HIV maintains an evolving and dispersed population among multiple tissues during suppressive cART with periods of rapid expansion corresponding to the onset of cancer

Running title: HIV Evolution During cART

Rebecca Rose[1*], Susanna L. Lamers[1*], David J. Nolan[1,2], Ekaterina Maidji[3], Nuno Faria[4], Oliver G. Pybus[4], James J. Dollar[2], Samuel A. Maruniak[2], Andrew C. McAvoy[2], Marco Salemi[2], Cheryl Stoddart[3], Elyse Singer[5], Michael S. McGrath#[3,6]

1. Bioinfoexperts, LLC, Thibodaux, LA, USA 70302. 2. Department of Pathology, Immunology and Laboratory Medicine and the Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA 32610. 3. Department of Laboratory Medicine, Pathology, and Medicine, University of California at San Francisco, CA, USA 94143. Department of Zoology, University of Oxford, Oxford, UK OX13PS. 5. The National Neurological AIDS Bank at the University of California at Los Angeles, CA, USA 90095. 6. The AIDS and Cancer Specimen Resource, University of California at San Francisco, CA 94143, USA.

* Equal contributors

Corresponding Author:

Michael McGrath
Department of Laboratory Medicine
University of California, San Francisco, California, USA 94110
Email: MMcGrath@php.ucsf.edu
Phone: 650-347-1258 Fax: 415-206-3765

Word Count: 4728

ABSTRACT

While combined antiretroviral therapy (cART) can result in undetectable plasma viral loads, it does not eradicate HIV infection. Furthermore, HIV-infected individuals while on cART remain at an increased risk of developing serious co-morbidities, such as cancer, neurological disease, and atherosclerosis, suggesting that during cART, tissue-based HIV may contribute to such pathologies.

We obtained DNA and RNA *env*, *nef* and *pol* sequences using single genome sequencing from *post mortem* tissues of three HIV+/cART+ individuals with undetectable viral load and metastatic cancer at death, and performed time-scaled Bayesian evolutionary analyses. We used a sensitive *in situ* hybridization technique to visualize HIV *gag-pol* mRNA transcripts in cerebellum and lymph node tissues from one patient.

Tissue-associated virus evolved at a similar rate in cART+ and cART- patients. Phylogenetic trees were characterized by two distinct features: 1) branching patterns consistent with constant viral evolution and dispersal amongst tissues; and 2) very recently derived clades containing both DNA and RNA sequences from multiple tissues. Cancer diagnoses were temporally associated with diversification of viral lineages. Rapid expansion of virus near death corresponded to wide-spread metastasis. HIV RNA+ cells clustered in cerebellum tissue but were dispersed in lymph node tissue, mirroring the evolutionary patterns observed for that patient. Activated, infiltrating macrophages were associated with HIV-expressing cells.

Our data provide evidence that tissues serve as a sanctuary for wild-type HIV during cART and suggest the importance of macrophages as an alternative reservoir and mechanism of virus spread.

IMPORTANCE

Combined anti-retroviral therapy (cART) reduces plasma HIV to undetectable levels; however, removal of cART results in plasma HIV rebound, thus highlighting its inability to entirely rid the body of infection. Additionally, HIV-infected individuals on cART remain at high risk of serious diseases, which suggests a contribution from residual HIV. Here, we isolated and sequenced HIV from *post mortem* tissues from three HIV+/cART+ individuals who died with metastatic cancer and had no detectable plasma viral load. Using high-resolution evolutionary analyses, we found that tissue-based HIV continues to replicate, evolve and migrate among tissues during cART. Furthermore, cancer onset and metastasis coincided with increased HIV diversity, suggesting a linked mechanism. HIV-expressing cells were associated with tissue macrophages, a target of HIV infection. Our results suggest the importance of tissues, and macrophages in particular, as a target for novel anti-HIV therapies.

INTRODUCTION

Although current combined antiretroviral therapy (cART) regimens can extend the lives of individuals infected with human immunodeficiency virus (HIV) by years or decades, two major problems remain. First, HIV is never eradicated in treated patients and plasma viral loads (VL) will rebound if treatment is stopped. Subsets of T-cells (7, 9, 10, 16, 17, 19, 40) and lymph (26, 33) are considered likely reservoirs of virus during cART. Virus obtained from circulating T-cells during suppressive therapy showed lack of evolution compared to pre-therapy virus, suggesting maintenance through latency rather than replication (4, 27). On the other hand, virus from lymph nodes appears to continue evolving throughout therapy (33) and is a heterogeneous population (26, 33). These contrasting patterns suggest multiple evolutionary strategies for maintenance during cART.

Second, significant HIV co-morbidities including cancer, lipid disorders and neurological diseases develop in cART-treated patients at a higher rate than the general population, despite fully suppressed VL and restored immunity. Of the major HIV-related co-morbidities, cancer is the leading cause of death for HIV-infected, cART-treated patients (6, 11, 48). However, the relationship between HIV infection, tumor growth, and metastasis is still largely unknown. In a previous study, we sequenced HIV DNA from tumor and non-tumor *post mortem* tissues from two cART-naïve patients and found that virus in tumors was evolutionarily distinct from virus in non-tumor tissues (43). We also recently quantified viral DNA in more than 65% of *post mortem* tissue samples obtained from a cohort of 20 HIV-infected patients who were virally suppressed at death, 15 of whom were diagnosed with cancer during treatment and/or at autopsy. HIV DNA was detected in twelve different anatomical locations, including tumor sites. Complete clinical

details of the study are reported elsewhere (S. Lamers *et al*. submitted for publication, *J Virol,* January 2016).

From this unique cohort, we selected a subset of five cancer patients to characterize tissue-based virus. We extracted HIV DNA and RNA from tissues and used single-genome sequencing to obtain a 3000bp region of the HIV genome spanning *env* through *nef*. We investigated whether viral populations were dispersed amongst tissues, if the viral populations were evolving during cART, and whether viral dynamics were associated with development and spread of cancer.

METHODS

NNAB participants. NNAB participants are recruited from the greater Los Angeles/California area. An IRB-approved informed consent is obtained from each participant or his/her legal guardian. Eligibility criteria are as follows: Participants must be aged 18 years or older and agree to participate in study examinations and donation of blood, urine, and (optional) cerebrospinal fluid (CSF) during life and to donate their brain and other tissues for research in the event of their death. HIV positive participants are selected because they have one or more of the following: a CD4+ count <50 cells/mm^3; systemic lymphoma or another widespread malignancy; mycobacterium avium complex infection; wasting with loss of ≥30% of body weight; primary central nervous system (CNS) lymphoma; progressive multifocal leukoencephalopathy; congestive heart failure; chronic renal failure with potential or current need for dialysis; chronic obstructive pulmonary disease; end-stage liver disease; serum albumin ≤3.2 g/dL; or any condition which, in the opinion of the study physician, was likely to lead to death within the period of study. In 2013, the criteria were amended to recruit a limited number

of older (age >60 years) HIV^+ participants. There are no exclusions based on race, ethnicity, language, immigration status, socioeconomic status, gender, substance use, or sexual preference. Most participants are recruited *pre mortem*; however, the remains of very recently deceased individuals may be donated by legal heirs, in which case information was obtained from guardians and/or by authorized release of medical records. Extensive details concerning NNAB guidelines, questionnaires and medical assessments are available in S. Lamers *et al*. (submitted for publication, *J Virol*, 2016).

The AIDS and Cancer Specimen Resource (ACSR). The ACSR is a National Cancer Institute funded program that supplies specimens from HIV positive participants with cancer to investigators interested in AIDS malignancy research. For this study, the ACSR independently verified the tissue diagnoses made by the NNAB and evaluated levels of HIV DNA. All of the cases described here reside at both the ACSR and the NNAB. The use of these materials is acknowledged to be a product of both NIH funded programs. This study has University of California at San Francisco IRB approval (#13-12020).

Patient cohort. For the present study, we selected a subset of five participants from the NNAB studies with tissues available from the post mortem autopsy. These patients were chosen because they had substantial cancer pathology at death and an undetectable (<40 HIV copies/mm fluid) VL at autopsy from the CSF and/or the blood (as measured from the cardiac aspirate). Extensive medical histories were available and described in detail elsewhere (S. Lamers *et al*, submitted for publication, *J Virol*, 2016). All tissues used in this study were histologically examined by ACSR pathologists and classified based on the pathology that was present in tissues extracted for the HIV DNA studies. For a subset of analyses, we also included viral sequences from two additional patients (designated KS1 and KS3), who died prior to the availability of

cART from advanced AIDS, including Kaposi's sarcoma. Similar to the participants in the current study, KS1 and KS3 had multiple *post mortem* tissues available through the AIDS and Cancer Specimen Resource (ACSR). Fewer clinical details were available for the two KS patients, although we also had exact dates of death and self-reported dates of infection. Details on all patients are reported in **Table 1**.

RNA/DNA extraction. Tissues were chosen for sequence analysis from the NNAB cohort participants if they were HIV DNA positive using digital droplet PCR (ddPCR; details for ddPCR experiments is presented in S. Lamers *et al*., submitted for publication, *J Virol*, 2016). An additional two tissues were included (lung from patient C02 and aorta from patient C05) that were ddPCR negative. KS1 tissues studied included skin tumor, lymph node tumor, lymph node and kidney. KS3 tissues included skin tumor, small bowel tumor, liver and stomach. All tissues from KS1 and KS3 were HIV positive. Total RNA and genomic DNA were isolated simultaneously from each tissue section (30-50ng) using the AllPrep DNA/RNA Mini Kit (Qiagen #80204). Tissues were homogenized just prior to extraction in Buffer RLT Plus (lysis buffer) using a TissueRupter rotor-stator homogenizer (Qiagen #9001271) with a fresh sterile disposable probe (Qiagen #990890) for each sample. Manufacturer's guidelines were followed, with the exception of two 50µl final elutions using RNase-free water during the RNA isolation. The 100µl final volume of RNA generated was concentrated using RNeasy MinElute Cleanup Kit (Qiagen #74204). RNA and DNA extractions, cDNA synthesis and first round PCR setup were conducted in a restricted-access, amplicon-free room with separate air handling and laboratory equipment where no amplified PCR products or recombinant cloned plasmids were allowed, and work surfaces and equipment were thoroughly cleaned before and after use with Eliminase® (Decon Labs, Inc.). cDNA was synthesized using the SuperScript® III First-Strand

Synthesis System (Invitrogen Life Technologies #18080-051) using the provided Oligo$(dT)_{20}$ primer according to manufacturer's recommendations. RNA was incubated at 65ºC for 5 minutes with deoxynucleoside triphosphates (0.5 mM) and 5μM Oligo$(dT)_{20}$, then cooled quickly to 4ºC. First-Strand cDNA synthesis was performed in a 40ul reaction volume containing 1X reverse transcription buffer (10mM Tris-HCl [pH 8.4], 25 mMKCl), 5mM $MgCl_2$, 10mM ditiothreitol, 2U/μl of RNase-OUT™ (RNase inhibitor), and 10U/µlSuperScript® III RT. The reaction was heated to 45ºC for 90 minutes, followed by 85ºC for 5 minutes. The reaction was cooled to 37ºC and 0.1U/μl of E. coli RNase H was added, followed by a 20-minute incubation. cDNA was stored at -20ºC.

Single Genome Sequencing. A modified single genome sequencing protocol based on previously published methods was used to prevent PCR resampling (38). cDNA and genomic DNA were serially diluted until an average of 30% or less of the nested PCR reactions were positive. During the first round PCR, diluted cDNA or genomic DNA was amplified in 20ul reactions containing 1X Platinum® Blue PCR SuperMix (Invitrogen Life Technologies) and 0.2μM of each primer: BEF1, 5'- TAATAGCAATAGTTGTGTGG-3' and BNR1, 5'-AGCTCCCAGGCTCAGATCT-3' (6111-6130 and 9558-9576 of HIV-1 HXB2, respectively) with the following cycling parameters – initial denaturation 94ºC for 3 minutes, then 40 cycles of 94ºC for 30 seconds, 56ºC for 30 seconds, 72ºoC for 4 minutes, followed by 72ºC for 10 minutes. Second round glycoprotein 120 (gp120) PCR consisted of 2ul of the first round PCR added to a 20ul second round reaction consisting of 1X Platinum® Blue PCR SuperMix (Invitrogen Life Technologies) and 0.2μM of each primer: BEF2, 5'- CAATAGTTGTGTGGTCCATAG-3' and BER2, 5'- CAACAGATGCTGTTGCGC-3' (6117-6137 and 7905-7922 of HIV-1 HXB2 respectively) with the following cycling parameters: initial denaturation 94ºC for 3 minutes, then

40 cycles of 94ºC for 30 seconds, 56ºC for 30 seconds, 72ºC for 3 minutes, followed by 72ºC for 10 minutes. This second round PCR generates a 1.8Kb product containing a complete envelope gp120 (*env*) sequence. Second round *env* PCRs were visualized on 1% agarose gels stained with ethidium bromide, and reactions containing a single 1.8Kb product were considered positive and selected for sequencing with BEF2 and BER2. Subsequently, the first round reactions that corresponded to positive second round *env* PCRs were then used to amplify the *nef* gene sequence; second round *nef* PCR consisted of 2ul of the first round PCR added to a 20ul second round reaction consisting of 1X Platinum® Blue PCR SuperMix (Invitrogen Life Technologies) and 0.2μM of each primer: BNF1, 5'-CTGGCTGTGGAAAGATACCT-3' and BNR2, 5'-ATCTGAGGGCTCGCCACT-3' (7965-7984 and 9488-9505 of HIV-1 HXB2, respectively) with the following cycling parameters: initial denaturation 94ºC for 3 minutes, then 40 cycles of 94ºC for 30 seconds, 58ºC for 30 seconds, 72ºC for 2 minutes, followed by 72ºC for 10 minutes. Second round *nef* PCR products were visualized on 1% agarose gels stained with ethidium bromide, and reactions containing single 1.5Kb products were considered positive and selected for sequencing with BNF1 and BNR2. The primers were designed using Primer3 and observing regions of conservation in alignments of published HIV-1 subtype B sequences downloaded from Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov).

In order to examine tissue-based HIV for drug resistance mutations, a 1.3 kb fragment of genomic DNA encompassing HIV-1 protease and the first 300 codons of RT (*pol*) was sequenced from DNA isolated from each tissue using a limiting dilution PCR modification of previously described methods (39, 45). During the first round PCR, diluted genomic DNA was amplified in 20ul reactions containing 1X Platinum® Blue PCR SuperMix (Invitrogen Life Technologies) and 0.2μM of each primer: MAW26, 5'- TTGGAAATGTGGAAAGGAAGGAC

-3' and RT21, 5'- CTGTATTTCTGCTATTAAGTCTTTTGATGGG -3' (2028-2050 and 3509-3539 of HIV-1 HXB2, respectively) with the following cycling parameters: initial denaturation 95ºC for 3 minutes, then 35 cycles of 94ºC for 15 seconds, 55ºC for 30 seconds, 72ºC for 2 minutes, followed by 72ºC for 10 minutes. Second round PCR consisted of 2ul of the first round PCR added to a 20ul second round reaction consisting of 1X Platinum® Blue PCR SuperMix (Invitrogen Life Technologies) and 0.2μM of each primer: PRO1, 5'-CAGAGCCAACAGCCCCACCA-3' and RT20, 5'-CTGCCAGTTCTAGCTCTGCTTC-3' (2147-2166 and 3441-3462 of HIV-1 HXB2, respectively) with the following cycling parameters: initial denaturation 95ºC for 3 minutes, then 35 cycles of 94ºC for 15 seconds, 63ºC for 30 seconds, 72ºC for 2 minutes, followed by 72ºC for 10 minutes. Second round *gp120* PCR products were visualized on 1% agarose gels stained with ethidium bromide, and reactions containing a single 1.3Kb product were considered positive and selected for sequencing with PRO1 and RT20.

Sequencing was performed on an Applied Biosystems 3730xl DNA Analyzer (Life Technologies) at the University of Florida Interdisciplinary Center for Biotechnology Research (UF ICBR). All sequences were assembled with the Geneious R7 software package (Biomatters http://www.geneious.com)

Sequence alignment. Sequences were aligned using ClustalW (51) in MEGA5 (49) with further optimization performed by hand. The final *env* and *nef* alignments spanned from positions 6213-7823 and positions 8797-9411 relative to the HXB2 genome, respectively. Due to a large number of insertions and deletions that are typically problematic to align and may bias phylogenetic analysis, hyper-variable regions in *env* (V1, V2 and V4 domains) were excluded. A preliminary maximum-likelihood phylogeny for each gene was estimated using sequences from

all participants to ensure no cross-contamination of patients was present. Sequences were tested for the presence of hyper-mutations using the HYPERMUTE tool (http://www.lanl.gov); sequences with a p-value of <0.01 were removed from the alignments. Sequences have been submitted to GenBank (Accession numbers KU708874-KU709831).

RNA/DNA population structure. Pairwise genetic distances (PND) were computed using the Tamura-Nei 93 (TN93) model of nucleotide substitution using a publically available script (https://github.com/veg/tn93). We used the Fst statistic, which measures population sub-division, to determine whether the RNA and DNA populations were significantly different in patients C02 and C05. 1000 bootstrap replicates were used to assess significance of the Fst values. This analysis was implemented in HYPHY (41).

Phylogenetic Analysis. To determine relationships among sequences, individual maximum-likelihood (ML) phylogenies for each patient were estimated using PhyML (25) under a general time reversible (GTR) nucleotide substitution model and gamma distributed rate variation among sites (+Γ). Statistical support was assessed with 200 bootstrap replicates.

Bayesian time-scaled analysis. The Bayesian phylogenetic approach implemented in BEAST v.1.8.2 (13) was used to generate rooted, time-measured phylogenies for three ART-treated patients (C02, C04, C05). To compare evolutionary patterns among cART-treated and cART-naïve cancer patients, we also undertook Bayesian phylogenetic analyses on tissue virus from two additional ART-naïve cancer patients (KS1, KS3). A relaxed molecular clock model was used with a log-normal model of among-branch rate variation. For each patient a normally distributed prior distribution was placed on the root height parameter; the mean of this distribution was set to the reported date of infection (**Table 1**) with a standard deviation of 0.5. For Bayesian phylogenetic analyses we used the Hasegawa-Kishino-Yano (HKY) model of

nucleotide substitution with gamma distributed among site rate variation (+Γ). Separate substitution models were specified for 1+2 codon positions versus 3rd codon positions. The Bayesian skygrid coalescent prior was used (99 categories). Analyses were run for 100 million steps or until chain convergence, as assessed using Tracer (ESS values >200). In some cases, the results from multiple analyses were combined. The maximum clade credibility (MCC) tree was reconstructed from the posterior distribution, minus a 10% burn-in. Each MCC tree was visualized and annotated in FigTree v.1.4.2 (programs available at http://tree.bio.ed.ac.uk/software/).

RNAScope and immunohistochemistry. We used an *in situ* hybridization technique (RNAscope) developed by Advanced Cell Diagnostics (ACD) to identify tissue-based cells expressing HIV *gag-pol* mRNA transcripts. This technology uses a "double Z" probe design, which greatly increases signal-to-noise ratio with visualization of single mRNA transcripts. RNAscope was performed using the 2.0 HD Reagent Kit-RED kit (ACD, cat. #310036) according to the manufacturer's instructions. HIV mRNA was detected using a HIV-*gag-pol* probe (ACD, cat. #317691). The RNAscope assay was followed by standard immunohistochemistry (IHC) for macrophage markers using mouse mAb to CD163 (Novocastra) and CD68 (Dako) (both brown), CD68 and CD163 Nuclei were counterstained with hematoxylin QS (Vector lab, cat. #H-3404)). Tissue sections were analyzed with a Leica DM6000 B microscope equipped with a Leica DFC 500 camera and LAS v4.3 software.

RESULTS

HIV RNA/DNA sequences were obtained from tissues in four of the five participants. Total DNA and RNA was extracted from each ddPCR+ tissue for five participants

and subjected to first-round single genome amplification of the HIV *env-nef* domains, followed by second-round amplification of separate *env* and *nef* gene sequences. HIV *pol* DNA was amplified from a subset of these tissues to examine sequence data for drug resistance mutations. Sequences were generated from 54% of the ddPCR HIV+ tissues (*n*=19), and at least one tissue in four of the five patients (**Table 2**).

RNA and DNA sequences arose from same population. For participants C02 and C05, we used the Fst statistic to test whether sequences obtained from RNA and DNA represented genetically distinct sub-populations. No significant differences were found in either participant for either *env* or *nef* (Fst = -0.012, p=0.894 for C02; Fst = 0.020, p=0.23 for C05). Therefore DNA and RNA sequences from each tissue were grouped together for the remainder of analyses. We did not apply this test between tissues, as in some cases too few sequences were present for meaningful analysis.

***Env* and *nef* from both cART+ and cART- patients showed similar diversity.** For each patient we then calculated NPD among all *env* and *nef* sequences (**Figure 1**). For the *env* gene, patients C02, C04, and C05 exhibited lower mean NPD than the two cART-naive patients, although the interquartile range (IQR) for C02 and C05 overlapped those for KS1 and KS3, while the IQR for C04 was less than both KS patients. However, for the *nef* gene, the mean NPD for C05 was similar to that of KS1 and KS3, and the IQR for all participants overlapped.

Anatomically dispersed viral populations evolve via different modes of evolution. In patient C02 (**Figure 2a**), many brain-derived sequences cluster in four distinct clades (A-C), each of which has a very recent common ancestor, and most of the remaining sequences from brain, lymph node tumor and lung have only a few or no closely related sequences and therefore descend from long terminal branches. The majority of DNA and RNA sequences in clades A-C

are identical. This pattern is consistent with a sudden and recent expansion of a population of viruses that has been genetically distinct from the remainder. However, note that these clades are not exclusively comprised of virus from the brain, as three lymph node tumor sequences are among them, and both RNA and DNA sequences are present. A few lymph node tumor derived sequences are present in clades A and B., with a potential migration event(s) between lymph node tumor and brain. This patient was reportedly on cART throughout infection, although clinical records indicate adherence may have been sub-optimal. However, the patient was in a managed care facility before death where medication was administered by staff. The patient's last VL measurement at death was undetectable and no virus was detected at autopsy in either plasma or CSF. His only cancer diagnosis occurred at autopsy, when plasmacytoma was diagnosed in multiple tissues, along with meningeal infiltrate into the brain. The lack of detection of this cancer while alive suggests a late onset and rapid spread. The *nef* gene showed a similar pattern (**Figure 2b).** We did not apply a test for significance of population structure among tissues as clearly there are multiple patterns of gene evolution within the tree, which would be masked by a simple statistic.

For patient C04, there were two well-supported clades, one containing only lung sequences (E), and the other containing lung and colon sequences (F, **Figure 3**). Note that the scale of substitution for *env* is much different compared to C02 and reflects the relatively low diversity within this patient (Figure 3a). The nef gene showed more variation and longer branches, and the lung-exclusive clade is now within the rest of the tree and contains a kidney sequence that is identical to the lung (Figure 3b). Additionally, clade G contained identical sequences from lung and colon, suggesting minimal barriers to gene flow among tissues. This patient, who was receiving fully suppressive cART since being diagnosed with HIV infection,

also experienced treated yet unresolved lymphoma until death. No RNA sequences were obtained for this patient, suggesting that proviruses were not being expressed at the time of death (or the RNA recovered from the autopsy tissue was degraded).

The patient C05 *env* phylogeny (**Figure 4a**) shows a pattern similar to patient C02, in which some clades (e.g. H) contain identical RNA and DNA sequences from, in this case, the spleen (H), or spleen and tumor lymph node (I), while other clades contain virus from multiple tissues that appears to be accumulating mutations (J). A clade with a very long branch (K) is in both trees and contains DNA sequences from kidney and pancreas, which suggests a potential compartmentalization and lack of replication, although in *nef* one spleen sequence is closely related and other kidney sequences are found scattered in the nef tree. The clades of identical sequences present in *env* did not have a *nef* counterpart, and thus those clades disappear in the *nef* tree. This patient received a lymphoma diagnosis around 2 years before death. Late in life, this patient's cancer spread to the liver, leading to death.

Evolutionary rates in *env* and *nef* are similar in ART-infected and ART-naïve viral populations. We then estimated rates of HIV sequence evolution for these three patients using a Bayesian analysis in which a strong prior was placed on the root of the tree, corresponding to the patients' reported date of diagnosis. We performed the same analysis using two additional patients, who died with KS prior to the era of cART. These patients also had known dates of death and HIV diagnosis. These results are shown in Figure 5. For comparison, we included on the graph viral evolutionary rates obtained from an earlier study (26) that analyzed HIV *gag* gene evolution between pre-therapy and during suppressive- therapy. Finally, we included the within-host evolutionary rates reported in (15) and (32).

The evolutionary rates for both *env* and *nef* in patients KS1 and KS3 were similar. Estimated *env* gene rates for patients C02 and C05 were not significantly different to the *env* rates for KS1 and KS3. C04 did exhibit a significantly lower rate than the other patients; however this rate was still $>1x10^{-3}$. All three cART-treated patients had *higher* mean rates for *nef* than the cART-naïve patients, although the distributions overlapped for all patients (**Figure 5**). These rates are all similar to those reported in two previous studies (shown on the chart). However, these results are in stark contrast to those found by Josefsson *et al* (26), who reported that the HIV evolutionary rate was reduced by orders of magnitude when HIV in T-cells during suppressive cART was compared to plasma-derived virus sampled pre-cART. The rates estimated from our patients are similar to that estimated from one ART-naïve patient using a similar approach (3).

In addition, in our study, we found that among-branch rate variation was extremely high in nearly all patients and genes (**Table 3**), consistent with the patients harboring viruses that evolve at different speeds. We note that while the viral population sampled at death may have a more recent temporal origin than the date of HIV diagnoses, the evolutionary rates would then be expected to increase, since the same amount of variation would need to have arisen in a shorter period of time. Although the date of HIV diagnosis may have been later than the true date of HIV infection, the precise clinical records of these patients indicates major health complications at the time of diagnosis, suggesting frequent health care and opportunities for testing.

Wild-type viruses evolve in tissues in ART-treated patients. We sequenced the *pol* gene from a subset of the tissues for each of the three patients to determine whether the viral populations in tissues were wild-type virus, and therefore shielded from the effects of ART, or whether they were drug resistant, which would suggest some mechanism by which viruses

replicated in tissues but not in blood T-cells. Only one sequence (in participant C05) out of a total 88 *pol* sequences analyzed carried a drug resistance mutation (against non-nucleoside reverse-transcriptase inhibitors; NNRTI). This patient was on nevirapine early in his treatment but was quickly switched to non-NNRTIs.

HIV RNA *gag-pol* transcripts are present in cellular clusters in cerebellum and singularly in lymph node tumor. Images from the RNAscope analysis showed that HIV mRNA transcripts for *gag-pol* were present within cells and in extra-cellular space in both the lymph node tumor and cerebellum in patient C02, the same tissues in which both viral DNA and RNA sequences were also recovered (**Figure 6**). In the lymph node tumor, HIV-expressing cells were typically localized distinct from each other (**Figure 6a-b**). In contrast, HIV-expressing cells were found in "clusters" in the cerebellum and appeared to be expressing high levels of viral transcripts (**Figure 6c-d**). In both tissues, HIV-expressing cells were surrounded by cells that co-stained for CD68/CD163, indicative of infiltrating activated macrophages. Furthermore, these macrophages appear to also express HIV RNA in the brain (**Figure 6e-f**). These results are consistent with the phylogenetic patterns, in which viral variants were dispersed among tissues and a rapid expansion of identical DNA and RNA sequences was evident in the cerebellum.

DISCUSSION

A striking result from this study was the *lack* of a substantial reduction in HIV evolutionary rates in anatomical tissues collected from fully-suppressed cART-treated patients, when compared to viral populations in cART-naïve patients. Several studies that have sampled T-cells or plasma after therapy interruption have found a genetically homogeneous HIV population that was similar, if not identical, to the pre-therapy virus (4, 27). One study, which

estimated the evolutionary rate of post-ART T-cells compared with pre-cART plasma virus using a similar method to that described here, found that the evolutionary rate was reduced orders of magnitude in post-cART virus, suggesting little to no virus replication/evolution in the T-cell population (26). Distinct subsets of T-cells are known to act as cellular reservoirs of HIV during cART (7, 9, 10, 16, 17, 19, 40), and the maintenance of non-evolving viral populations has been shown to occur through cellular expansion of memory cells with integrated but non-expressing provirus (35, 52). Therefore, an alternative site of viral maintenance during cART, apart from memory T-cells, must be present in order to explain the results presented here.

It is possible that lymphatic tissues, including lymph node and spleen, may be conducive to viral replication during cART. A recent study used the same analytical method as the one in this study to estimate evolutionary history and found evidence of on-going replication of wild-type virus in lymph nodes and continual seeding of the blood from lymph node sanctuary sites where drugs are not fully suppressive (33). Another group reported evidence for HIV RNA+ cells in lymph nodes prior to therapy interruption and a diverse population of rebound virus, suggesting that the rebounding viral population originated from multiple sources including lymph nodes (26). Lower concentration of drugs are also found in lymphatic tissues (18). Our finding of on-going replication of HIV RNA and DNA in lymph node and spleen is consistent with these observations. The lymph system may serve as an alternative route for HIV-infected cells to migrate (30) which would explain our observation of the dispersal of viral population among tissues (i.e. no compartmentalization), as well the lack of detectable virus in the blood. Furthermore, a recent study suggested that the meninges may be part of the lymphatic system (34), which could provide an explanation of late infection in the brain noted previously (31) and shown here in the recent and rapid expansion in the cerebellum. We note that our intention in

this study was not to determine routes of migration among tissues or the extent of compartmentalization among tissues, as the numbers of sequences among sites was unequal. In addition, clearly there were several types of evolution occurring among these sequences, which could result from infection of different cell types with varying dynamics.

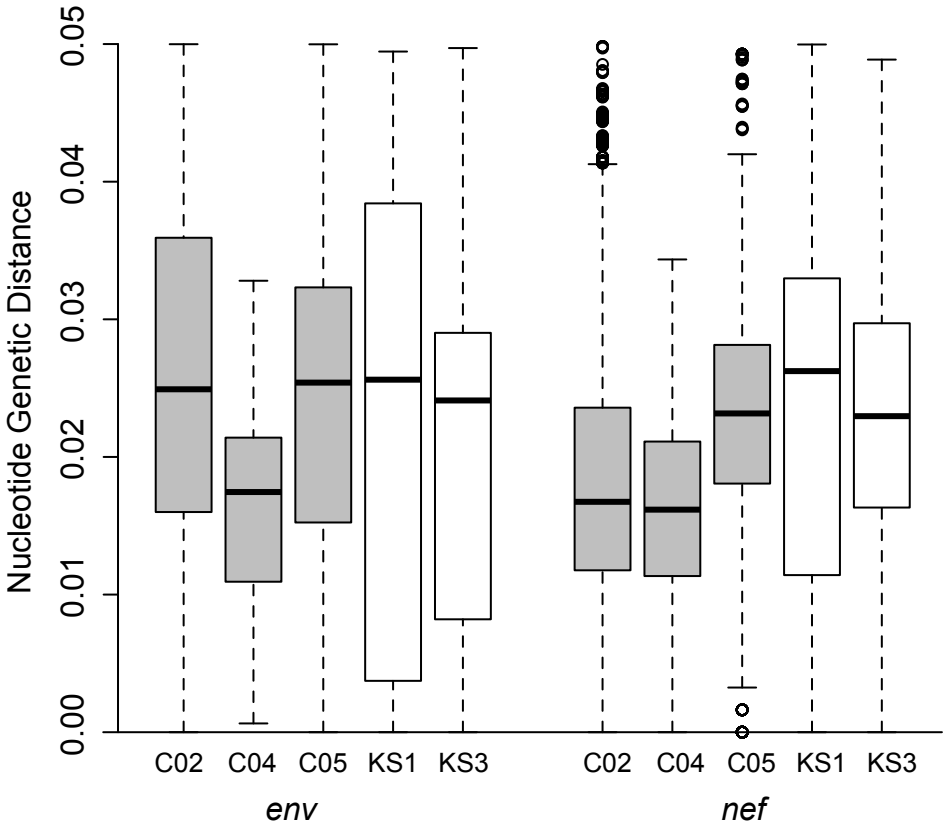
For example, in addition to T-cells, tissue-resident macrophages are long-lived cells capable of harboring virus (29, 46). Macrophages are less resistant to the cytopathic effects of the virus (22), contain lower intra-cellular concentrations of cART (20, 21, 44), and participate in efficient cell-to-cell transfer of virus among macrophages and from macrophages to T-cells (24), thus undermining the effectiveness of cART (14). We did not test the identity of the cells from which virus was extracted; however, we found evidence of activated infiltrating macrophages in very close proximity to HIV-expressing cells in two tissues, and potential HIV-expression by these macrophages in the brain. Infection of multiple cell types, i.e. T-cells and macrophages, may provide explanations for the two distinct branching patterns found in all patients in this study, consistent with on-going evolution in long-lived cells and clonal expansion.

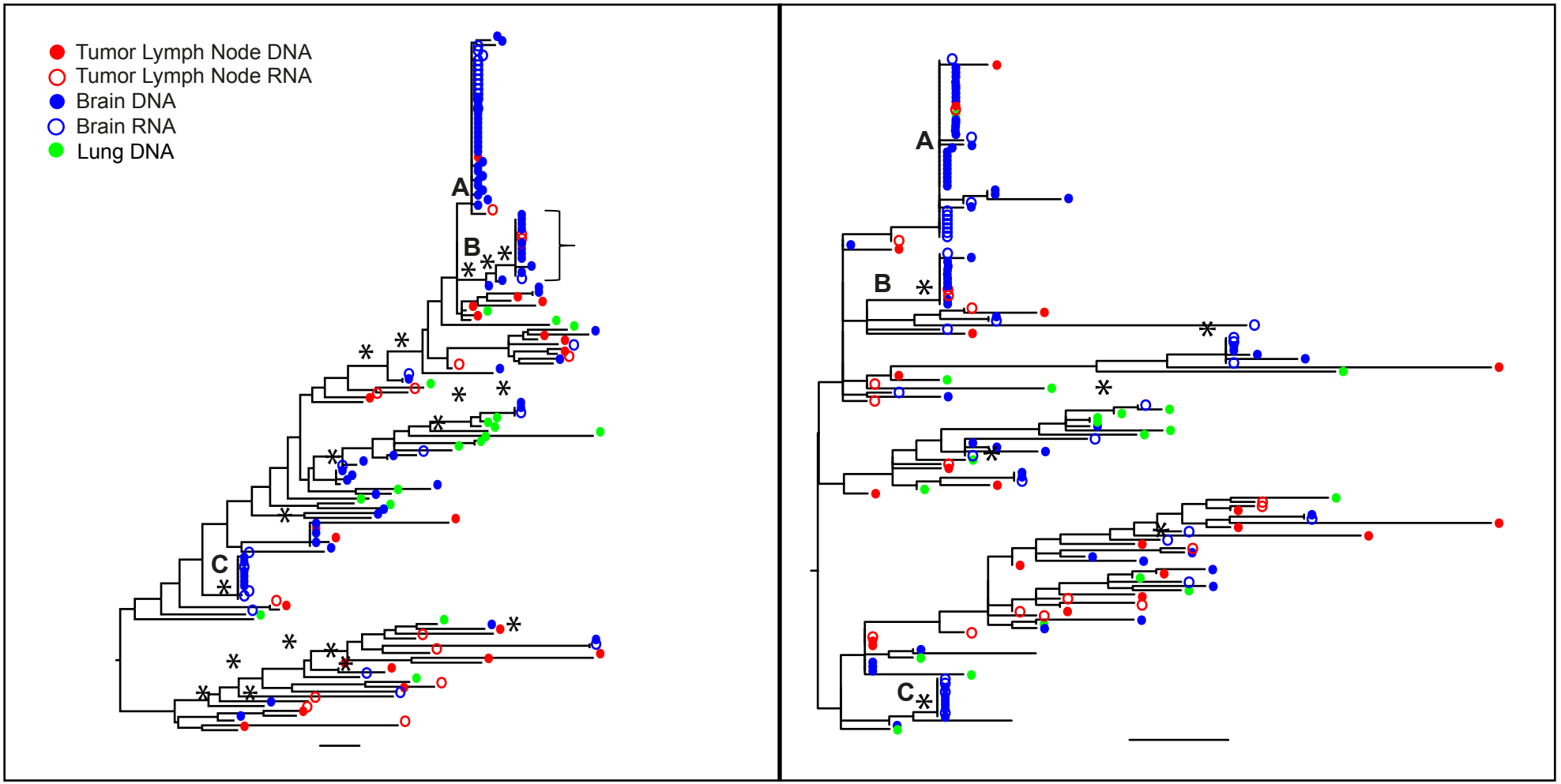
All of the participants in this study, as well as the two cART-naïve participants, were diagnosed with at least one type of cancer. In a previous study, we found that tumors harbored viral strains that were distinct from those in non-tumor tissues, and that macrophages were identified as a productively infected cell in the tumors (43). Tumor-associated macrophages may contribute to on-going HIV replication by orchestrating tumor progression, angiogenesis, tumor growth, and metastasis (2, 12, 42, 47) and migration of cells (37). The presence of activated infiltrating macrophages surrounding HIV-expressing cells in the two tissues studied here is consistent with recent migration of these cells, and potentially of the HIV-infected cells as well. If these cells were derived from the tumor as a result of metastasis, this again would explain the

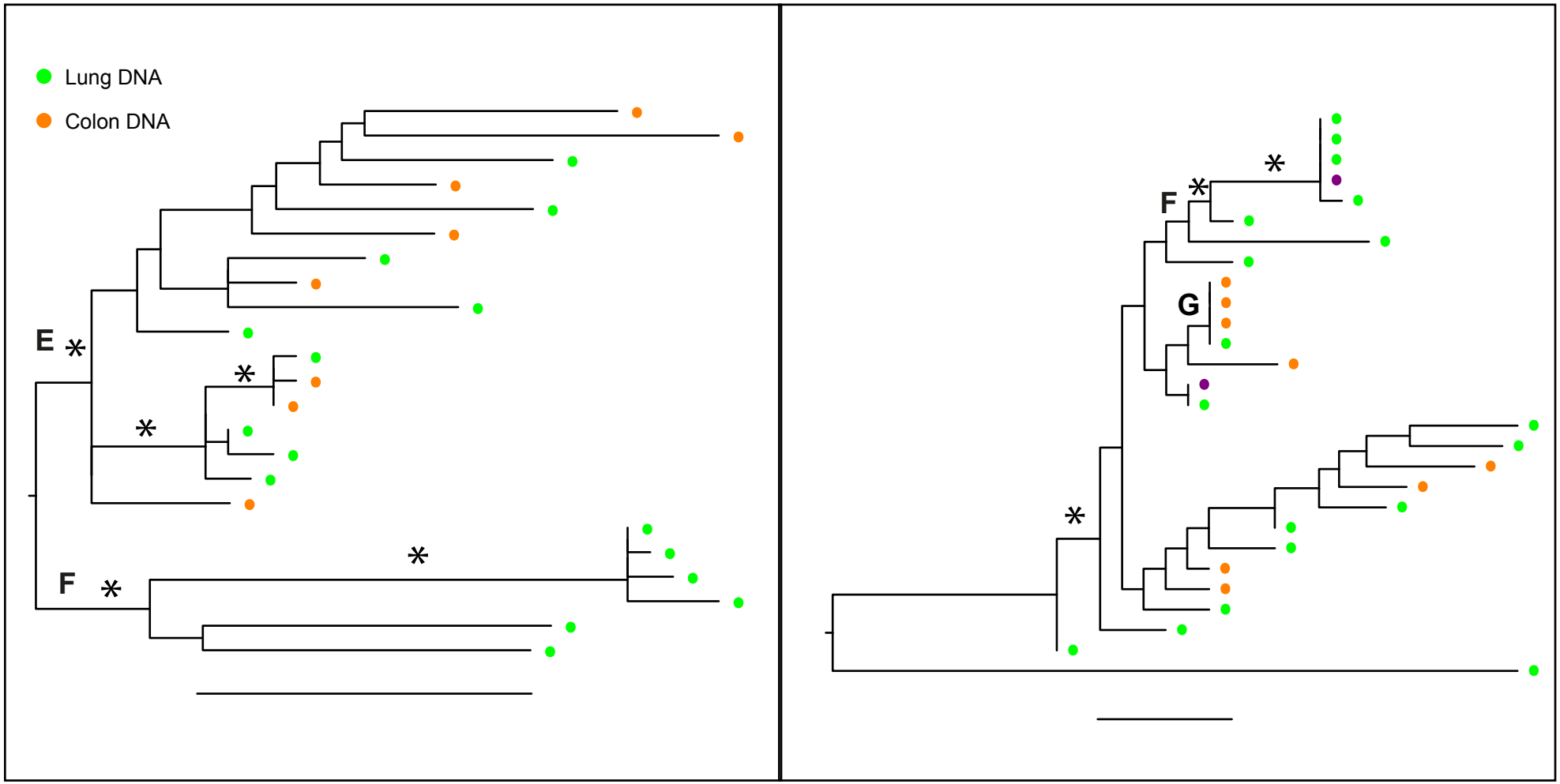
ability of the virus to replicate despite cART, dispersal of virus among tissues, and evidence of rapid and recent expansion of a replicating viral population. Furthermore, cancers can induce Type I IFN production, which may play a role in anti-tumor immunity (54), drive T cell dysfunction (8), and contribute to other aspects of HIV disease pathogenesis. For example, virus control (MX2 post-entry inhibition (23), APOBEC3 restriction of replication (28), BST-2 virus release (53), NK cell-activation (36)) or immune activation or dysfunction (CD38 expression on CD8+ T cells (5), CD4+ T cell/Fas-dependent apoptosis (1), decreased IL-7/increased activation-induced T cell proliferation (50)). These beneficial or deleterious effects may or may not relate to the evolution and dispersion of virus populations amongst tissues during the onset of cancer.

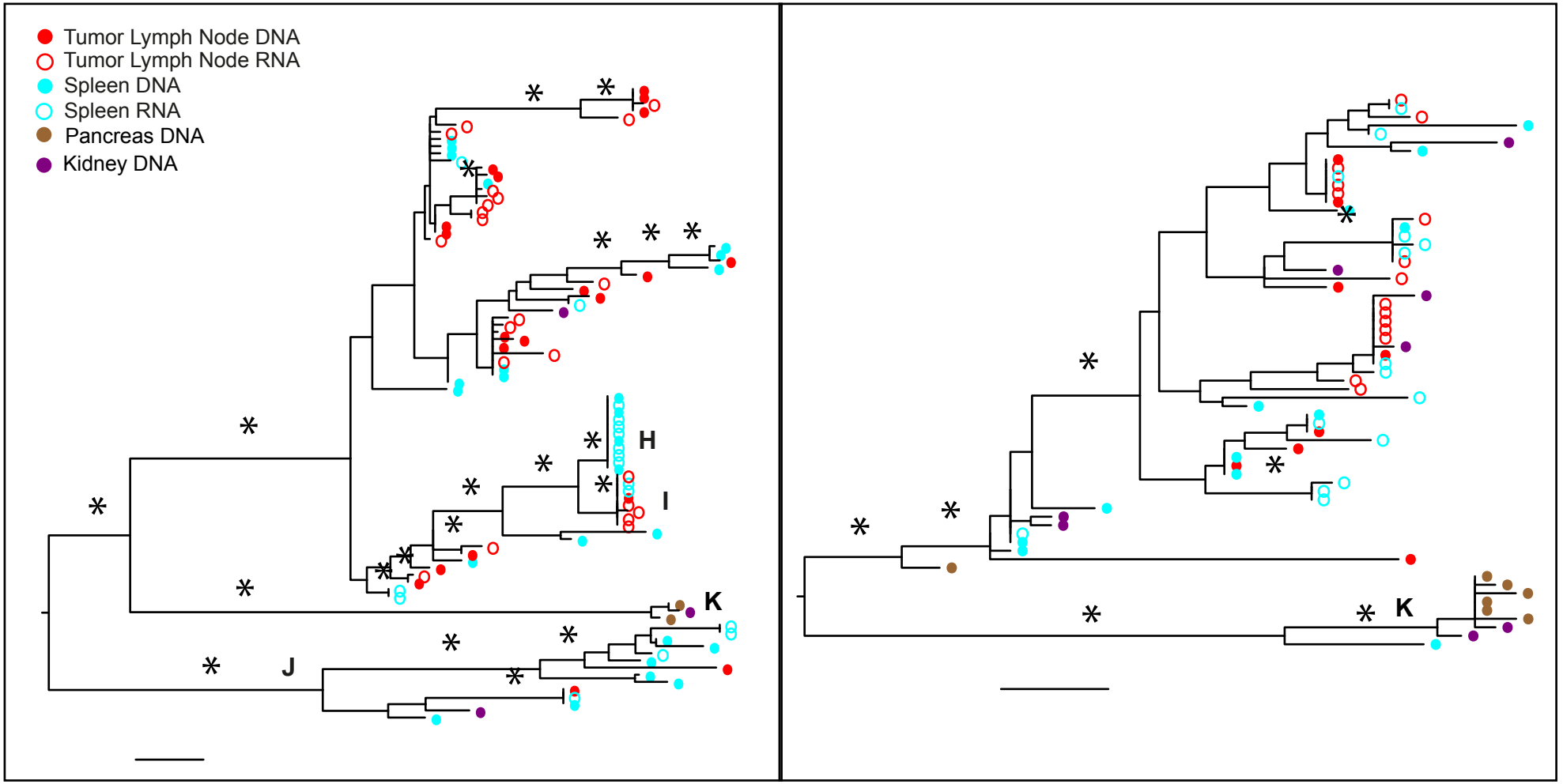
There are of course limitations to this study. In particular, the individuals selected do not represent a random or representative sample of all HIV infected persons, as the NNAB preferentially selects participants thought to have serious life-threatening illnesses. Furthermore, one selection criteria for inclusion was a short PMI, which could exclude healthier participants who expire at some distance from the study site, while preferentially including participants who expire due to poor health in a medical facility. Although the medical information reported in this study is complete to the best of our ability, some clinical variables are unkwown. The precise location of tissue sampling could clearly impact HIV recovery, for example, Lamers *et al* (J Virol, 2016, submitted) found high variability in the amount and detection of HIV in tissues could within an individual. We used a sequencing technique (SGA) and multiple extractions and amplifications in order to ensure that we sampled the broadest variability possible, but deep-sequencing studies could provide even more resolution. Additional studies included more patients are clearly required to completely understand the trajectory of viral evolution during cART.

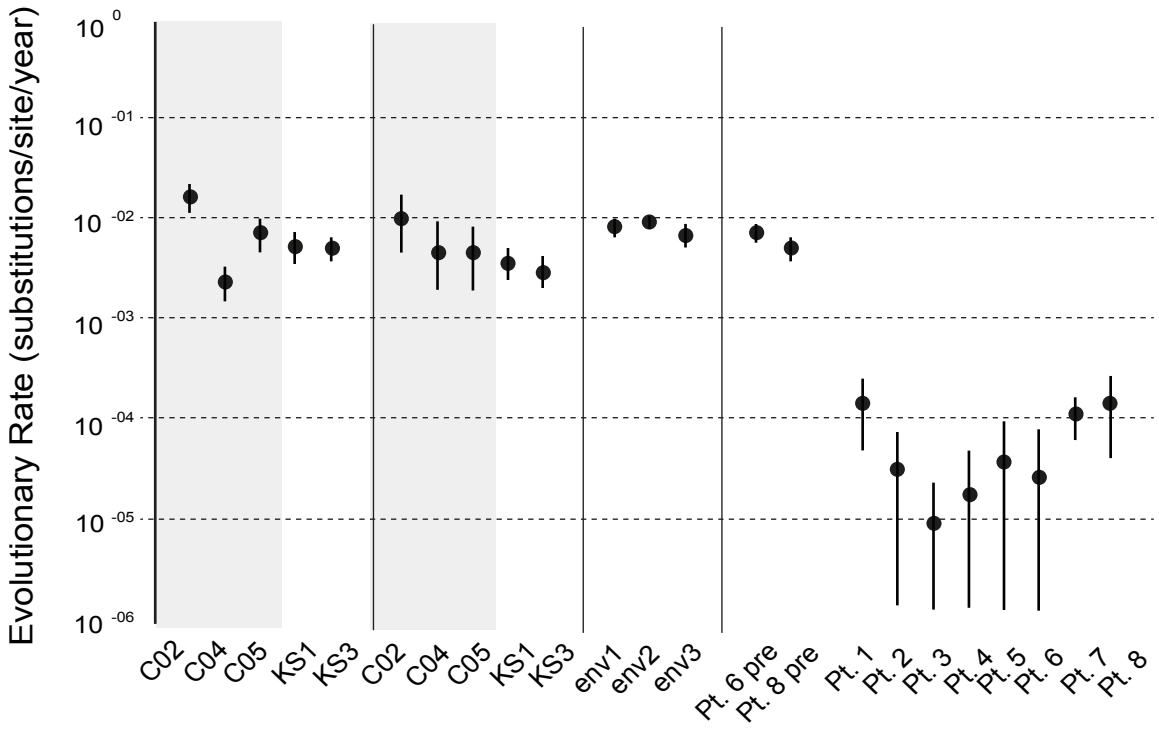
Acknowledgements. This study depended on the generous voluntary efforts of human participants and their caregivers to donate time and tissues. SLL, DJN, RR, MS, EJS and MSM are funded by the National Institute of Mental Health grant #NIH R01 MH100984. EM, CS and MSM are funded by National Cancer Institute grant #UM1 CA181255. EJS is funded by the National Institute of Mental Health grant # NIMH U24MH100929. SLL, DJN, MS are funded by the National Institute of Neurological Disorders and Stroke grant #NS063897.

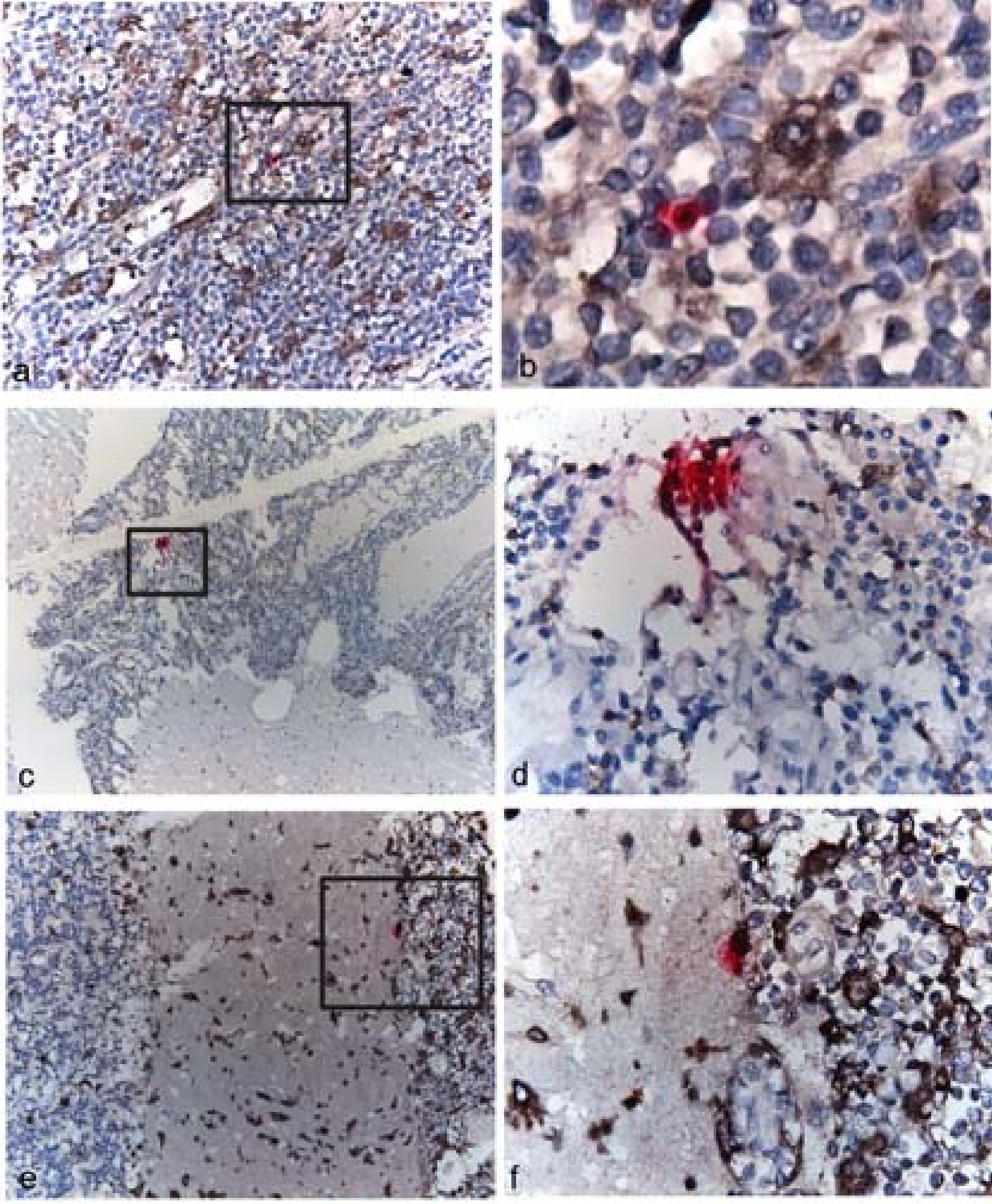

Nucleotide Genetic Distance
0.00
0.01
0.02
0.03
0.04
0.05
C02
C04
C05
KS1
KS3
C02
C04
C05
KS1
KS3
env
nef


Tumor Lymph Node DNA
Tumor Lymph Node RNA
Brain DNA
Brain RNA
Lung DNA
A
B
C
A
B
C


Lung DNA
Colon DNA
E
F
F
G


Tumor Lymph Node DNA
Tumor Lymph Node RNA
Spleen DNA
Spleen RNA
Pancreas DNA
Kidney DNA
H
I
J
K
K


Evolutionary Rate (substitutions/site/year)
10 0
10 -01
10 -02
10 -03
10 -04
10 -05
10 -06
C02
C04
C05
KS1
KS3
C02
C04
C05
KS1
KS3
env1
env2
env3
Pt. 6 pre
Pt. 8 pre
Pt. 1
Pt. 2
Pt. 3
Pt. 4
Pt. 5
Pt. 6
Pt. 7
Pt. 8


a
b
c
d
e
f