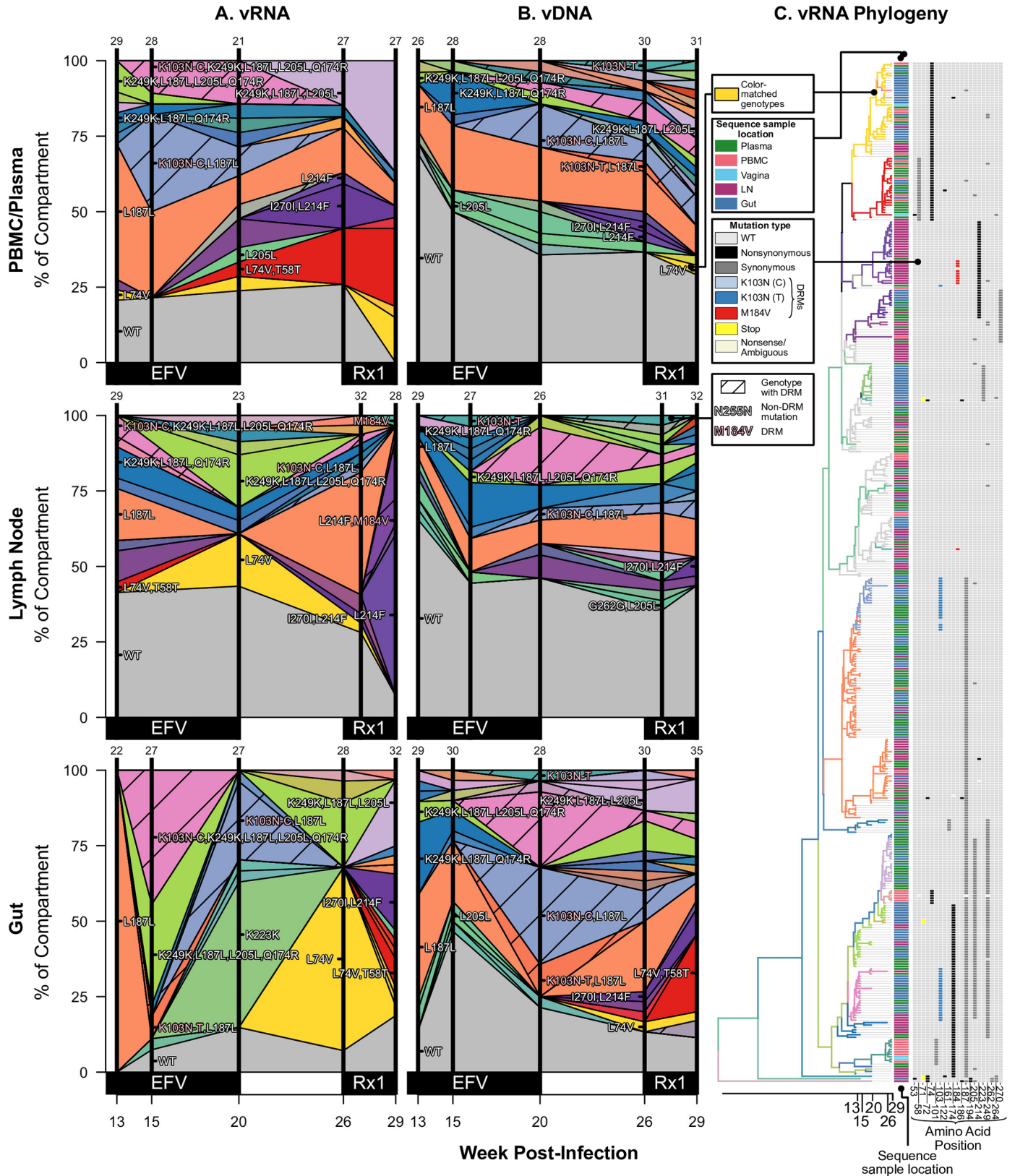
Fig 4. Drug resistance spreads dynamically across several compartments in macaque A99165 over the course of treatment. Colored polygons show the prevalence of different genotypes of RT-SHIV RNA (A) or DNA (B) over time in different compartments. Vertical distance at black vertical lines

indicates the frequency of the sample with an exact set of mutations (i.e., genotype) at the time of sampling. Grey indicates WT virus, and striations on the background of a color indicate a drug-resistant genotype. Common mutations are labeled, and known DRMs are labeled in pink. For plasma/PBMC, plasma vRNA and PBMC vDNA are shown. Two sequences are considered the same genotype when all mutations found above frequency 1% in the macaque are shared. Sample sizes for at each sampling point are given above each vertical line. (C) Time-structured phylogenies reveal relationships between vRNA sequences (shown as leaf nodes). Branch lengths indicate sequence sampling time, and branches are colored to match (A) and (B). Sampling location is indicated for each sequence, and the identity of all mutations at frequency >1% is shown to the right. Colors indicating mutation type (synonymous, nonsynonymous, DRM, stop or nonsense/missense) are shown in the legend. Note, except for WT, the coloration is not preserved between figures (i.e., the yellow in Fig 3 does not represent the same genotype as the yellow in Fig 4). Rx1 is treatment FTC+TFV+EFV.

estimates using standard diversity based methods ($N_{e,d}$, see Materials and Methods). This finding is addressed in more length in the discussion.

Two other mutations are notable in relation to the spread of drug resistance. T98133 had the polymorphic mutation V179I emerging on multiple backgrounds (**Fig 3**). This mutation can cause resistance to some NNRTIs, which T98133 did not receive. V179I appeared on most drug resistant backgrounds (M184V, M184V+N255N, and M184I), possibly indicating a selective benefit. Also notable was that M184V+N255N+V179I appeared first in the vaginal vRNA and appeared only later across other tissues. This is one of the few instances in this data where a simple migration direction is suggested. Similarly, A99165 had the mutation L74V rapidly reach high frequency across tissues (**Fig 4**), despite L74V conferring resistance to select NRTIs, which A99165 was not given until later time points. However, L74V has been shown to improve fitness of HIV-1 by increasing levels of RT in virions [50]. Both of these mutations support that the selective landscape within RT is composed of more than known DRMs.

Frequencies of common viral variants in viral RNA suggest compartments are not completely well-mixed

We have shown that the same genotypes containing DRMs appeared simultaneously across multiple compartments, suggesting a substantial amount of migration. However, these same genotypes are found at different frequencies in different compartments (e.g., the frequency of M184V+N255N across compartments in **Fig 3**). We asked whether the observed differences are due to real differences between the compartments or whether they are due to random sampling effects and fairly small sample sizes. In this context, we define two compartments to be **compartmentalized** if their vRNA genotype frequencies are significantly different when using a randomization test, and **well-mixed** otherwise.

To assess the extent of compartmentalization, we compared the frequency of vRNA genotypes in different compartments using a variance partitioning statistic, K_{ST} [51] and two other tests: AMOVA [52] and the Slatkin-Maddison test [53]. Between each pair of compartments sampled at the same time point (±1 week) in each macaque, we repeatedly subsampled 10 sequences from each compartment without replacement and computed the test statistic. The significance of each test statistic was assessed via permutation test with larger values of K_{ST} and AMOVA's φ and lower values of Slatkin-Maddison inferred migration events indicating evidence of more compartmentalization. The proportion of tests significant at the 5% level is reported. Therefore, the reported value is the probability of a significant test when drawing 10 sequences randomly. Note that we subsample compartments to the same number of sequences to equalize power across comparisons. The subsample size of 10 was chosen as a tradeoff between power to detect compartmentalization and the number of comparisons that can be computed given that some of the sample sizes are very small (see Materials and Methods).

Across all pairwise tests where each compartment had at least 10 sequences, the probability of rejecting the compartments as well-mixed using a K_{ST} permutation test was 0.47 (**Fig 6**).
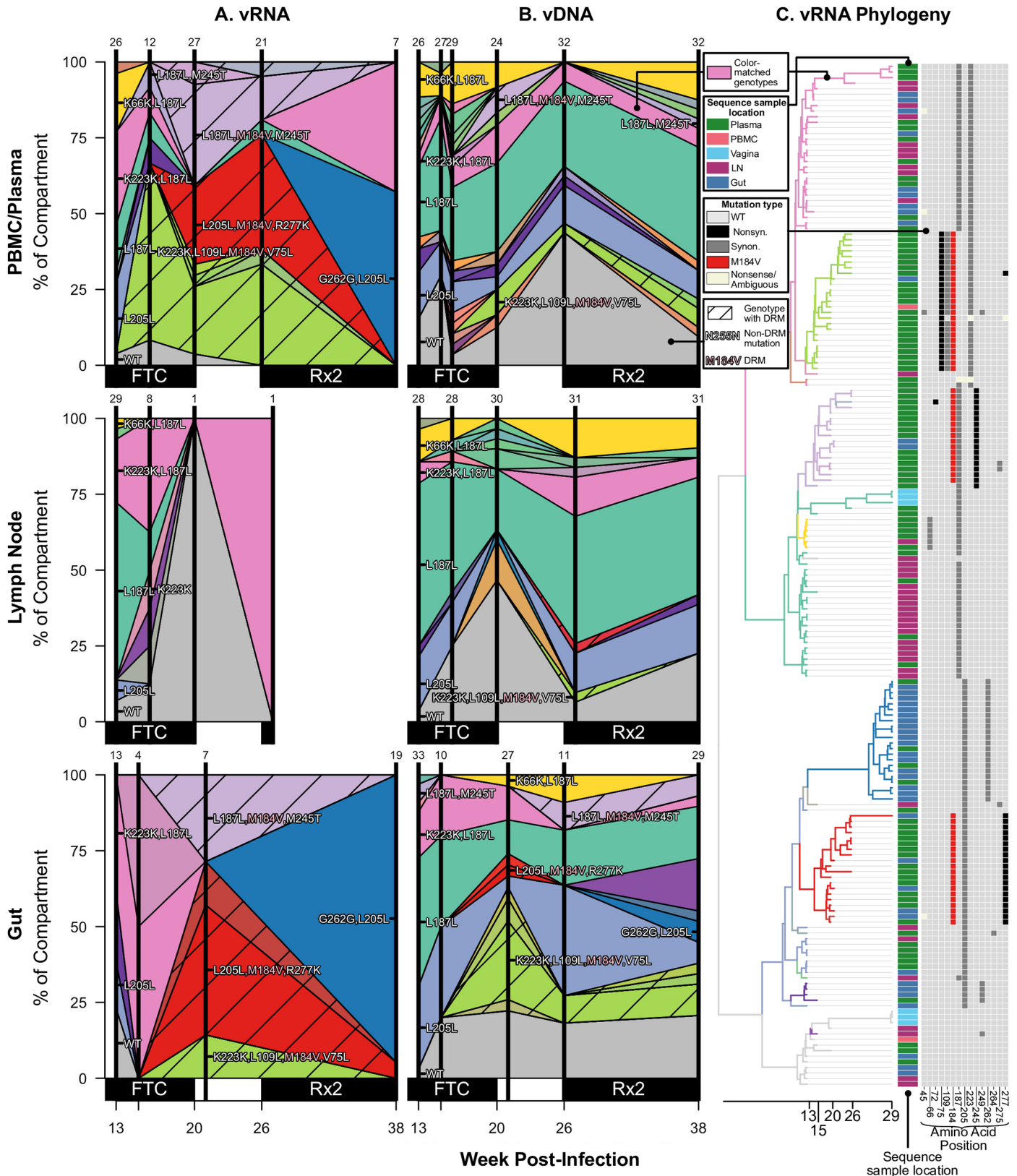
Fig 5. Drug resistance spreads dynamically across several compartments in macaque A99039 over the course of treatment. Colored polygons show the prevalence of different genotypes of RT-SHIV RNA (A) or DNA (B) over time in different compartments. Vertical distance at black vertical lines

indicates the frequency of the sample with an exact set of mutations (i.e., genotype) at the time of sampling. Grey indicates WT virus, and striations on the background of a color indicate a drug-resistant genotype. Common mutations are labeled, and known DRMs are labeled in pink. For plasma/PBMC, plasma vRNA and PBMC vDNA are shown. Two sequences are considered the same genotype when all mutations found above frequency 1% in the macaque are shared. Sample sizes for at each sampling point are given above each vertical line. (C) Time-structured phylogenies reveal relationships between vRNA sequences (shown as leaf nodes). Branch lengths indicate sequence sampling time, and branches are colored to match (A) and (B). Sampling location is indicated for each sequence, and the identity of all mutations at frequency >1% is shown to the right. Colors indicating mutation type (synonymous, nonsynonymous, DRM, stop or nonsense/missense) are shown in the legend. Note, except for WT, the coloration is not preserved between figures. Rx2 is the treatment TFV+L870812+DRV/r.

This result shows that although DRMs appear to spread quickly from one compartment to another, migration is not pervasive enough to make the populations entirely well-mixed.

We find that compartmentalization is dynamic through time. While some pairs of compartments in some macaques have a consistently high probability of observing compartmentalization (e.g., A99165 gut and lymph node vRNA) or a consistently low one (e.g., A01198, plasma and PBMC vRNA), most often, compartmentalization is transient through time (**Fig 6**). The Slatkin-Maddison test and AMOVA produced similar results (S5 **Fig** **and** **S6 Fig**, respectively).

We also characterized which vRNA samples appeared most similar to each other throughout time (**Fig 7**)**.** The blood (plasma and PBMC) and lymphoid systems appear well connected with few rejections of being well-mixed in pairwise comparisons. The most compartmentalized comparisons were between the lymph nodes and mucosal tissues (gut and vagina) or between mucosal tissues themselves. Similar results hold using the Slatkin-Maddison test and AMOVA (**Fig 7**). Biologically, these observations are in line with expectations about system connectivity.

The differences and similarities between vRNA and vDNA

We have characterized the dynamics of the vRNA in different compartments, but this may or may not be representative of the viral reservoirs. We therefore also sampled vDNA in all compartments, which allows us to compare vDNA and vRNA from the same location. There were marked differences in genetic composition between vRNA and vDNA. Most notably, the vDNA retained a higher percentage of WT viruses than the vRNA in T98133 and A99165 (**Figs 3**, **4 and 5**). In A99039, very little WT vRNA or vDNA was present, even at the initial time point, and the population was dominated by viruses with synonymous variants (L187L, K223K, L205L and combinations thereof). However, there were many fewer DRMs in the vDNA samples than in the vRNA samples, suggesting that, like in the other monkeys, the vDNA is changing more slowly than the vRNA (**Fig 5**).

To more directly test whether the rate of change was higher in vRNA than vDNA, we computed K_{ST} between consecutive time points separately from both the vRNA and vDNA from

Table 2. Estimates of N_e taken after the spread of RT-SHIV drug resistance to monotherapy using three different estimation methods (see Materials and Methods).

	$N_{e,d}$ (Plasma)	$N_{e,DRM}$ (DR vRNA)	$N_{e,geno}$ (DR vRNA)	$N_{e,DRM}$ (DR all)	$N_{e,geno}$ (DR all)
T98133	$8.51*10^1$	$1.17*10^4$	$5.31*10^4$	$1.08*10^4$	$4.72*10^4$
A99165	$7.46*10^2$	$4.14*10^4$	$2.38*10^5$	$6.15*10^4$	$3.29*10^5$
A99039	$1.46*10^3$	$1.47*10^4$	$5.71*10^5$	$9.61*10^3$	$5.62*10^5$
A01198*	ND	ND	ND	ND	ND

ND: Not done

*Because no drug resistance arose in A01198 during monotherapy, the number of origins of resistance could not be used to estimate effective population size.
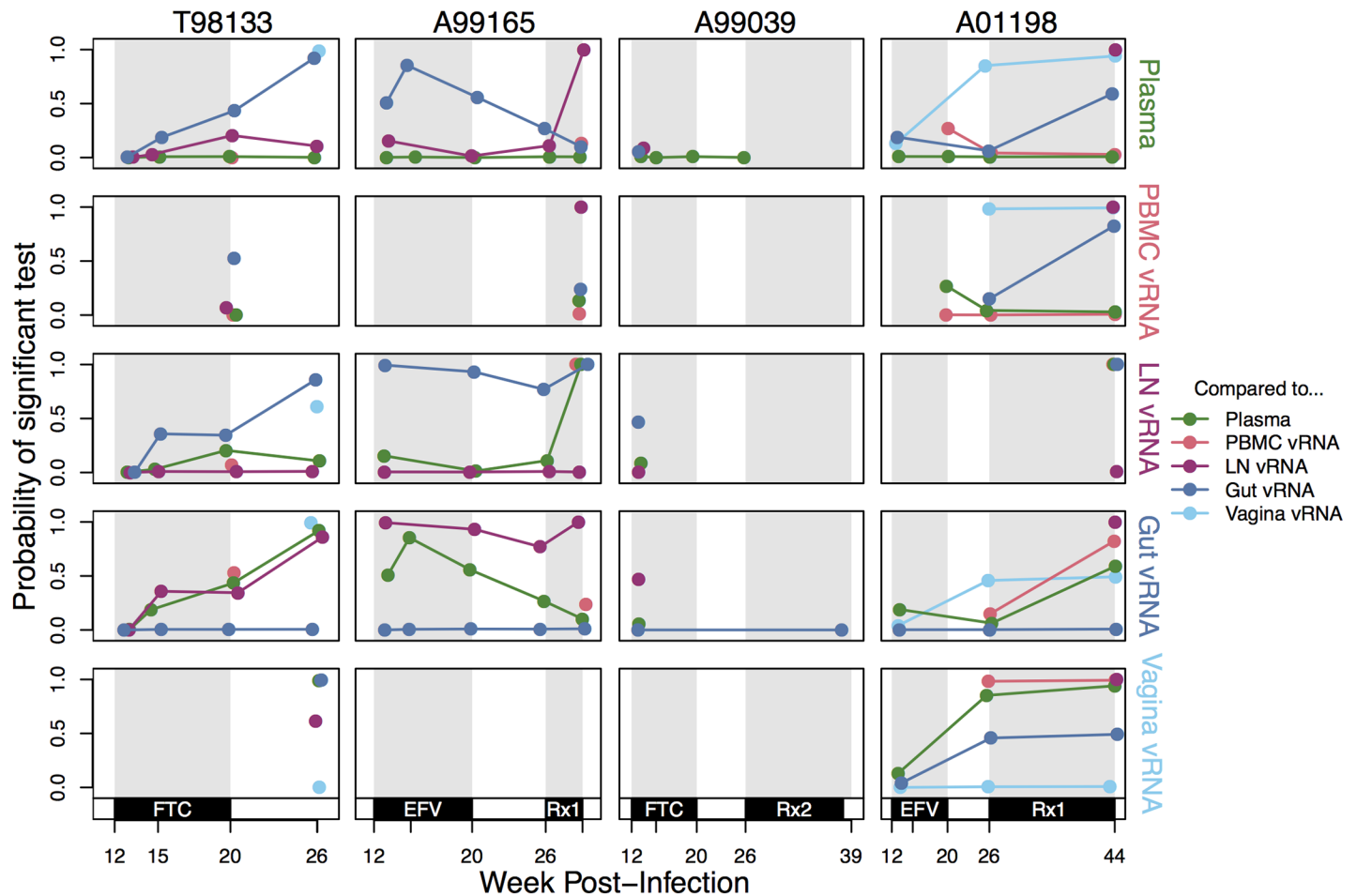
Fig 6. Compartmentalization relationships between different samples reveal transient and stable patterns. Probability of a significant pairwise compartmentalization test between vRNA subsampled from each compartment for the four macaques (columns) over time (x-axis). Each row represents comparisons between the vRNA in each compartment marked at the right (from top to bottom: plasma, PBMC, LN, gut and vagina vRNA) to all other compartments. Because all pairwise relationships are shown for each compartment, lines are repeated (i.e., the lymph node versus plasma line is present in both the lymph node and plasma rows). The y-axis indicates the proportion of K_{ST} tests significant at the 5% significance level when subsampled to 10 sequences per compartment 1000 times. Coloration indicates that the comparison was done between the focal compartment for the row and plasma (green), PBMC (pink), LN (purple), gut (dark blue) or vagina (light blue). Rx1 is treatment FTC+TFV+EFV and Rx2 is the treatment TFV+L870812+DRV/r.

the same compartment of each macaque (Fig 8). All pairs of adjacent time points were treated equivalently, even though they were separated by different amounts of time. However, because the average difference in weeks between time points was not significantly different between vRNA and vDNA (7.39 weeks versus 6.53 weeks respectively, t-test p-value = 0.50), we could still compare the number of significantly compartmentalized comparisons as we did in Fig 7. As in the pairwise compartment analysis, we subsampled comparisons to 10 sequences per time point 1000 times and report the proportion of those tests resulting in a permutation test p-value ≤ 0.05. We find that the probability of a significant test between consecutive vRNA time points from the same compartment (indicating a different viral composition) was 0.55 as compared to 0.25 between consecutive vDNA comparisons. A t-test between the relative probabilities provides evidence that vDNA changes more slowly than vRNA ($p = 6 * 10^{-4}$). This approach also revealed that vRNA from successive time points in the gut was more often compartmentalized than those from the plasma (0.85 probability of significant test compared to 0.37, $p = 2.92 * 10^{-3}$).
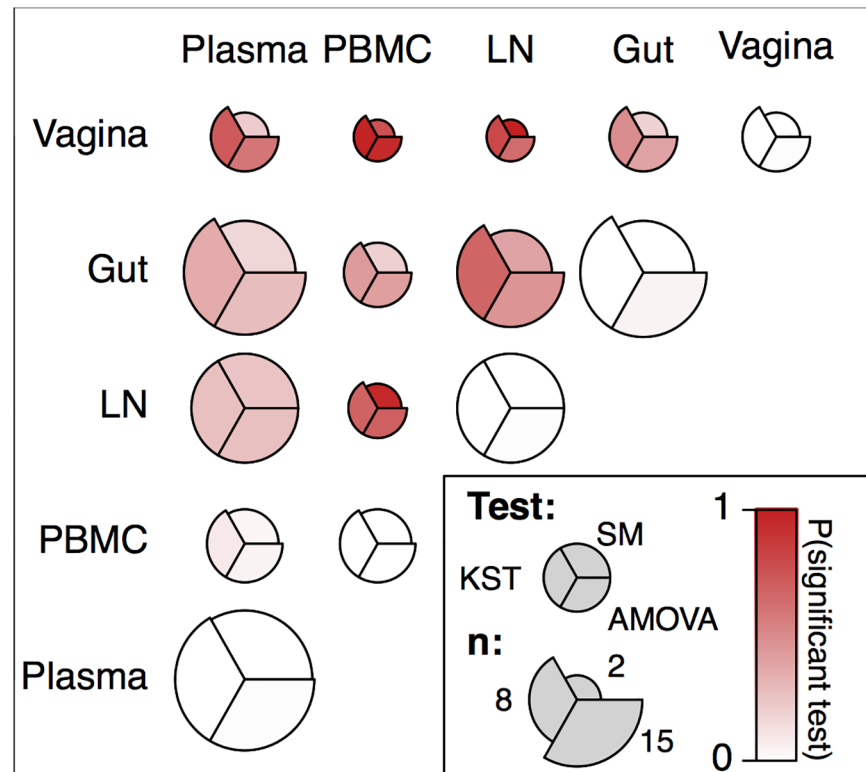
Fig 7. Compartmentalization test results reveal pairwise RT-SHIV RNA compartmentalization relationships between blood, lymph node, gut, and vagina. Each circle summarizes the results of three compartmentalization tests performed pairwise between the vRNA from compartments within a macaque at a given time, and the position of the segment represents the test performed (left: KST, top-right: Slatkin-Maddison (SM), bottom-right: AMOVA). The coloration indicates the probability of a significant test result at the 5% significance level when subsampling 10 sequences from each compartment averaged across each macaque and all time points with red indicating a high probability and white indicating a low probability of a significant test. The area of the segment indicates the number of comparisons contributing to that average (i.e., the number of time points with a sufficient number of sequences in each compartment within a macaque). Because we used 10 unique subsampled sequences to perform the SM test, the SM test could not be performed on some macaque time points, resulting in smaller sample sizes (see Materials and Methods for details).

We then characterized the relationship between the sampled vRNA and vDNA variants by examining viral variants found in the vRNA but not the vDNA and vice versa. We compared all time points for which there were vDNA and vRNA samples ($n \geq 3$ for both vRNA and vDNA) from a single compartment in a macaque. The proportion of vRNA observed without corresponding vDNA was 25% on average across macaques (S7 Fig), suggesting a large portion of the vRNA was transcribed from vDNA at a frequency below the limit of sampling ($< \approx 1/30$). A similar proportion of vDNA is sampled without any concurrent vRNA. This may mean that the provirus is latent (i.e., not currently producing vRNA) or replication incompetent.

In three of the four macaques (T98133, A99165, A99039) vDNA became less representative of the vRNA over time (i.e., a larger proportion of vDNA was sampled without vRNA) (S7 Fig), although this effect appeared to plateau in A99165 and A99039. This is consistent with the vDNA building up a "fossil record" or "seed bank" of proviruses that are not transcribed at a given time but that may later be reactivated.

We also examined the extent of G to A hypermutation and premature stop codons found in vRNA versus vDNA sequences. We found that vDNA sequences were approximately five
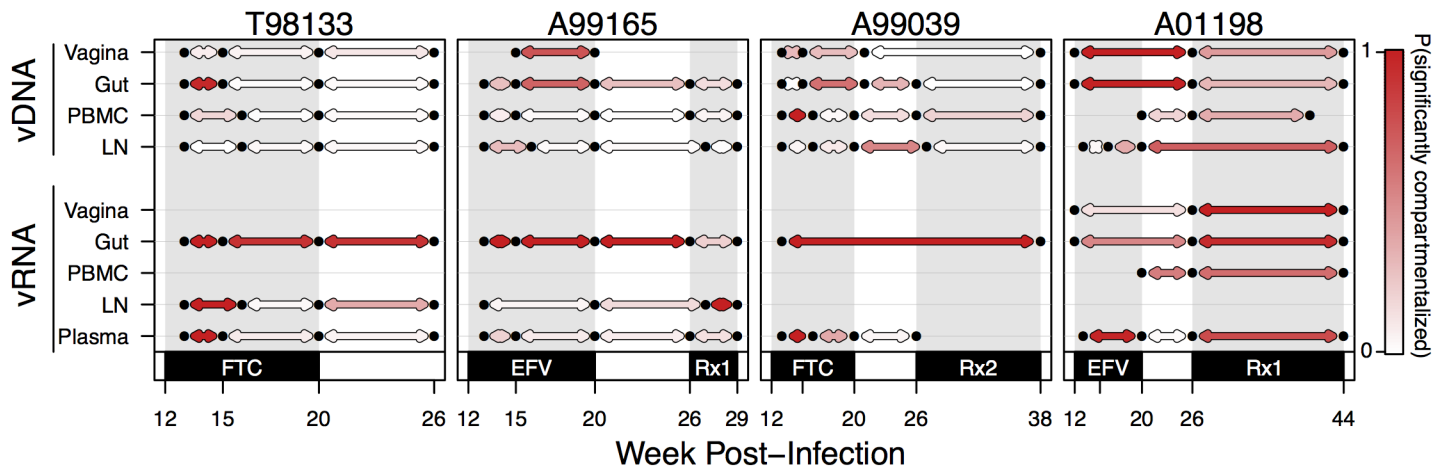
Fig 8. Change in compartmental viral composition over time varies by compartment and drug pressure. For each macaque, K_{ST} tests are plotted between all time points with adjacent samples (i.e., week 12 or 13 compared to week 15 or 16) for each compartment (vRNA top, vDNA bottom). Black dots indicate time points with samples of at least 3 sequences. Red lines indicate that a K_{ST} test comparing those two samples is significantly different (with a 5% FDR). Grey lines indicate a failure to reject the null hypothesis of being well-mixed at the 5% FDR level. Grey shading indicates monotherapy or combination therapy, as indicated below the x-axis. Rx1 is treatment FTC+TFV+EFV and Rx2 is the treatment TFV+L870812+DRV/r.

times more likely to have >2% of available G residues mutated to A when compared to vRNA sequences (Fisher's Exact Test odds-ratio: 4.79, p(true OR = 1) < $9.066*10^{-11}$). VDNA sequences were approximately three times more likely to have a premature stop codon as vRNA sequences (Fisher's Exact Test OR: 2.98, p(true OR = 1) < $1.68*10^{-5}$). As expected, this suggests a considerably higher level of selective constraint on the vRNA as compared to the vDNA.

Discussion

In this paper, we present high-resolution sequencing data from a SIV-HIV chimeric virus (RT-SHIV) sampled across space and time in infected macaques. Although our data is sampled from RT-SHIV-infected macaques, this system corresponds well to HIV-1 in humans in many ways. First, when untreated, RT-SHIV viremia in macaques plateau at a similar set point as HIV-1 in many humans, suggesting that census population sizes are similar between the two systems. Second, RT-SHIV reaches levels of genetic diversity that are similar to levels in HIV-1, as quantified by average pairwise diversity [40]. Third, drug resistance arises via the same DRMs that are common in HIV-1 (M184V/I and K103N) and, fourth, like in HIV-1, the mutations spread via soft sweeps after occurring independently on many backgrounds [49, 54]. These observations suggest that an intra-host RT-SHIV population closely mimics that of an intra-patient HIV-1 population undergoing monotherapy, which may occur during drug nonadherence.

Similarities between RT-SHIV in macaques and HIV-1 in humans make RT-SHIV a good model system for HIV-1. In addition, it is possible to sample RT-SHIV much more intensively, including from multiple, longitudinal tissue biopsies, than HIV-1. Thanks to viral RNA and viral DNA single-genome sequences from different time points and different compartments, we have detailed information about evolution in time and space which allows us to make more detailed observations about the dynamics of the system. We have data from 4–5 sampled time points from blood, gut, lymph node, and vagina (both vRNA and vDNA) from four infected macaques. This level of detail allows us to explore several open questions concerning intra-host spatial dynamics and drug resistance.

First, we can ask how well virus from the blood plasma represents the remainder of the intra-host population. This is important because often only the blood is sampled in HIV-infected individuals. We find that while the blood plasma often has the same viral variants as other compartments, it is not necessarily representative of the dynamics occurring in other compartments. For example, the virus population in the gut changes faster than that in the plasma, both among the vRNA and the vDNA. This may be because the CD4+ T cell composition of the gut is different than that of the blood. Specifically, the gut has a higher proportion of effector memory cells CD4+ T cells (T_{EM}) than the blood or lymphoid tissues [55, 56]. Infection with a CCR5-tropic SIV, including the SIV_{mne027} backbone used in this study [57], leads to significant loss of CD4+ T_{EM} cells in mucosal tissues, such as the gut [58, 59]. These CD4 + T_{EM} cells are also shorter-lived than the CD4+ central memory T cells (T_{CM}) found at higher frequencies in the plasma and lymphoid tissues [58, 60]. This may explain the faster rate of change in the expressed virus (vRNA) and in virally infected cells (vDNA) in the gut. Our data also suggest that drug-resistant genotypes may spread more slowly among the lymph node vRNA than the vRNA from other compartments. Previous studies have found that antiretroviral drug concentrations are lower in the lymph node than other tissues [43, 44]. Although our sample sizes are underpowered to test for transient DRM frequency differences and we were unable to measure drug concentrations from the lymph node, these results suggest that unequal drug penetration may be another source of differing dynamics among compartments. This is in contrast to a rhesus macaque study showing higher frequencies of LN vDNA with cytotoxic T cell epitope escape mutants compared to PBMC vDNA, as LNs are main sites of both virus replication and cellular immune pressure [12]. Consistent with our results, their data show mutation frequency is correlated with selective pressure. More broadly, these findings suggest that while surveying the plasma may accurately reveal the types of mutations present in the HIV-1 quasispecies, it does not necessarily reflect the frequency or dynamics of these mutations in other compartments.

Second, we sampled RNA and DNA, which allows us to observe not just the part of the viral population that is actively transcribing but also the part of the viral population that is not expressed and may be latent (DNA not represented in RNA). This is of interest because it is not well known how the viral reservoir changes in patients undergoing monotherapy. Previous studies of HIV-1 or RT-SHIV populations during suppressive combination antiretroviral therapy (ART) have shown that the reservoir changes very little, suggestive of little or no ongoing replication during ART [18, 61–63]. While studies of actively replicating RT-SHIV populations have suggested that the reservoir is changing through examining changes in vRNA [18, 64, 65], this study directly measures changes in vDNA over time. We find that while new vDNA is produced, the viral composition of the reservoir still remains relatively stable compared to the vRNA. While the majority of vRNA from different compartments of the macaques can contain DRMs, only a small proportion of the vDNA acquires drug resistance, and the majority remains WT or genotypes that were present before treatment. This observation suggests that only a small proportion of the vDNA is actively producing virus, and the majority is defective or latent, even during active replication.

Third, our detailed sampling over space and time helps explain results from existing studies on compartmentalization in HIV-1. As summarized in Table 1, most studies that characterize within-patient population structure within the tissues sampled in this study find that evidence for compartmentalization is found for some comparisons, and not for others. In addition, those studies that sample at different points in time often observe mixed results over time [5–7, 12]. Similarly, our data also show that compartmentalization is a transient occurrence in the blood and tissues we sampled, that arises as the population adapts and declines through frequent and fast migration. This transiency may contribute to the equivocal findings across

HIV-1 compartmentalization studies. In fact, studies that found either ample support for compartmentalization or no evidence for compartmentalization may have found the opposite result had they sampled at a different time.

Fourth, because our study sampled both across time and from multiple compartments, we were able to determine when and where compartmentalization is most likely. We find that particular compartments, specifically the blood and the lymph nodes, usually harbor the same mutant frequencies, likely because of high rates of migration of infected cells between those compartments. We rarely find evidence for compartmentalization between those compartments. On the other hand, some compartments, specifically the blood and the mucosal tissues, harbor significantly different mutant frequencies, likely because of limited migration of cells between those compartments, such as CD4+ T_{EM} and tissue resident T cells. We usually observe evidence for compartmentalization between those compartments.

In addition to answering these open questions, our data are well suited to provide insight into population genetic parameters of viral populations. Knowing these parameters is important in order to use existing theory to make predictions about HIV-1 evolution, which could help improve antiretroviral therapies. For example, because most theory assumes that beneficial mutations are rare, it is important to know if this holds true for HIV-1, including for the emergence and dissemination of drug resistance. Next to the rate at which mutations occur in the population, the migration rate and selection strength are important parameters in population genetic models. Together, these parameters control the dynamics of mutation frequencies over space and time.

In the RT-SHIV sequence data, we observe that both migration and selection are powerful drivers of allele frequency changes. Selection causes drug-resistant variants to reach high frequencies in the course of just weeks, and migration causes these same variants to be present throughout the body. However, it is interesting to note that neither force overwhelms the other. If selection was much stronger than migration, beneficial variants should reach high frequencies in their originating compartment before migrating to other compartments, but we observe the same beneficial variants in all compartments nearly simultaneously. On the other hand, if migration was a much stronger force than selection, we would expect to see very similar allele frequencies in all compartments. Instead we observe transient but significant compartmentalization, particularly after the onset of a selective pressure. This suggests that while there are instances where selection or migration may be acting more quickly, neither force is much stronger than the other. Since we can estimate the selection coefficient of certain mutations to be around 0.5 (0.41–0.7), migration must be of the same order of magnitude. Taken together, the balanced forces of strong selection and rapid migration create an incredibly dynamic system.

The number of beneficial mutations that arise per generation in the population is an important driver of how fast a population will evolve. This rate (the population mutation rate or θ) is the product of the number of genomes that can mutate (i.e., the population size) and the beneficial mutation rate. If the population mutation rate is much less than 1 mutation/generation, a long period of time will be required for a beneficial mutation to occur. If the population mutation rate is much larger than 1, mutations will happen often and we will observe multiple beneficial mutations spreading in the population before any mutation reaches 100% frequency (i.e., sweeps will be soft). Since the mutation rate is fixed, observing the number of adaptive mutations can inform our estimates of the population size.

For HIV-1, estimates for the census population size are very large (10^7–10^8 productively infected cells [66, 67]) Given the mutation rate of HIV-1, this would imply a large enough population mutation rate for beneficial mutations to occur on hundreds of backgrounds, and sweeps should therefore be exceptionally soft. We do not observe this in our data, which

implies that the census population size does not fully determine population dynamics. It is therefore useful to estimate an 'effective' population size (N_e) that better captures the number of replicating viruses over a certain timescale. One common approach to estimate N_e is by using the amount of neutral diversity, which led to the first estimates of N_e in HIV-1 to be $< 10^3$ [68]. Indeed, when we use population genetic diversity to estimate population size in our data, we also estimate values of N_e on this order of magnitude. However, if N_e was this small, the multiple beneficial mutations we observe spreading simultaneously in our data would be extremely unlikely. This shows that similar to the census population size, the diversity-based N_e also does not explain the dynamics of drug resistance emergence. We believe that the diversity-based N_e captures a long-term effective population size, which may be very small because it is strongly influenced by bottlenecks and selective sweeps, including the strong population bottleneck of initial infection. It may therefore greatly underestimate the size of the population that actually contributes to resistance evolution.

To estimate an N_e that appropriately captures the population dynamics at the time of drug resistance emergence, we use the number of origins of resistance, following Pennings, et al. [49]. Given that we observe 3–4 different drug-resistant backgrounds per population, we estimate that effective population size is on the order of 10^5, which is much lower than the census population size but higher than the diversity-based effective population size. The estimate is similar to the estimate of the HIV-1 effective population size [49], which further shows how RT-SHIV and HIV-1 have similar dynamics. Note that our estimate is a lower bound, given that we cannot distinguish when the same DRM occurs on the same background repeatedly, and so are likely undercounting the number of independent origins of drug resistance. In macaques like T98133, where there is little pre-existing variation before drug treatment, many origins may be missed. In macaques with much pre-existing variation, such as A99039, our estimates for the number of independent origins are likely to be more accurate, given that each drug-resistant genotype contains several unique non-DRMs.

Our findings of soft sweeps also support previous theoretical results about populations undergoing evolutionary rescue (i.e., populations decreasing in size that can only be 'rescued' by a beneficial mutation which make the growth rate positive). Wilson et al. predict that when there is a high probability that the population acquires an adaptive mutation before it becomes very small (i.e., rescue is likely), the adaptive mutation is likely to spread via a soft sweep [69]. In the RT-SHIV populations during drug pressure, all three macaques experienced rescue by development of DRMs and all three experienced soft sweeps.

As described above, we believe that the characterization of the RT-SHIV dynamics is relevant for HIV-1 populations, and specifically those HIV-1 populations treated with ineffective therapy, but this may not extend to suppressive ART. Treating HIV-1 with multiple drugs may cause viral populations to evolve with many fewer beneficial mutations. This can change the dynamics through which drug resistance spreads within a patient. An example of this was previously demonstrated, in which it was shown that more effective combinations of drugs decrease the number of origins of DRMs at treatment failure [54]. Further studies will be needed to determine if the level of dynamism that we observe here is also found in HIV-infected individuals treated with combinations of effective drugs.

Future studies have much to contribute to our growing understanding of intra-patient evolutionary dynamics. Single-genome sequencing provides full linked genotypes, which allow for the evolutionary origin of new mutations to be tracked. However, next generation sequencing could provide a greater depth of data with more sensitive allele frequency estimates, which in turn could be used to better differentiate population compositions across compartments. In addition, isolating and sequencing different types of CD4+ cells could reveal that different cell types form sub-compartments within those examined in this paper, as previous studies have

suggested they might (Table 1). To determine the physical source of DRMs, samples would likely need to be taken on shorter intervals and at greater depth to reveal very low frequency variants. Finally, in order to be able to observe additional origins of drug resistance, it may be possible in the future to use barcoded initial viral strains [70].

Our study reveals a dynamic intra-patient evolutionary process in RT-SHIV, featuring abundant mutations, fast migration, and strong selection. As we strive for better care for HIV-infected individuals, a thorough understanding of this process will be essential. RT-SHIV is a powerful tool for quantifying how and where drug resistance emerges, and locating potential drug sanctuary sites, which may accelerate the evolution of resistance. Further, the ability to sample in space presents new opportunities to characterize the latent reservoir and ultimately develop strategies to find a cure.

Materials and methods

Ethics statement

All animal-related work was conducted according to the Public Health Services (http://grants.nih.gov/grants/olaw/references/PHSPolicyLabAnimals.pdf). Washington National Primate Research Center (WaNPRC) is accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC) International and registered as a USDA Class R research facility. WaNPRC is certified by the NIH Office of Laboratory Animal Welfare (OLAW A3464-01). All animal-related experiments were performed under protocol 2370–25, approved by the University of Washington Institutional Animal Care and Use Committee. Animals were under the care of a licensed veterinarian and all efforts were made to minimize animal pain and suffering, in accordance with the recommendations of the Weatherall report (http://www.acmedsci.ac.uk/viewFile/publicationDownloads/1165861003.pdf). Temperature in animal quarters is maintained at 75–78˚F and humidity and air quality are controlled. Animals were monitored by veterinary staff at least once daily. Prior to infection, animals were pair housed. Following infection, macaques were housed separately to prevent cross-transmission but were provided regular visual, aural and olfactory contact with other macaques. Waste trays and cages were cleaned and sanitized daily and biweekly, respectively. Animals were fed three meals daily of a commercial monkey chow, which was supplemented with fresh fruits and vegetables, and given access to water ad libitum. Environmental enrichment was provided by a biweekly rotation of toys, use of foraging boards, social housing where permitted and/or positive human interaction, including food treats, with veterinary and research staff. All procedures were conducted while the animals were sedated with intramuscular injection of Telazol (tiletamine/zolazepam, 3-6mg/kg) for minor surgeries (e.g. biopsies) or with Ketamine (10–15 mg/kg) for blood draws and other minor procedures. Animals that were diagnosed by the veterinarian to be experiencing more than momentary pain and distress were evaluated and treated with appropriate analgesic drugs as indicated. Animals scheduled for necropsy were euthanized by intravenous injection of pentobarbital (80 mg/kg) in the saphenous vein after sedation with 3 mg/kg Telazol.

Animals

Four female pigtailed macaques (labeled T98133, A99165, A99039, and A01198) were infected intravenously with 1×10^5 infectious units of RT-SHIV$_{mne027}$, a simian immunodeficiency virus (SIV$_{mne027}$) encoding HIV-1$_{HxB2}$ RT [38, 39] and hereafter referred to as RT-SHIV. The animals were negative at study initiation for serum antibodies to HIV type 2, SIV, type D retrovirus, and simian T-lymphotropic virus type 1. Daily monotherapy was administered for 8 weeks between 12–20 weeks post-infection. T98133 and A99039 received FTC subcutaneously

(s.c.; 40 mg/kg) and animals A99165 and A01198 received 400mg EFV orally (p.o.). After a 6 week treatment interruption, daily combination ART was administered until necropsy. T98133 and A99039 received suppressive ART consisting of tenofovir (20 mg/kg s.c.), L-870812 (100mg p.o. twice daily), and darunavir/ritonavir (375 mg/50 mg p.o. twice daily). A99165 and A01198 received recycled therapy consisting of EFV (400 mg p.o.), FTC (40 mg/kg s.c.), and tenofovir (20 mg/kg s.c).

Blood was obtained in EDTA-treated tubes weekly or bi-monthly for peripheral lymphocyte counts, peripheral blood mononuclear cell (PBMC) pellets, and plasma pelletable vRNA quantification, as previously described [71–73]. The lower limit for quantification was 30 vRNA copies/ml plasma. Tissue biopsies were also taken at multiple time points throughout the study. Superficial LNs (i.e. axillary, inguinal) were taken through a <1 inch incision in the skin and CD4+ T cells were isolated from LNs by magnetic separation (Miltenyi). Rectal (gut) pinch biopsies were taken as previously described [74]. Briefly, sterile 2.0 mm biopsy forceps collected biopsies via an 8.9 mm diameter video gastroscope. Vaginal pinch biopsies (not full thickness) were obtained with a rigid vagiscope using biopsy forceps for a maximum of 4–6 biopsies (approximately 2–3 mm diameter) per time point. Plasma, PBMC, LN CD4+ T cells, and mucosal biopsies were stored at -80 C until use for nucleic acid extraction or drug concentration measurements.

Viral DNA and RNA isolation and single-genome sequencing (SGS)

Viral RNA and viral DNA were isolated from frozen samples as previously described [75]. SGS was performed on vRNA and vDNA by limiting dilution of cDNA made from RNA or DNA, as previously described [18, 39, 62]. Briefly, cDNA produced from viral RNA or viral DNA was diluted to a single copy and an amplicon of the first 770bp of RT was generated and sequenced, with a median of 28 sequences per sample (among samples with sequences) (**S1 Table**). Sequences containing mixtures were discarded. Sequences that identified more than one genome were not included in the analysis. DRMs were determined from the IAS-USA 2015 Update of the Drug Resistance Mutations in HIV-1 [76]. DRMs listed for drugs not being used to treat macaques were not assessed as DRMs.

Antiretroviral drug concentrations

Drug concentrations were quantified by LC-MS/MS analysis using methods similar to those previously published [77, 78]. Tissues were homogenized in 1mL of 70:30 acetonitrile-1mM ammonium phosphate (pH 7.4) with a Precellys 24 tissue homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France). Plasma underwent protein precipitation with an organic solution containing a stable-labeled internal standard. For all compounds, a Shimadzu high-performance liquid chromatography system was used for separation, and an AB SCIEX API 5000 mass spectrometer (AB SCIEX, Foster City, CA, USA) equipped with a turbo spray interface was used as the detector. Samples were analyzed with a set of calibration standards (0.02 to 20 ng/mL for tissue and 5–5,000ng/mL for plasma) and quality control (QC) samples. The precision and accuracy of the calibration standards and QC samples were within the acceptable range of 15%. Tissue density was assumed to be 1.06 g/cm3. All LC-MS/MS data underwent QC by a designated individual not directly involved in this study to ensure accuracy.

Muller plots

Muller plots show the change in frequency of different sampled genotypes over time separately by compartment and macaque. Time points were included in which at least five sequences of vRNA or vDNA were available in a compartment within a macaque (black vertical lines).

Time points with fewer than five sequences of vRNA or vDNA in a single compartment were excluded from the plot. Genotypes were called using all positions with minor allele variants at frequency higher than 1% within each macaque (across all compartments and time points). Genotypes with at least one DRM are marked with black hatching.

Phylogenetic analysis

Phylogenetic analysis of base pairs 135–900 using BEAST [79] using a HKY substitution model [80], estimated base frequencies, a strict clock and no site heterogeneity model. Default priors were used as provided by BEAUti 1.8.3, with a uniform clock rate prior of [0, 1] and a starting frequency of 0.005. The chain length was $5^{*}10^{7}$ with parameters recorded every $2.5^{*}10^{4}$ states resulting in 2000 trees from which maximum clade credibility trees were derived using the option "keep target heights." This process was repeated separately for vRNA and vDNA from each macaque. The trees were plotted using ggtree [81] and annotated with substitutions as described in **Materials and Methods: Muller Plots** and the location from which the samples were taken (i.e., PBMC, plasma, vagina, gut and LN). Branches were colored to match Muller plot coloration and internal branches with no predicted state were plotted in black.

Assessing viral compartmentalization

Compartmentalization of viral genotypes and change in their frequencies over time is assessed using K_{ST} [51, 82]. K_{ST} is computed as follows:

$$K_{ST} = \frac{\pi_t - \pi_s}{\pi_t}$$

where the total average pairwise diversity of the sample is given by

$$\pi_t = \sum_{i,j} x_i x_j \delta_{ij}$$

and the average pairwise diversity within each of k compartments is given by

$$\pi_s = \sum_k \frac{n_k}{n} \sum_{i,j} x_{i,k} x_{j,k} \delta_{ij}$$

and $x_{i,k}$ is the frequency of genotype i in compartment k, $x_i = \sum_k x_{i,k}$ is the frequency of genotype i across all compartments and δ_{ij} is the fraction of sites at which $i \neq j$.

The significance of K_{ST} was assessed via permutation test, in which compartmental identities were repeatedly reassigned without replacement. Labels were permuted 1,000 times, and significance was assessed with a one-sided test with larger values of K_{ST} indicating stronger evidence for compartmentalization. Because the power of this test was found to vary when compartments were subsampled to different sample sizes (S8 **Fig**), we performed a subsampling procedure to make compartmentalization results more comparable across compartments. We used a sample size of 10, which allowed us to retain much of the data for analysis while still having power to detect signals of compartmentalization visible at larger sample sizes (S8 **Fig**).

Each compartment was subsampled without replacement 1000 times, and these subsampled compartments were used to perform the compartmentalization tests. When compartments had fewer than 10 sequences at a given time point, they were excluded from the compartmentalization analysis. We report the proportion of these 1000 tests that result in a permutation test p-value ≤ 0.05. This can be interpreted as the probability of a pairwise significant test if 10 sequences are drawn randomly from each compartment.

K_{ST} permutation testing subsampled to 10 sequences was also used to assess compartmentalization between subsequent time points taken from the same compartment (following Achaz, et al. [83]).

For verification, compartmentalization tests were also performed using AMOVA [52] and the Slatkin-Maddison test [53]. AMOVA was performed in R pairwise on compartments subsampled to 10 sequences 1000 times using the package poppr [84] with 1,000 permutations and a quasi-euclidean transformation to convert distance matrices. AMOVA was found to have a low false positive rate and power at a sample size of 10 (S9 **Fig**).

The Slatkin-Maddison test was performed with 1,000 permutations as implemented in HyPhy [85] from trees constructed via UPGMA. For the Slatkin-Maddison test, we excluded duplicate sequences before performing the tests, following other papers [5, 33, 86]. When duplicate sequences were not removed, the test had a high false positive rate in assessing compartmentalization between two samples from the same compartment. Because compartments were required to have 10 sequences for comparison, not all tests performed with AMOVA and KST could be performed with the Slatkin-Maddison test (i.e., there were ≥ 10 sequences in a given compartment but <10 **unique** sequences). Because sample sizes of unique sequences were smaller than full sample sizes, only 100 sets of sequences of size 10 were used to perform the Slatkin-Maddison test. The resulting test was found to a have a low false positive rate and had power to detect compartmentalization in at least a subset of comparisons at a subsample size of 10 (S10 **Fig**).

Estimating selection coefficients for DRMs

Selection coefficients were estimated by assuming logistic growth of the drug resistant variants starting at the first treatment administration time point (week 12). This relies on the assumption that DRMs had reached establishment frequency ($1/Ns$) before treatment. We assumed generation times of 1/day, similar to HIV-1 and previously published [41] and effective population size (N_e) of $1.5*10^5$ [49] and then logistic growth of the allele frequency x as follows:

$$x = \frac{e^{st}}{e^{st} + 2N_e s}$$

where t is measured in generations. We fit s values using two time points–absence (or near absence) of DRMs at week 13 and the DRM frequency at the next available time point (week 15 for T98133 and A99165 and week 16 for A99039). In cases where multiple DRMs were present (e.g., M184V and M184I in T98133), a single s value was fit for both DRMs.

Viral diversity measurements

Intra-compartment diversity was measured for each compartment using the average pairwise diversity (see the definition for π_t in **Assessing Viral Compartmentalization** above). The relationship between diversity and time within each macaque was fit with a linear regression of intra-compartment diversity predicted by time (i.e., weeks after infection), separately for vDNA and vRNA. The difference between vRNA and vDNA was assessed by fitting two nested linear models of time on diversity–one in which an indicator variable discriminated between whether a diversity observation was from the vDNA or the vRNA, and one without an indicator variable. An ANOVA was performed to assess the significance of the indicator variable.

Estimating effective population size (N_e) of RT-SHIV

We used both nucleotide level diversity and genotype structure to estimate the effective population size of the intra-macaque RT-SHIV population. Pairwise nucleotide diversity relates to

effective population size N_e via the formula:

$$2N_e\mu = \pi_t$$

as described in **Assessing Viral Compartmentalization** above. Following Leigh-Brown [68], we estimated N_e using vRNA sequences taken from the blood plasma. It is well known that the estimate of N_e is much lower than the census population size, not only in HIV-1 but also in humans and fruit flies.

Aware of the problems with using neutral diversity to estimate N_e, Pennings et al. identified an alternative approach that should lead to an N_e estimate that is much closer to the short term N_e that is relevant for drug resistance evolution that can be estimated directly from the nucleotide level encodings of DRMs [49]. Briefly, the diversity of different DRM variants i through j can be summarized by the complement of their summed squared frequencies (or heterozygosity):

$$H = 1 - \sum_{x=i...j} x^2$$

N_e can then be estimated from the heterozygosity:

$$2N_e\mu = \frac{H}{1-H}$$

We calculate the heterozygosity using two different definitions of DRM variants: 1) DRM variants that are different by state (e.g., AAU versus AAC for K103N) and 2) DRM variants that can be distinguished by linkage to a common segregating variant (non-DRM mutation at a frequency of at least 10%). We label the resulting estimates of the effective population size $N_{e,DRM}$ and $N_{e,geno}$ respectively. Both $N_{e,DRM}$ and $N_{e,geno}$ are computed among the vRNA variants only and among all sampled variants (vRNA and vDNA) at the first time the vRNA is over 90% drug-resistant, or maximally drug-resistant if 90% drug resistance is never achieved. These times for T98133, A99165 and A99039 are weeks 15/16, week 15 and weeks 20/21, respectively.

All estimates rely on the HIV-1 mutation rate of $\mu = 1.4 * 10^{-5}$ for transitions (such as M184V) and $2 * 10^{-6}$ for transversions (both forms of K103N) [87].

Supporting information

S1 Table. Summary of available sequences from RT-SHIV RNA and DNA across compartments separated by macaque and sampling week.
(XLSX)

S2 Table. Linear model fits predicting the change in diversity (average pairwise distance within a compartment (π) over time (week post-infection)) separately for vRNA and vDNA by macaque.
(XLSX)

S1 Fig. Average pairwise nucleotide diversity within each compartment over time. The average pairwise diversity (π) increases over time in both the vRNA (top row) and vDNA (bottom row) in each macaque (shown in the columns). Each colored line marks a different compartment with the diversity of all compartments combined shown in black. The asterisk marks where diversity decreases at the final sampling point for A99039, and corresponds to viral suppression (see Fig 1). Grey shading indicates monotherapy or combination therapy, as indicated below the x-axis. Rx1 is treatment FTC+TFV+EFV and Rx2 is the treatment TFV+L870812+DRV/r.
(TIF)

S2 Fig. FTC and EFV plasma and tissue concentrations detected during monotherapy.
FTC and FTC-TP (A) and EFV (B) concentrations were measured in limited plasma (top) and
PBMC and mucosal biopsy samples (bottom) taken from all animals at multiple time points.
Grey shading indicates monotherapy (weeks 12–20). Open symbols represent measurements
that were below the lower level of quantitation. NA indicates sample was not available. Unfor-
tunately LN tissues were not available for drug measurements.
(TIFF)

**S3 Fig. Time-structured phylogenies among vDNA sequences from macaques with drug
resistance show multiple origins of drug resistance.** Branch lengths indicate sequence sam-
pling time from vDNA of macaques T98133 (A), A99165 (B), and A99039 (C). Branches are
colored to match the corresponding plots of Figs 3, 4 and 5(A) and 5(B) for T98133, A99165
and A99039, respectively. Sampling location is indicated for each sequence, and the identity of
all mutations at frequency >1% is shown to the right. Colors indicating mutation type (synon-
ymous, nonsynonymous, DRM, stop or nonsense/missense) are shown in the legend.
(TIF)

**S4 Fig. Time-structured phylogenies among vDNA and vRNA sequences A01198 shows
multiple origins of drug resistance.** Branch lengths indicate sequence sampling time from
the vRNA (A) and vDNA (B) of macaque A01198. Sampling location is indicated for each
sequence, and the identity of all mutations at frequency >1% is shown to the right. Colors
indicating mutation type (synonymous, nonsynonymous, DRM, stop or nonsense/missense)
are shown in the legend.
(TIF)

**S5 Fig. Compartmentalization relationships between different samples reveal transient
and stable patterns by Slatkin-Maddison tests.** The y-axis indicates the proportion of Slat-
kin-Maddison tests significant at the 5% significance level when subsampled to 10 sequences
per compartment 100 times, but the figure caption is otherwise shared with Fig 6.
(TIF)

**S6 Fig. Compartmentalization relationships between different samples reveal transient
and stable patterns by AMOVA tests.** The y-axis indicates the proportion of AMOVA tests
significant at the 5% significance level when subsampled to 10 sequences per compartment
1000 times, but the figure caption is otherwise shared with Fig 6.
(TIF)

**S7 Fig. vRNA and vDNA samples from the same compartment harbor variants private to
each other.** For a given compartment, the percentage of vRNA genotypes without correspond-
ing vDNA genotypes (top) as well as the the percentage of vDNA genotypes without corre-
sponding vRNA genotypes (bottom) is plotted at each time point, in which a genotype was
present in both vRNA and vDNA for a given compartment, macaque and time point at a fre-
quency greater than 3 sequences. Plasma vRNA is compared to PBMC vDNA (green), and all
other samples are compared within their sampled compartment (lymph node in purple, vagina
in light blue, gut in dark blue and PBMC vRNA verus vDNA in pink). The time point means
for both vDNA without vRNA and vRNA without vDNA are shown in black and show overall
trends across compartments. Time points that were within a week of each other were consid-
ered close enough to compare. Grey shading indicates monotherapy or combination therapy,
as indicated below the x-axis. Rx1 is treatment FTC+TFV+EFV and Rx2 is the treatment
TFV+L870812+DRV/r.
(TIF)

S8 Fig. Power of K_{ST} test for compartmentalization is sensitive to sample size. For each pairwise compartmental comparison, the proportion of 100 subsampled K_{ST} tests resulting in a permutation test p-value ≤ 0.05 is plotted versus the down sampling number separately by macaque and paired time points (i.e., time points within ±1 week). Sequences sampled from different macaques are shown in different colors, and compartments sampled from different weeks are represented with different symbols. Lines connect tests from the same macaque and the same week at different levels of subsampling. Because subsampling is done without replacement, lines extend only as far as the minimum of the sample size of the two compartments being compared. The time point represents the relative time of sampling, with time point 1 representing weeks 12 or 13, time point 2 representing weeks 15 or 16, time point 3 representing weeks 20 or 21, time point 4 representing weeks 26 or 27, and time point 5 representing week 29+.
(TIF)

S9 Fig. Power of Slatkin-Maddison test for compartmentalization is sensitive to sample size. The y-axis represents the proportion of 100 subsampled Slatkin-Maddison tests resulting in a permutation test p-value ≤ 0.05, but the caption is otherwise shared with S6 Fig. Slatkin-Maddison tests are done using only unique sequences, as described in the Materials and Methods.
(TIF)

S10 Fig. Power of AMOVA test for compartmentalization is sensitive to sample size. The y-axis represents the proportion of 100 subsampled AMOVA tests resulting in a permutation test p-value ≤ 0.05, but the caption is otherwise shared with S6 Fig.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: AFF SLH DAP PSP ZA.

Formal analysis: AFF MC ADMK BFK SLH DAP PSP ZA.

Investigation: CK PP BFK.

Writing – original draft: AFF DAP PSP ZA.

Writing – review & editing: AFF MC ADMK BFK SLH DAP PSP ZA.

References

1. Rhee S-Y, Blanco JL, Jordan MR, Taylor J, Lemey P, Varghese V, et al. Geographic and Temporal Trends in the Molecular Epidemiology and Genetic Mechanisms of Transmitted HIV-1 Drug Resistance: An Individual-Patient- and Sequence-Level Meta-Analysis. PLoS Med. 2015; 12(4):1–29.

2. Cory TJ, Schacker TW, Stevenson M, Fletcher CV. Overcoming pharmacologic sanctuaries. Current opinion in HIV and AIDS. 2013; 8:190–5. https://doi.org/10.1097/COH.0b013e32835fc68a PMID: 23454865

3. Lorenzo-Redondo R, Fryer HR, Bedford T, Kim E-Y, Archer J, Kosakovsky Pond SL, et al. Persistent HIV-1 replication maintains the tissue reservoir during therapy. Nature. 2016; 530(7588):51–6. https://doi.org/10.1038/nature16933 PMID: 26814962

4. Moreno-Gamez S, Hill AL, Rosenbloom DIS, Petrov DA, Nowak MA, Pennings PS. Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:E2874–83. https://doi.org/10.1073/pnas.1424184112 PMID: 26038564

5. Bull ME, Heath LM, McKernan-Mullin JL, Kraft KM, Acevedo L, Hitti JE, et al. Human immunodeficiency viruses appear compartmentalized to the female genital tract in cross-sectional analyses but genital lineages do not persist over time. The Journal of infectious diseases. 2013; 207:1206–15. https://doi.org/10.1093/infdis/jit016 PMID: 23315326

6. Kelley CF, Sullivan ST, Lennox JL, Evans-Strickfaden T, Hart CE. Lack of effect of compartmentalized drug resistance mutations on HIV-1 pol divergence in antiretroviral-experienced women. AIDS (London, England). 2010; 24:1361–6.

7. Fulcher JA, Hwangbo Y, Zioni R, Nickle D, Lin X, Heath L, et al. Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection. Journal of virology. 2004; 78:7883–93. https://doi.org/10.1128/JVI.78.15.7883-7893.2004 PMID: 15254161

8. Potter SJ, Lemey P, Dyer WB, Sullivan JS, Chew CB, Vandamme A-M, et al. Genetic analyses reveal structured HIV-1 populations in serially sampled T lymphocytes of patients receiving HAART. Virology. 2006; 348:35–46. https://doi.org/10.1016/j.virol.2005.12.031 PMID: 16455126

9. Delobel P, Sandres-Saune K, Cazabat M, L'Faqihi F-E, Aquilina C, Obadia M, et al. Persistence of distinct HIV-1 populations in blood monocytes and naive and memory CD4 T cells during prolonged suppressive HAART. AIDS (London, England). 2005; 19:1739–50.

10. van Marle G, Church DL, Nunweiler KD, Cannon K, Wainberg MA, Gill MJ. Higher levels of Zidovudine resistant HIV in the colon compared to blood and other gastrointestinal compartments in HIV infection. Retrovirology. 2010; 7:74-. https://doi.org/10.1186/1742-4690-7-74 PMID: 20836880

11. Rozera G, Abbate I, Vlassi C, Giombini E, Lionetti R, Selleri M, et al. Quasispecies tropism and compartmentalization in gut and peripheral blood during early and chronic phases of HIV-1 infection: possible correlation with immune activation markers. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2014; 20:O157–66.

12. Vanderford TH, Bleckwehl C, Engram JC, Dunham RM, Klatt NR, Feinberg MB, et al. Viral CTL escape mutants are generated in lymph nodes and subsequently become fixed in plasma and rectal mucosa during acute SIV infection of macaques. PLoS Pathog. 2011; 7(5):e1002048. Epub 2011/05/19. PubMed Central PMCID: PMCPMC3098234. https://doi.org/10.1371/journal.ppat.1002048 PMID: 21625590

13. Heeregrave EJ, Geels MJ, Brenchley JM, Baan E, Ambrozak DR, van der Sluis RM, et al. Lack of in vivo compartmentalization among HIV-1 infected naive and memory CD4+ T cell subsets. Virology. 2009; 393:24–32. https://doi.org/10.1016/j.virol.2009.07.011 PMID: 19698967

14. Wang YM, Dyer WB, Workman C, Wang B, Sullivan JS, Saksena NK. Molecular evidence for drug-induced compartmentalization of HIV-1 quasispecies in a patient with periodic changes to HAART. AIDS (London, England). 2000; 14:2265–72.

15. Lewis MJ, Frohnen P, Ibarrondo FJ, Reed D, Iyer V, Ng HL, et al. HIV-1 Nef sequence and functional compartmentalization in the gut is not due to differential cytotoxic T lymphocyte selective pressure. PloS one. 2013; 8:e75620-e. https://doi.org/10.1371/journal.pone.0075620 PMID: 24058696

16. Imamichi H, Degray G, Dewar RL, Mannon P, Yao M, Chairez C, et al. Lack of compartmentalization of HIV-1 quasispecies between the gut and peripheral blood compartments. The Journal of infectious diseases. 2011; 204:309–14. https://doi.org/10.1093/infdis/jir259 PMID: 21673043

17. van Marle G, Gill MJ, Kolodka D, McManus L, Grant T, Church DL. Compartmentalization of the gut viral reservoir in HIV-1 infected patients. Retrovirology. 2007; 4:87-. https://doi.org/10.1186/1742-4690-4-87 PMID: 18053211

18. Kearney MF, Anderson EM, Coomer C, Smith L, Shao W, Johnson N, et al. Well-mixed plasma and tissue viral populations in RT-SHIV-infected macaques implies a lack of viral replication in the tissues during antiretroviral therapy. Retrovirology. 2015; 12:93. PubMed Central PMCID: PMCPMC4642622. https://doi.org/10.1186/s12977-015-0212-2 PMID: 26559632

19. Josefsson L, Palmer S, Faria NR, Lemey P, Casazza J, Ambrozak D, 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 Pathog. 2013; 9(6):e1003432. https://doi.org/10.1371/journal.ppat. 1003432 PMID: 23818847

20. Boritz Eli A, Darko S, Swaszek L, Wolf G, Wells D, Wu X, et al. Multiple Origins of Virus Persistence during Natural Control of HIV Infection. Cell. 166(4):1004–15. https://doi.org/10.1016/j.cell.2016.06.039 PMID: 27453467

21. Haddad DN, Birch C, Middleton T, Dwyer DE, Cunningham AL, Saksena NK. Evidence for late stage compartmentalization of HIV-1 resistance mutations between lymph node and peripheral blood mononuclear cells. AIDS (London, England). 2000; 14:2273–81.

22. Overbaugh J, Anderson RJ, Ndinya-Achola JO, Kreiss JK. Distinct but related human immunodeficiency virus type 1 variant populations in genital secretions and blood. AIDS Res Hum Retroviruses. 1996; 12(2):107–15. https://doi.org/10.1089/aid.1996.12.107 PMID: 8834460

23. Si-Mohamed A, Kazatchkine MD, Heard I, Goujon C, Prazuck T, Aymard G, et al. Selection of drug-resistant variants in the female genital tract of human immunodeficiency virus type 1-infected women receiving antiretroviral therapy. J Infect Dis. 2000; 182(1):112–22. Epub 2000/07/06. https://doi.org/10. 1086/315679 PMID: 10882588

24. Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, et al. Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations. PLoS one. 2009; 4:e7122-e. https://doi.org/10.1371/journal.pone.0007122 PMID: 19771165

25. Chaudhary S, Noel RJ, Rodríguez N, Collado S, Munoz J, Kumar A, et al. Correlation between CD4 T cell counts and virus compartmentalization in genital and systemic compartments of HIV-infected females. Virology. 2011; 417(2):320–6. Epub 2011/07/13. PubMed Central PMCID: PMCPMC3204360. https://doi.org/10.1016/j.virol.2011.06.018 PMID: 21745672

26. Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, Nyange P, et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol. 1995; 69(12):8118–22. PubMed Central PMCID: PMCPMC189765. PMID: 7494333

27. Poss M, Rodrigo AG, Gosink JJ, Learn GH, de Vange Panteleeff D, Martin HL, et al. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J Virol. 1998; 72(10):8240–51. PubMed Central PMCID: PMCPMC110179. PMID: 9733867

28. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C, et al. HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc Natl Acad Sci U S A. 2003; 100(22):12972–7. Epub 2003/10/13. PubMed Central PMCID: PMCPMC240729. https://doi.org/10.1073/pnas.2134064100 PMID: 14557540

29. Tirado G, Jove G, Kumar R, Noel RJ, Reyes E, Sepulveda G, et al. Compartmentalization of drug resistance-associated mutations in a treatment-naive HIV-infected female. AIDS Res Hum Retroviruses. 2004; 20(6):684–6. https://doi.org/10.1089/0889222041217509 PMID: 15242547

30. Tirado G, Jove G, Kumar R, Noel RJ, Reyes E, Sepulveda G, et al. Differential virus evolution in blood and genital tract of HIV-infected females: evidence for the involvement of drug and non-drug resistance-associated mutations. Virology. 2004; 324:577–86. https://doi.org/10.1016/j.virol.2004.04.013 PMID: 15207642

31. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C, et al. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intra-patient recombination. Journal of virology. 2005; 79:353–63. https://doi.org/10.1128/JVI.79.1.353-363. 2005 PMID: 15596829

32. Lerner P, Guadalupe M, Donovan R, Hung J, Flamm J, Prindiville T, et al. The gut mucosal viral reservoir in HIV-infected patients is not the major source of rebound plasma viremia following interruption of highly active antiretroviral therapy. Journal of virology. 2011; 85:4772–82. https://doi.org/10.1128/JVI. 02409-10 PMID: 21345945

33. Evering TH, Mehandru S, Racz P, Tenner-Racz K, Poles MA, Figueroa A, et al. Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. PLoS Pathog. 2012; 8(2):e1002506. Epub 2012/02/02. PubMed Central PMCID: PMCPMC3271083. https://doi.org/10.1371/journal.ppat.1002506 PMID: 22319447

34. Barnett SW, Barboza A, Wilcox CM, Forsmark CE, Levy JA. Characterization of human immunodeficiency virus type 1 strains recovered from the bowel of infected individuals. Virology. 1991; 182(2):802–9. PMID: 2024498

35. Zhang L, Rowe L, He T, Chung C, Yu J, Yu W, et al. Compartmentalization of surface envelope glycoprotein of human immunodeficiency virus type 1 during acute and chronic infection. Journal of virology. 2002; 76:9465–73. https://doi.org/10.1128/JVI.76.18.9465-9473.2002 PMID: 12186928

36. Potter SJ, Dwyer DE, Saksena NK. Differential cellular distribution of HIV-1 drug resistance in vivo: evidence for infection of CD8+ T cells during HAART. Virology. 2003; 305(2):339–52. PMID: 12573579

37. Potter SJ, Lemey P, Achaz G, Chew CB, Vandamme A-M, Dwyer DE, et al. HIV-1 compartmentalization in diverse leukocyte populations during antiretroviral therapy. Journal of leukocyte biology. 2004; 76:562–70. https://doi.org/10.1189/jlb.0404234 PMID: 15218056

38. Ambrose Z, Boltz V, Palmer S, Coffin JM, Hughes SH, Kewalramani VN. In vitro characterization of a simian immunodeficiency virus-human immunodeficiency virus (HIV) chimera expressing HIV type 1 reverse transcriptase to study antiviral resistance in pigtail macaques. Journal of virology. 2004; 78:13553–61. https://doi.org/10.1128/JVI.78.24.13553-13561.2004 PMID: 15564466

39. Ambrose Z, Palmer S, Boltz VF, Kearney M, Larsen K, Polacino P, et al. Suppression of viremia and evolution of human immunodeficiency virus type 1 drug resistance in a macaque model for antiretroviral therapy. Journal of virology. 2007; 81:12145–55. https://doi.org/10.1128/JVI.01301-07 PMID: 17855539

40. Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, et al. Population genomics of intrapatient HIV-1 evolution. Elife. 2015; 4. PubMed Central PMCID: PMCPMC4718817.

41. Ambrose Z, Palmer S, Boltz VF, Kearney M, Larsen K, Polacino P, et al. Suppression of viremia and evolution of human immunodeficiency virus type 1 drug resistance in a macaque model for antiretroviral therapy. J Virol. 2007; 81(22):12145–55. https://doi.org/10.1128/JVI.01301-07 PMID: 17855539

42. Melody K, McBeth S, Kline C, Kashuba AD, Mellors JW, Ambrose Z. Low Frequency of Drug-Resistant Variants Selected by Long-Acting Rilpivirine in Macaques Infected with Simian Immunodeficiency Virus Containing HIV-1 Reverse Transcriptase. Antimicrob Agents Chemother. 2015; 59(12):7762–70. https://doi.org/10.1128/AAC.01937-15 PMID: 26438501

43. Solas C, Lafeuillade A, Halfon P, Chadapaud S, Hittinger G, Lacarelle B. Discrepancies between protease inhibitor concentrations and viral load in reservoirs and sanctuary sites in human immunodeficiency virus-infected patients. Antimicrob Agents Chemother. 2003; 47(1):238–43. PubMed Central PMCID: PMCPMC149042. https://doi.org/10.1128/AAC.47.1.238-243.2003 PMID: 12499197

44. Fletcher CV, Staskus K, Wietgrefe SW, Rothenberger M, Reilly C, Chipman JG, et al. Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. Proc Natl Acad Sci U S A. 2014; 111(6):2307–12. Epub 2014/01/27. PubMed Central PMCID: PMCPMC3926074. https://doi.org/10.1073/pnas.1318249111 PMID: 24469825

45. Pennings PS, Hermisson J. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Molecular biology and evolution. 2006; 23:1076–84. https://doi.org/10.1093/molbev/msj117 PMID: 16520336

46. McColl DJ, Margot N, Chen SS, Harris J, Borroto-Esoda K, Miller MD. Reduced emergence of the M184V/I resistance mutation when antiretroviral-naive subjects use emtricitabine versus lamivudine in regimens composed of two NRTIs plus the NNRTI efavirenz. HIV Clin Trials. 2011; 12(2):61–70. https://doi.org/10.1310/hct1202-61 PMID: 21498149

47. Schuurman R, Nijhuis M, van Leeuwen R, Schipper P, de Jong D, Collis P, et al. Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). J Infect Dis. 1995; 171(6):1411–9. PMID: 7539472

48. Palmer S, Boltz V, Maldarelli F, Kearney M, Halvas EK, Rock D, et al. Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant HIV-1 in patients starting and stopping non-nucleoside therapy. AIDS. 2006; 20(5):701–10. https://doi.org/10.1097/01.aids.0000216370.69066.7f PMID: 16514300

49. Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV. PLoS genetics. 2014; 10:e1004000-e. https://doi.org/10.1371/journal.pgen.1004000 PMID: 24465214

50. Wang J, Li D, Bambara RA, Yang H, Dykes C. L74V increases the reverse transcriptase content of HIV-1 virions with non-nucleoside reverse transcriptase drug-resistant mutations L100I+K103N and K101E+G190S, which results in increased fitness. J Gen Virol. 2013; 94(Pt 7):1597–607. Epub 2013/03/27. PubMed Central PMCID: PMCPMC3709637. https://doi.org/10.1099/vir.0.050914-0 PMID: 23535575

51. Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. Mol Biol Evol. 1992; 9(1):138–51. PMID: 1552836

52. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992; 131(2):479–91. PubMed Central PMCID: PMCPMC1205020. PMID: 1644282

53. Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 1989; 123(3):603–13. PMID: 2599370

54. Feder AF, Rhee S-Y, Holmes SP, Shafer RW, Petrov DA, Pennings PS. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. eLife. 2016; 5:e10670-e. https://doi.org/10.7554/eLife.10670 PMID: 26882502

55. Sathaliyawala T, Kubota M, Yudanin N, Turner D, Camp P, Thome JJ, et al. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. Immunity. 2013; 38 (1):187–97. PubMed Central PMCID: PMCPMC3557604. https://doi.org/10.1016/j.immuni.2012.09.020 PMID: 23260195

56. Yukl SA, Shergill AK, Girling V, Li Q, Killian M, Epling L, et al. Site-specific differences in T cell frequencies and phenotypes in the blood and gut of HIV-uninfected and ART-treated HIV+ adults. PLoS One. 2015; 10(3):e0121290. PubMed Central PMCID: PMCPMC4374729. https://doi.org/10.1371/journal.pone.0121290 PMID: 25811360

57. Kimata JT, Kuller L, Anderson DB, Dailey P, Overbaugh J. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence AIDS progression. Nat Med. 1999; 5(5):535–41. https://doi.org/10.1038/8414 PMID: 10229230

58. Okoye A, Meier-Schellersheim M, Brenchley JM, Hagen SI, Walker JM, Rohankhedkar M, et al. Progressive CD4+ central memory T cell decline results in CD4+ effector memory insufficiency and overt disease in chronic SIV infection. J Exp Med. 2007; 204(9):2171–85. PubMed Central PMCID: PMCPMC2118701. https://doi.org/10.1084/jem.20070567 PMID: 17724130

59. Picker LJ, Hagen SI, Lum R, Reed-Inderbitzin EF, Daly LM, Sylwester AW, et al. Insufficient production and tissue delivery of CD4+ memory T cells in rapidly progressive simian immunodeficiency virus infection. J Exp Med. 2004; 200(10):1299–314. PubMed Central PMCID: PMCPMC2211921. https://doi.org/10.1084/jem.20041049 PMID: 15545355

60. Picker LJ. Immunopathogenesis of acute AIDS virus infection. Curr Opin Immunol. 2006; 18(4):399–405. https://doi.org/10.1016/j.coi.2006.05.001 PMID: 16753288

61. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, Wind-Rotolo M, 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. J Virol. 2006; 80(13):6441–57. https://doi.org/10.1128/JVI.00591-06 PMID: 16775332

62. Kearney M, Spindler J, Shao W, Maldarelli F, Palmer S, Hu SL, et al. Genetic diversity of simian immunodeficiency virus encoding HIV-1 reverse transcriptase persists in macaques despite antiretroviral therapy. J Virol. 2011; 85(2):1067–76. https://doi.org/10.1128/JVI.01701-10 PMID: 21084490

63. Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, et al. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. PLoS Pathog. 2014; 10(3):e1004010. PubMed Central PMCID: PMC3961343. https://doi.org/10.1371/journal.ppat.1004010 PMID: 24651464

64. Maldarelli F, Kearney M, Palmer S, Stephens R, Mican J, Polis MA, et al. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. J Virol. 2013; 87(18):10313–23. PubMed Central PMCID: PMCPMC3754011. https://doi.org/10.1128/JVI.01225-12 PMID: 23678164

65. Kearney M, Spindler J, Shao W, Maldarelli F, Palmer S, Hu S-L, et al. Genetic diversity of simian immunodeficiency virus encoding HIV-1 reverse transcriptase persists in macaques despite antiretroviral therapy. Journal of virology. 2011; 85:1067–76. https://doi.org/10.1128/JVI.01701-10 PMID: 21084490

66. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature. 1997; 387:183–8. https://doi.org/10.1038/387183a0 PMID: 9144289

67. Haase AT. POPULATION BIOLOGY OF HIV-1 INFECTION: Viral and CD4+ T Cell Demographics and Dynamics in Lymphatic Tissues. Annu Rev Immunol. 1999; 17(1):625–56.

68. Brown AJL. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. Proceedings of the National Academy of Sciences of the United States of America. 1996; 94(5):1862–5.

69. Wilson B, Pennings P, Petrov D. Soft selective sweeps in evolutionary rescue. bioRxiv. 2016.

70. Del Prete GQ, Park H, Fennessey CM, Reid C, Lipkey L, Newman L, et al. Molecularly tagged simian immunodeficiency virus SIVmac239 synthetic swarm for tracking independent infection events. J Virol. 2014; 88(14):8077–90. PubMed Central PMCID: PMCPMC4097795. https://doi.org/10.1128/JVI.01026-14 PMID: 24807714

71. Li Y, Cleveland B, Klots I, Travis B, Richardson BA, Anderson D, et al. Removal of a single N-linked glycan in human immunodeficiency virus type 1 gp120 results in an enhanced ability to induce neutralizing antibody responses. J Virol. 2008; 82(2):638–51. PubMed Central PMCID: PMCPMC2224603. https://doi.org/10.1128/JVI.01691-07 PMID: 17959660

72. Polacino P, Cleveland B, Zhu Y, Kimata JT, Overbaugh J, Anderson D, et al. Immunogenicity and protective efficacy of Gag/Pol/Env vaccines derived from temporal isolates of SIVmne against cognate

virus challenge. J Med Primatol. 2007; 36(4–5):254–65. https://doi.org/10.1111/j.1600-0684.2007.00243.x PMID: 17669214

73. Suryanarayana K, Wiltrout TA, Vasquez GM, Hirsch VM, Lifson JD. Plasma SIV RNA viral load determination by real-time quantification of product generation in reverse transcriptase-polymerase chain reaction. AIDS Res Hum Retroviruses. 1998; 14(2):183–9. https://doi.org/10.1089/aid.1998.14.183 PMID: 9462929

74. Ho O, Larsen K, Polacino P, Li Y, Anderson D, Song R, et al. Pathogenic infection of Macaca nemestrina with a CCR5-tropic subtype-C simian-human immunodeficiency virus. Retrovirology. 2009; 6:65. Epub 2009/07/14. PubMed Central PMCID: PMCPMC2720380. https://doi.org/10.1186/1742-4690-6-65 PMID: 19602283

75. Kline C, Ndjomou J, Franks T, Kiser R, Coalter V, Smedley J, et al. Persistence of viral reservoirs in multiple tissues after antiretroviral therapy suppression in a macaque RT-SHIV model. PLoS one. 2013; 8: e84275-e. https://doi.org/10.1371/journal.pone.0084275 PMID: 24367650

76. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D, et al. 2015 Update of the Drug Resistance Mutations in HIV-1. Topics in Antiviral Medicine. 2015:132-. PMID: 26713503

77. Thompson CG, Bokhart MT, Sykes C, Adamson L, Fedoriw Y, Luciw PA, et al. Mass spectrometry imaging reveals heterogeneous efavirenz distribution within putative HIV reservoirs. Antimicrob Agents Chemother. 2015; 59(5):2944–8. PubMed Central PMCID: PMCPMC4394831. https://doi.org/10.1128/AAC.04952-14 PMID: 25733502

78. Cottrell ML, Yang KH, Prince HM, Sykes C, White N, Malone S, et al. A Translational Pharmacology Approach to Predicting Outcomes of Preexposure Prophylaxis Against HIV in Men and Women Using Tenofovir Disoproxil Fumarate With or Without Emtricitabine. J Infect Dis. 2016; 214(1):55–64. PubMed Central PMCID: PMCPMC4907409. https://doi.org/10.1093/infdis/jiw077 PMID: 26917574

79. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012; 29:1969–73. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

80. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of molecular evolution. 1985; 22:160–74. PMID: 3934395

81. Yu G, Smith D, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017; 8:28–36.

82. Charlesworth B, Charlesworth D. Elements of Evolutionary Genetics: Roberts & Company; 2010.

83. Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, et al. A robust measure of HIV-1 population turnover within chronically infected individuals. Molecular biology and evolution. 2004; 21:1902–12. https://doi.org/10.1093/molbev/msh196 PMID: 15215321

84. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014; 2:e281. Epub 2014/03/04. PubMed Central PMCID: PMCPMC3961149. https://doi.org/10.7717/peerj.281 PMID: 24688859

85. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21(5):676–9. https://doi.org/10.1093/bioinformatics/bti079 PMID: 15509596

86. Kumar SB, Handelman SK, Voronkin I, Mwapasa V, Janies D, Rogerson SJ, et al. Different regions of HIV-1 subtype C env are associated with placental localization and in utero mother-to-child transmission. J Virol. 2011; 85(14):7142–52. Epub 2011/05/04. PubMed Central PMCID: PMCPMC3126595. https://doi.org/10.1128/JVI.01955-10 PMID: 21543508

87. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. Journal of virology. 2010; 84:9864–78. https://doi.org/10.1128/JVI.00915-10 PMID: 20660205