HIV persists throughout deep tissues with repopulation from multiple anatomical sources

Antoine Chaillon¹*, Sara Gianella¹*, Simon Dellicour²,³, Stephen A Rawlings¹, Timothy E. Schlub⁴, Michelli Faria De Oliveira¹, Caroline Ignacio¹, Magali Porrachia¹, Bram Vrancken³^ and Davey M Smith¹^

1. Department of Medicine, University of California San Diego, La Jolla, California, United States of America.
2. Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, CP160/12 50, av. FD Roosevelt, 1050 Bruxelles, Belgium
3. KU Leuven, Department of Microbiology and Immunology, Rega Institute, Laboratory of Computational and Evolutionary Virology, Leuven, Belgium;
4. University of Sydney, Faculty of Medicine and Health, Sydney School of Public Health, Sydney 2000, NSW, Australia

*contributed equally
^contribute equally

Keywords
Discrete phylogeography; generalized linear model (GLM); viral dynamics; within host; reservoirs.

Corresponding author and requests for reprints:

Antoine Chaillon, M.D. Ph.D.
Department of Medicine
University of California San Diego
9500 Gilman Drive MC 0679
La Jolla, CA 92093-0679, USA
Tel: (858) 552-8585 #2673
Fax: (858) 552-7445
E-mail: achaillon@ucsd.edu
ABSTRACT

**Background.** Understanding HIV dynamics across the human body is important for cure efforts. This goal has been hampered by technical difficulties and the challenge to obtain fresh tissues.

**Methods.** This observational study evaluated 6 persons with HIV (4 virally suppressed with antiretroviral therapy and 2 with rebound viremia after stopping therapy) who provided blood serially before death and their bodies for rapid autopsy. HIV reservoirs were characterized by digital droplet PCR and single genome amplification and sequencing of full-length (FL) envelope HIV. Phylogeographic methods reconstructed HIV spread and generalized linear models tested for viral factors associated with dispersal.

**Results.** Across participants, HIV DNA levels varied from ~0 to 659 copies/10⁶ cells (IQR:22.9-126.5). A total of 605 intact FL env sequences were recovered in antemortem blood cells and across 28 tissues (IQR:5-9). Sequence analysis showed: 1) emergence of large, identical, intact HIV RNA populations in blood after stopping therapy, which repopulated tissues throughout the body, 2) multiple sites acted as hubs for HIV dissemination but blood and lymphoid tissues were the main source, and 3) viral exchanges occurred within brain areas and across the blood brain barrier, and 4) migration was associated with low HIV divergence between sites and higher diversity at the recipient site.

**Conclusion.** HIV reservoirs persist in all deep tissues, and blood is the main source of dispersal. This may explain why eliminating HIV susceptibility in circulating T cells via bone marrow transplants allowed some people with HIV to have therapy free remission, even though deeper tissue reservoirs were not targeted.

**Trial registration.** Not applicable.

**Funding.** National Institute of Health Grants (P01 AI31385, P30 AI036214, AI131971-01, AI120009AI036214,HD094646, AI027763, AI134295, AI68636).
INTRODUCTION

Upon transmission to a new host, HIV-1 disseminates to the lymph nodes within days, and to the bloodstream and throughout the human body within 1-2 weeks(1-4). In this process, HIV establishes reservoirs throughout the body including: the central nervous system (CNS), lymphoid tissues (spleen, thymus, lymph nodes, gut associated lymphoid tissue), bone marrow, lungs, kidneys, liver, adipose tissue, gastrointestinal tract and genitourinary systems(5). The persistence of replication competent HIV in these anatomical sites, even in the setting of longstanding, potent antiretroviral therapy (ART) that suppresses HIV replication, is the main barrier to curing HIV(6).

Despite extensive investigations in humans (5, 7-18) much remains unclear about HIV reservoirs that persist during ART(19-23). In part, this is because of technical limitations and limited access to appropriately collected tissues for such studies. In general, most studies aiming to characterize HIV DNA reservoirs in persons with HIV (PWH) have necessarily focused on blood with a few other compartments that are relatively easy to sample, such as the gut via endoscopy (5, 10, 11, 24), genital tract through genital secretions(12, 25-30) or CNS through cerebrospinal fluid (31-37). Of note, one study in PWH diagnosed with cancer allowed the analysis of HIV reservoirs in selected anatomic tissues collected during autopsy (38, 39). More recently, De Scheerder et al. investigated the origins of HIV rebound, and although this study was limited by the number of anatomical sites, they showed that viral rebound originated from diverse cellular and tissue reservoirs (40). To address tissue availability, a peri-mortem observational research cohort, ‘The Last Gift’, was developed and enrolled PWH diagnosed with a terminal illness from a non-HIV condition. Participants consented to blood sampling before death and donated their bodies for a rapid autopsy after death(41-43). This allowed us to collect ante-mortem blood and post-mortem tissues across 28 anatomical compartments from six participants. Post-mortem tissues were collected within 6 hours of death to maximally preserve tissue and viral integrity(43). In the Last Gift cohort, some participants decided that they no longer wanted to continue their ART in the days and weeks before death. This provided the opportunity to observe and characterize rebounding viral populations in blood and compare them to viral populations found in tissues.

To characterize the HIV reservoirs in collected specimens, we combined new technologies that could deep sequence near full-length (FL) env HIV genomes and sensitively quantify HIV DNA(44, 45) with established molecular epidemiology inference methods to assess the viral diversity, divergence, predicted cellular tropism, replication competence, compartmentalization, and migration across the human body.

RESULTS

Cohort. The design of the proposed study is summarized in Figure 1. This study evaluated six male participants (see Supplementary Table S1 and Supplementary Figure S1). The median age at the time of death was 57.5 years (range: 52-72). All initiated ART during chronic infection and were virally suppressed for a mean 7.5 years before enrolling into the Last Gift study (range: 3-14 years). Two participants, LG01 and LG04, voluntarily interrupted ART 53 days and ~70 days prior to death respectively. The last HIV RNA blood
plasma level prior to death for LG01 and LG04 was 280 copies/ml and 48,000 copies/ml, respectively. LG03, LG05, LG06 and LG08 had undetectable HIV RNA in blood plasma (<20 copies/ml) until death and at the time of the autopsy. The median CD4 T cell count prior to death was 338/mm³ (range: 174 to 868/mm³).

**HIV DNA levels.** Overall, HIV DNA levels varied in sampled blood and tissues from ~0 to 658.8 copies/10⁶ cells (median=55.9, IQR:22.9-126.5) (Supplementary Table 2, Figure 2F and Figure 3). As expected, HIV DNA levels was lowest in CNS samples varying from a range of 0-4.6 copies/10⁶ cells in the occipital lobe to 1.8-33.6 copies/10⁶ cells in the frontal lobe. Although the sample size was small, there was no discernible difference in HIV DNA levels in tissue between the participants who did and did not interrupt ART (p>0.05).

**Defective proviruses (HIV DNA).** Across all samples from all participants, we recovered 676 single genome FL env sequences (Supplementary Table 2). Hypermutated and otherwise defective (i.e. non-intact) proviruses (FL env) accounted for 10.5% of all sequences (71/676), varying from 0 to 18.3% of all FL env sequences in a given tissue (Supplementary Table 2 and Figure 2B). Non-intact FL env sequences were excluded from further analyses, resulting in 605 evaluated intact FL env sequences (55 to 152 sequences per participant, median of 7 sequences /site, interquartile range: [IQR] 5-9). Overall, there was evidence that the proportion of genetically intact FL env variants varied between anatomical compartments (P = 0.016); however, the pattern of this variation was not consistent across participants (P = 0.96). To further investigate this, the 28 compartments were collapsed to 11 semi-related categories (blood plasma, cardiac serum, gut, CNS, kidney, liver, lymphoid, pancreas, PBMC, pericardial adipose, and genital tract), and still no pattern was detected. There was also weak evidence for a difference in the proportion intact FL env variants across participants (P = 0.04), but this association was weakened further after adjusting for anatomical compartment (P = 0.27) and anatomical compartment category (P = 0.051). The prevalence of intact FL env variants for each participant and anatomical compartment is presented in Figure 2 and Supplementary Table 2.

**Viral diversity and clonality.** Viral diversity of HIV RNA populations in blood and HIV DNA populations in tissues varied between tissues across participants and across tissues within participants (Figure 2E and Supplementary Table and Figure S2). In participants who decided to stop therapy, most of the rebounding intact HIV RNA in blood (65% and 80% of variants) was >99% identical, with the mean pairwise nucleotide Tamura Nei 93 distance (46) of 0.015 for LG01, and 0.001 for LG04. Overall, viral diversity did differ significantly across participants (p=0.057) but not across tissues (p=0.51); however, our power to detect differences across tissues was poor due to the large number of compartments and small number of participants. Considering this, we also looked at individual tissue comparisons to blood plasma HIV RNA populations and found that viral diversity was overall higher for all HIV DNA in tissues but the higher viral diversity was only significant for prostate (p=0.046), rectum (p=0.045), right colon (p=0.013), spleen (p=0.013), and occipital cortex (p=0.006).

We also found evidence of identical intact FL env proviruses within and across tissues. For example, monoclonal FL env populations (100% identical) were found in ileum (3 out of 3 intact FL env sequences sampled) in LG03, jejunum (7 out of 7) in LG04 and testes (6 out of 6) in LG06. Identical FL env sequences
were also found across compartments, with LG03 having 30 identical intact envelopes recovered from 8 anatomical compartments and LG04 having 14 identical intact envelopes from 5 tissue compartments. We also found a large nearly identical population (≥99% identical) of 25 FL env proviruses sampled across 10 compartments in LG08 (Figure 4 and Supplementary Figure S4). The proportion of identical FL env sequences varied significantly across participants (p<0.01) with the proportion of identical FL env variants throughout the body being: LG01 21/75 28%, LG03 (54/113, 48%), LG04 (28%, 42/152), LG05 (14/55, 25%) , LG06 (19/104 18%) and LG08 (25/106 24%). While the proportion of identical FL env sequences differed significantly across compartments, effect modification was also detected between participant and compartment (P < 0.001). In other words, the differences across compartments varied according to participant with no consistent trends of clonality across tissues observed with or without sustained viral suppression.

**Viral tropism and re-population.** HIV DNA populations in ante-mortem PBMC and post-mortem tissues were predominantly X4-tropic based on genotypic analysis for four participants (LG01 84%, LG03 75.5%, LG05 78.2% and LG06 54.5%), while X4-tropic proviruses comprised only 14.9% and 6.6% sequences obtained from LG04 and LG08 (Supplementary Table 2 and Supplementary Figure S5). Next, we evaluated tissue re-population during rebounding viremia in the 2 participants who voluntarily decided to interrupt their therapy:

Participant LG04 who stopped his ART more than 2 months prior to his death and his last HIV RNA level prior to death in blood plasma was 48,000 copies/ml. The rebounding HIV RNA was 100% intact in env and nearly clonal (10 out of 10 intact nearly identical [≥99% identity] FL env proviruses) and 80% of variants were R5 tropic (Supplementary Figures S4 and S5). Interestingly, identical R5 tropic FL env sequences were also found in the HIV DNA population in ante-mortem PBMC (17/23 sequences), and across almost all other sampled tissues collected post-mortem, including lymph nodes, genital tract and gut tissues (Supplementary Figure S4). While we cannot be sure of the origin of the rebounding R5-tropic population as ante-mortem samples were only available for PBMC and blood plasma, this observation strongly suggests that many tissue reservoirs were re-populated by R5-tropic virus during rebound.

Participant LG01 stopped his ART 53 days prior to his death and his last HIV RNA level prior to death in blood plasma was 280 copies/ml. Again, the rebounding HIV RNA population was 100% intact (17/17) and predominantly clonal (13 out of 17 intact nearly identical [≥99% identity] FL env). Phylogenetic analyses also revealed a large nearly identical FL env population of 20 X4 tropic proviruses in post-mortem gut (n=5), PBMC (n=4) prostate (n=3) and blood plasma (n=8) (Figure 4 and Supplementary Figure S5).

**Viral dispersal and re-population.** As expected, all sequences clustered by participant when combined in a single phylogeny (Supplementary Figure S3), and using phylogeny-trait (i.e. anatomical site) association measures, we found evidence of viral compartmentalization by anatomical site for all six participants (P<0.01) (Supplementary Table S3). Discrete phylogeographic models showed various patterns of viral dispersal (Figures 4 and 5, only transition events between locations for which the adjusted BF ≥ 3 are shown).

Among participants who remained virally suppressed (LG03, LG05, LG06 and LG08), the majority of transition events originated from lymph nodes and gut tissues (Figures 5-6 and Supplementary Figure S6).
There was also strong evidence for viral dispersal from lymph nodes to prostate in LG03 (BF = 568) and from rectum to seminal vesicles in LG08 (adjusted BF = 74). Interestingly, there was also positive support for viral transitions from PBMC to frontal lobes in LG05 (adjusted BF = 11) and LG08 (adjusted BF = 5.5).

Among the participants who stopped ART (LG01 and LG04), we sequenced FL env from HIV RNA from blood plasma collected ante-mortem during viral rebound (HIV RNA levels of 13,500 copies/mL and 48,000 copies/mL respectively). For LG01, there was strong evidence of viral dispersal from PBMC of all transition events (BFs≥20) toward gut sites, prostate, spleen, and blood plasma (adjusted BFs of 449.7, 62, 27 and 22.7 respectively). There was also good support (BFs≥10) for transition events from blood plasma toward lymph nodes (adjusted BF = 16.6) and liver (adjusted BF = 10.3) and from liver toward lymph nodes (adjusted BF = 11.8). We also showed positive support for viral migration from frontal lobe toward PBMC (adjusted BF = 3.9) and from lymph nodes toward frontal lobe (adjusted BF = 10.4) (Figures 5-6). For LG04, 9 sources of viral migration were strongly-supported (BFs≥20), including migration from blood plasma toward gut tissues and intense viral exchanges between the sampled gut sites (Figure 5 and Supplementary Figure S6). Interestingly for LG04, we also found evidence of bidirectional transition events within the CNS, from the occipital lobe toward frontal lobe (adjusted BF = 25.4) and from the frontal lobe toward the occipital lobe (adjusted BF = 10.1). The lack of dominance of particular transition types shows that a diverse set of anatomical compartments can act as the source of reservoir virus.

Predictors of viral dispersal. We next used a GLM extension of the discrete phylogeographic model to investigate if the number of proviruses, viral diversity, proportion of X4-tropic viruses, and viral divergence between compartments were associated with potential dispersal rates. We first assessed collinearity between all the factors evaluated, and all correlation coefficients were <0.8. Consistently for LG03, LG04 and LG06, the model revealed strong evidence that viral migration was associated with limited divergence between the source and recipient compartment (respective adjusted BFs of 389.7, 54.4 and 1690.4, Supplementary Figure S7). For LG03 and LG04, there was also evidence of association between the transition rate and higher viral diversity in the recipient compartment (BFs of 373.9 and ≥31.5.10^4 respectively). Finally, for LG03 only, there was a positive association between lower proportion of X4 tropic viruses in the source compartment and viral dispersal (BF=5.8). For LG01, LG05 and LG08, there was no evidence that any of the tested factors associated with viral dispersal (Supplementary Figure S7).

DISCUSSION

Understanding how HIV populates tissues throughout the human body is crucial for the development of strategies to clear or otherwise lock-down these reservoirs(47-49). Such investigations have been hampered by limitations in technology and access to fresh sampling of tissues throughout the body. This study used samples and data from six participants in a novel, peri-mortem cohort (The Last Gift) of altruistic PWH who provided blood samples before they died and their whole bodies immediately after death. Given the freshness of samples, viral populations likely remained intact, so virologic data were generated and analyzed using state-of-the-art methods. The main findings of this study were: 1) The emergence of large, clonal, intact, HIV RNA populations in blood plasma after stopping ART, and these populations repopulated tissues throughout
the body. 2) Multiple sites can act as hubs for dissemination of HIV within the host, predominately blood, gut and lymph nodes. 3) Viral exchanges occur within CNS areas and between the CNS and blood. 4) Viral dynamics are associated with low HIV divergence between sites and high HIV diversity at the recipient site. Together, these data provide new insights in how HIV populates reservoirs across the human body.

Previous studies showed that lymphoid tissues (e.g. spleen, thymus, lymph nodes, gut-associated lymphoid tissue) are the sites with highest level of viral replication during active infection, and HIV DNA can be readily cultured from lymphoid tissues after years of treatment (5). HIV-infected cells have also been found in the CNS, the lungs, kidneys, liver, adipose tissue, genitourinary tract, and bone marrow (5, 50-53). In agreement with these observations, we found intact FL HIV-1 env throughout the 28 anatomic compartments sampled from 6 Last Gift participants. It was interesting that participant LG03 had very high levels of HIV DNA in his pancreas (391.5 gag copies /10^6 cells [329.2-453.7]), which is also where he had his terminal cancer. Since T cells infiltrate tumor microenvironments, it is possible that the resident tumor attracted HIV-infected T cells (54). More tumor tissues will need to be evaluated to confirm this observation.

Consistent with other reports that investigated blood from PWH (55-57), our study found various proportion of identical FL env sequences in PBMC (varying from 19% [CI95%:4-46%] to 81%[CI95%:54-96%] for LG08 and LG01 respectively). We also found identical FL env sequences in various tissues (e.g. ileum and jejunum for LG03 and LG04) and across multiple tissues. This observation could be from either monoclonal HIV RNA infecting susceptible cells at multiple sites, as seen in LG04 during rebound viremia, or infected cells that clonally expanded at one site then migrated into a different site. These scenarios are not mutually exclusive. While identical env sequences may reflect cellular proliferation, without integration site analysis, we cannot determine if identical env sequences are derived from cellular clonal expansion (58). Future integration site analyses are needed to confirm this possibility, as illustrated by Patro et al (59).

Concerning viral migration, this study found that not only blood but also mainly gut and lymphoid tissues can act as sources for the dissemination of HIV, which is consistent with previous studies suggesting that both are major HIV reservoirs during ART (60-65). Interestingly, we also saw that genital tract tissues (i.e. prostate, seminal vesicles, testis) were potential sources of viral dispersal to other tissues. Although limited by the number of intact FL env proviruses recovered from the frontal and occipital lobes (5 out of 6 participants), this study shows that viral migration occurs within the CNS (LG03 and LG04), between blood and the CNS in 4 participants (LG01, LG03, LG05, LG08) and from the lymph nodes toward the CNS (LG01).

Assessing the factors that may contribute to the viral dynamics within and across compartments is likely to assist in developing strategies targeting the reservoirs; therefore, we used a GLM model to evaluate the role of viral characteristics (e.g. viral diversity, proportion of X4 tropic viruses and viral divergence between compartments) as potential drivers of viral dispersal. Although these findings were limited to 6 participants and only on viral population characteristics, we showed that limited viral divergence between compartment was strongly associated with viral dispersal in 3 out of 6 participants. We also showed that viral dispersal was associated with higher viral diversity in the recipient (2/6 participants) and lower proportion of X4 tropic viruses.
in the source (1/6 participants). While these results remain limited and require validation in a larger cohort, it suggests that HIV dynamics across the body is also influenced by molecular characteristics of the HIV populations. Further analyses evaluating pharmacological and immunological factors in the HIV micro-environment would be of interest.

Some PWH at the end of their lives want to stop their ART before they die. This unique situation allowed us to observe viral rebound and tissue re-population in two participants who voluntarily interrupted their ART 53 and 70 days prior to death with rebound viremia. Consistent with De Scheerder et al (40), our phylogenetic and statistical analyses show that rebounding viral populations can originate from various anatomical compartments with no prominent source of rebound. Also consistent with previous studies (66), we showed that rebounding HIV RNA populations were comprised of intact nearly identical FL env sequences, and thus we found that HIV RNA populations were more homogenous than HIV DNA populations in tissues in the two participants who stopped therapy (P = 0.025). Interestingly, when LG04 stopped his ART, he had an asymptomatic rebound of a large, intact, nearly identical R5-tropic HIV RNA population (FL env) in blood. This population was identical to a HIV DNA population seen in PBMC ante-mortem. After death, we found identical FL env sequences in tissues throughout the body including lymph nodes, genital tract and gut tissues. Similarly, LG01 had rebound viremia with an intact nearly identical X4-tropic population, and we found identical FL env sequences in gut, PBMC and prostate. Although not conclusive because we did not sample tissues before stopping ART, this strongly suggest that the rebounding HIV population in blood plasma is the source of viral replenishment in tissue reservoirs. It also supports the notion that preventing HIV rebound in the blood may allow for ART-free remission, like what was observed in Timothy Brown and the London patient, when their previously HIV-susceptible circulating CD4 T cells were replaced by bone marrow transplant with CD4 T cells that were no longer susceptible to HIV infection because of the CCR5 32Δdeletion (67, 68).

Our study has a number of limitations. The main limitation is the small number of participants, especially having only two participants who stopped their ART. Nonetheless, it allowed us to unprecedentedly observe the population of tissue reservoirs from the blood during rebound viremia. Another limitation is that the participants were all in the process of dying, which may limit the generalizability to healthy PWH. Also, this study, as others(69, 70), focused on the HIV env gene, which has the greatest amount of molecular diversity and evolution of all coding regions(71-76), but we acknowledge that by sequencing only the env region, we may have incorrectly inferred that some viruses intact in the env coding region were replication competent, when they may have had defects in other genome regions. Further, the study found many identical HIV DNA env single genomes, consistent with previous reports,(77-80) but we did not directly assess cellular clonal expansion, which is thought to be an important mechanism for HIV persistence(81-83). However, others have shown that multiple identical env sequences in proviruses provide a strong indication for clonal expansion(84, 85). Sequencing near FL proviral genome (86-88) would increase the sensitivity of the analyses of intactness and clonality but these approaches require a high cellular input which may limit the ability to explore reservoirs such as the central nervous system. Therefore, we are confident that our approach provided a good surrogate for the extensive analysis of HIV reservoirs. Further, our study did not perform phenotypic testing of CCR5.
tropism, and every participant in our study had X4-tropic virus by genotypic analysis. This is likely because participants were infected for a long time before cohort enrollment (89), and emphasizes that people with longstanding HIV infection are unlikely to benefit from measures focusing only on CCR5 tropic viruses, as was the case for the two PWH who received CCR5 mutated bone marrow transplants and have been cured of HIV so far (68, 90). Finally, we cannot exclude the possibility of blood T cell contamination in tissues obtained during autopsy. This is likely to be a small impact on our analysis given the small size of capillaries compared to overall tissue mass and settling of blood in tissues, i.e. livor mortis. While we cannot completely rule out such contamination, our sequence analyses showed viral compartmentalization for all participants, which suggested that possible blood contamination did not significantly impact our analyses.

In conclusion, this study leveraged the unique Last Gift cohort of participants who continued and discontinued ART before dying, after which multiple tissues were rapidly collected at autopsy. From these samples, the study reconstructed the spatial dynamics of HIV across the body and found that multiple body compartments, especially the gut and lymphoid tissues, can act as hubs for dissemination of HIV, and that blood is the likely conduit for dispersal. This study also uniquely clarified the HIV dynamics within the CNS and across the blood brain barrier. Importantly for cure efforts, our analyses strongly indicated that rebounding virus was derived from circulating reservoirs, and that rebounding virus quickly populated deep tissue reservoirs throughout the body via the blood. Given the extent of intact FL env HIV populations throughout the body, it is likely that any of the tissue reservoirs could facilitate re-seeding of virus to the blood and thus repopulate tissue reservoirs. These findings might have consequences for treatment interruption studies although clinical significance is unclear (91). Thus, for cure efforts aimed at eradication, then all such reservoirs would need to be cleared. Alternatively, functional cure efforts could focus on making circulating CD4 T cells impervious to infection to prevent rebound viremia and subsequent viral dispersal and reseeding, although this approach would not address local viral reactivation and damage.

METHODS

Study cohort and sampling. Peripheral mononuclear cells (PBMC) and tissue samples were obtained from the first 6 PWH enrolled in ‘The Last Gift cohort. (92, 93) Study participants were followed closely near the end of their lives (approximatively every other week with closer intervals as death approached). During these visits, participants provided: (i) Detailed clinical and socio-demographic information before their death (use of ART, chemotherapy and other therapies, surgical procedures, co-infections, etc.), (ii) blood collections while they were alive, and (iii) their entire bodies after they died for a rapid autopsy. Clinical characteristics of study participants are summarized in Supplementary Table S1 and Supplementary Figure S2.

Rapid Autopsy. The Last Gift rapid autopsy protocol was designed to collect tissues within 6 hours of death so that post-mortem tissue degradation is minimized. At the time of demise, the body is rapidly transported to the UCSD morgue where the team performs a complete autopsy to sample all organs for formalin-fixed
paraffin embedded tissue histological analysis and snap-frozen tissue collected in liquid nitrogen. Fluids collected include CSF and blood drawn from cardiac puncture.

**HIV DNA Quantification and Sequencing.**

*DNA extraction, purification and quantification.* Genomic DNA was extracted from 5 million PBMCs and snap-frozen tissues using QIAamp DNA Mini Kit (Qiagen cat#51306) per manufacturer's protocol. After extraction, precipitation was performed to concentrate DNA. Concentrations of DNA were determined using NanoDrop One (ThermoScientific). Levels of HIV DNA were quantified by droplet digital PCR (ddPCR) from extracted DNA using the BIO-RAD QX200 Droplet Reader. Copy numbers were calculated as the mean of 3 replicate PCR measurements and normalized to one million cells, as determined by RPP30 (total cell count). (94, 95)

*RNA Extraction and Quantification.* RNA was extracted from blood plasma layering 500-700µl of plasma on top of 200µl of 20% sterile filtered sucrose solution. Sample was spun at 23,500xg for 1 hour at 4°C to pellet the virus. Supernatant was removed and the pellet resuspended in 140µl of PBS. RNA was extracted using Qiagen’s QIAamp Viral RNA mini kit (cat# 52904) according to the manufacturer’s recommendation. cDNA from HIV RNA was generated using Bio-Rad One-Step RT-ddPCR Advanced Kit for Probes (cat# 186-4021), and levels were quantified by ddPCR (BIO-RAD QX200 Droplet Reader). Copy numbers were calculated as the mean of 3 replicates.

*Nested PCR.* To amplify single genome FL env, DNA extracted from antemortem PBMCs and post-mortem tissues was diluted using ddPCR quantification data. This limited dilution PCR reaction can prevent PCR recombination, ambiguous base calls and allow the amplification of viral single genomes. (87, 96) For HIV RNA in blood plasma, cDNA was generated from RNA using SuperScript III First Strand Synthesis System (cat# 18080-051). Template cDNA and HIV DNA extracted from tissues were diluted until approximately 30% of the second-round reactions were positive for the correctly-sized amplification product. Primers used for the first round were 5’FENVouter (forward) TTAGGCATCTCCTATGGCAGGAA and 3’RENVouter (reverse) TCTTAAAGGTACCTGAGGTCTGACTGG. First round PCRs were performed using the Advantage 2 PCR Kit from Takara (cat# 639206) following manufacturer’s recommendations using the 10X SA Buffer. Cycling conditions were 95°C for 1 min, 35 cycles of 95°C 15sec, 57°C 30sec, 68°C 3min with a final extension at 68°C for 10min. The second round PCR were done using 5’FENVinner: GAGCAGAAGACAGTGGCAATGA (forward) and 3’RENVinner: CCACTTGCCACCCATBTTATAGCA (reverse). Cycling conditions were 95°C for 1 min, 30 cycles of 95°C 15sec, 64°C 30sec, 68°C 3min with a final extension at 68°C for 10min. PCR clean ups were done on the second round reaction products using QIAquick PCR Purification Kit (cat# 28106). DNA was quantified using Qubit dsDNA HS Assay Kit (Invitrogen cat#Q32854). Quality and integrity were measured using Genomic DNA Screen Tape in combination with the 2200 TapeStation System (Genomic DNA Reagents cat#5067-5366, Genomic DNA Screen Tape Cat #5067-5365).
Nextera XT Library Preparation. Single Genome amplicons were prepared for deep sequencing using the Nextera XT DNA Library Preparation Kit (Illumina FC-131-1096) with indexing of 96-samples per run (Nextera XT index kit set A FC-131-2001) per manufacturer’s protocols.

Assembly of full-length HIV env proviruses. We used a custom designed pipeline to recover FL env HIV sequences from the paired end reads. The pipeline included a preliminary step of quality control, which included trimming reads for PHRED quality above or equal 30 and removal of Illumina adapters. Next, overlapping identical paired forward and reverse reads were merged and pre-mapped to HXB2 reference genome. Cleaned reads were re-mapped to the de novo assembled near FL env sequence before generating the final consensus sequence. The minimum acceptable coverage was set to 10,000 reads. To identify mixtures (i.e. suggesting multiple amplified HIV templates), we screened all generated FL env contigs. Mixtures were identified based on read coverage and variant calling. Contigs with evidence of single nucleotide polymorphisms (SNPs) with a frequency >1% were considered mixtures and excluded from further analyses.

Test for cross-contamination A maximum likelihood (ML) phylogeny including all sequences from the four participants was estimated using using IQtree with the general time reversible (GTR) substitution model (97) to test for contamination, which would show as intermixed clustering of taxa between participants.

Identification of defective or hypermutant sequences. FL envelope containing large deletions (>100bp) were considered defective(86, 98). Deleterious stop codons were identified using the Gene Cutter tool (Los Alamos HIV Database). Any contigs containing a stop codon were considered defective. APOBEC-induced G-A hypermutations were identified using the Los Alamos HIV Database Hypermut 2 program and the participant’s consensus sequence(99, 100). Proviruses with a number of mutations significantly higher than the participant’s consensus (Fisher exact test p < 0.05) were considered hyper-mutant and were not included in the downstream analyses described below.

Sequence analyses.

Identification of Identical full length env sequences. For determining the sequences which were ≥99% or 100% genetically identical, we used the ElimDupes tool from Los Alamos HIV database using ≥99% and 100% genetic identity threshold as the analysis parameter(101). A sequence was classified as identical if it was a 100% match against another sequences sampled from the same participant. Once identified, the proportion of identical and nearly identical sequences were calculated by dividing the total number of sequences classified as identical/nearly identical for each participant or compartment by the total number of sequences for that group.

Diversity and Divergence. Viral diversity was defined as the average pairwise genetic distance between sequences from a compartment using the Tamura Nei 93 (TN93) correction for multiple hits(46). Viral
divergence was assessed by computing the mean pairwise distance (TN93) between viral populations sampled across anatomical sites.

**Co-receptor tropism.** Viral tropism of each variant was inferred from the V3 amino acid sequence using geno2pheno(102) We applied a conservative 10% false positive rate threshold for co-receptor CXCR4 usage based on recommendation from the European Consensus Group on clinical management of HIV-1 tropism testing.

**Viral migration.** Prior to evaluating the within-host migration processes, the level of spatial structure was quantified using the Simmonds association index (AI) implemented in BaTS v1.0(103). We used the BEAST software package v1.10(104) for all evolutionary analysis. All sequences were considered isochronous, i.e. branch lengths were estimated in units of substitutions per site. For this, a strict molecular clock was specified and the clock rate was fixed to 1. The substitution process was described with a HKY+Γ model(105, 106) and a constant population size was assumed. Discrete trait analyses were performed using the asymmetric diffusion model(107, 108). To identify the subset of migration rates that were most informative to reconstruct the dispersal history we used a model averaging procedure (Bayesian stochastic search variable selection, BSSVS)(107). Bayes factor (BF) support for all possible types of location exchanges were calculated with SpreaD3(109). BF between 3-20, 20-150, and above 150 were considered positive, strong, and decisive support respectively(110). Estimates of the posterior expected number of migration events between all pairs of locations (Markov jumps) were computed through stochastic mapping techniques(105, 111). To identify individual-level differences in within-host viral diffusion patterns, the trait analyses were performed on a patient-specific basis.

To investigate whether support for migration links followed from the relative abundance of the involved trait states, analyses were repeated whilst randomly permuting the compartment states between tips during the MCMC sampling(112), a technique analogous to the tip-date randomization procedure for testing the significance of the temporal signal(113, 114). Indeed, if the support for a particular migration rate persisted after randomizing the tip-to-location assignments, one cannot rule out that the support is due to sampling intensity differences. Furthermore, of the migration links that pass the above filter (good BF support in the ‘as is’ analysis and poor BF support in the ‘tip-state-swap’ analysis), only those that remain significant after accounting for the sampling heterogeneity were taken into account to further reduce the false positive rate. To this end, the inclusion frequencies from the tip-state-swap analysis instead of those of the ‘as is’ analysis were treated as the prior inclusion probabilities when recomputing the Bayes factor support, thereafter referred to as “adjusted BF”. Only the results with adjusted BF with positive support (BFs≥3) were further discussed in this study. This approach was also used to investigate the sensitivity of the Simmonds AI to the sampling heterogeneity.

Multiple MCMC chains were run sufficiently long to ensure convergence and adequate mixing (Effective Sample Size [ESS]>200), which was inspected using Tracer 1.7(115). The chains were sampled every 500,000 generations and combined after removal of the burn-in. Maximum clade credibility (MCC) trees were obtained with TreeAnnotator 1.10(104).
Identifying predictors of within-host spread. The generalized linear model (GLM) extension of the discrete trait model implemented in BEAST(116) was used to investigate the relevance of potential explanatory variables (predictors) to explaining the dispersal across body compartments. The following variables were included: (1) the number of FL intact env provirus recovered in each compartment to control for sampling bias effects(116), (2) the level of HIV DNA (ddPCR gag copy/10^6 cells for all antemortem PBMC and post-mortem tissue samples) or level of HIV RNA for antemortem blood plasma samples (the number of copy/µL), (3) viral diversity estimated using average pairwise genetic distance between each sequences, and (4) the proportion of X4 tropic provirus in each compartment. For each of these variables, the values in the compartment source and recipient were considered. We also included a matrix of pairwise measures of TN93 distance between compartments (i.e. viral divergence). Collinearity between the variables included in the GLM model was evaluated. When 2 variables showed a collinearity coefficient ≥0.8, the analysis was repeated with only one variable in the model.

Statistical analyses. Multivariable logistic regression carried out in R version 3.6.1 using function glm and binomial link function, was used to compare the proportion of sequences that were intact and identical (clonal). The independent variables in these analyses were participant and anatomical location. For the continuous average pairwise distance (diversity) outcome, a multiple linear regression was used with assumptions of constant variance and normality of residuals checked and met. Mixed models for both binary and continuous outcomes were analyzed with glmer and lmer from the lme4 R library. These mixed models used participant as the grouping factor and included a random term for either intercept alone, or intercept and compartment. Due to the sparse nature of the data, all mixed models had difficulty converging with poor model fits. As such the results and P-values of these models are not reliable and so the results were not presented. Given the large number of tissue compartments, the sparse sampling across participants, and the expected effect modification by participant and compartment, comparisons across specific compartments is not informative.

Study approval. All study participants were at least 18 years of age and provided written, informed consent. Our study (protocol number 160563) was approved by The University of California San Diego (UCSD) Office of Human Research Protections Program. One of the participants exercised his legal right to die option in the state of California(117). No government funds were used as part of the option.
AUTHOR CONTRIBUTIONS
Conceived and designed the experiments: AC, SG, CI, MP, BV, SD, SR, DS
Performed the experiments: AC, CI, MP
Analyzed the data: AC, SG, DS, BV, SD, TS
Enrolled participants: SG, DS, SR
Wrote and revised the manuscript: AC, SG, DS, MFO, BV, SD, SR, TS

ACKNOWLEDGEMENTS
We are very grateful to the Last Gift participants and their families. We would like to acknowledge the investigators and team members in the CFAR Genomic and Sequencing and Translational Virology Cores. We would like to acknowledge Gemma Caballero, Laura Layman and Brianna Scott for their technical assistance. We would like to acknowledge Steven Hendrix and Susanna Concha Garcia for their assistance in participant enrollment, and Susan Little and Matt Strain for the very helpful discussions.

FUNDING
AC acknowledges funding from the National Institute of Health (AI131971-01, AI036214) and from the University of San Diego Center for AIDS Research (CFAR), a NIH funded program and San Diego CFAR TV Core. SG acknowledges funding from the National Institute of Health (HD094646, AI027763, AI134295, AI68636). DS acknowledges funding from the National Institute of Health (P01 AI31385, P30 AI036214).

MP was supported by the San Diego The Veterans Medical Research Foundation (VMRF). S.D. was and B.V. is supported by a postdoctoral fellowship from the Fonds voor Wetenschappelijk Onderzoek (FWO, Belgium). S.D. is currently supported by the Fonds National de la Recherche Scientifique (FNRS, Belgium).

COMPETING FINANCIAL INTERESTS:
AC, SG, SD, BV, MFO, SR, CI, MP, SJL and DMS do not have any commercial or other associations that might pose a conflict of interest.
REFERENCES


101. HIV Los Alamos National Laboratory. ElimDupes. [https://www.hiv.lanl.gov/content/sequence/ELIM Dupes/El im DupesExplain.html](https://www.hiv.lanl.gov/content/sequence/ELIMDupes/ElimD upesExplain.html).


FIGURE 1. Flow Diagram. Six participants enrolled in the Last Gift cohort (n=2 who stopped ART and n=4 who remained virally suppressed on ART until death) were included in this study. (1) Pre-mortem blood plasma (n=2 participants) and peripheral blood mononuclear cell (PBMC) samples (n=6 participants) were collected before death and tissues were collected during the rapid autopsy procedure. (2) HIV RNA and DNA were extracted from blood plasma and PBMC/tissues for quantification of HIV DNA/RNA (digital droplet PCR); (3) and (4) HIV full length envelopes were sequenced via single genome amplification and sequencing. Intact full length env sequences from all samples were used to (5) characterize the HIV populations within each compartments and in blood, (6) assess viral dispersal across tissues using Bayesian phylodynamic models and (7) evaluate factors associated with viral dispersal.
The proportion of nearly identical (any sequences with ≥99% nucleotide identity with at least one other sequence from the same participant); 2D. Proportion of identical (100%); 2E. Pairwise genetic distance between sequences and 2F. levels of HIV DNA. The pairwise genetic distance between sequences from a compartment was measured using the Tamura Nei 93 (TN93) algorithm (46). Level of HIV DNA was quantified by droplet digital PCR (ddPCR). Copy numbers were calculated as the mean of 3 replicate PCR measurements and normalized to one million cells, as determined by RPP30 (total cell count)(94, 95). See method section for details.
FIGURE 3. HIV DNA Levels in Compartments for each Participant. Level of HIV DNA was quantified by droplet digital PCR (ddPCR). Copy numbers were calculated as the mean of 3 replicate PCR measurements and normalized to one million cells, as determined by RPP30 (total cell count)(94, 95). See method section for details. The dots and horizontal bars represent the mean and range (minimum and maximum) of the HIV DNA levels.
FIGURE 4. Maximum Likelihood (ML) Phylogenies and Clonal Populations (Full Length Envelope) for 2 Participants Who Stopped ART (A) or Remained Virally Suppressed on ART (B). Phylogenies were estimated using IQtree(97) from the full length HIV env sequences obtained from pre-mortem blood plasma and from tissues and PBMC collected during rapid autopsy. A. ML phylogeny for Individual LG01 who stopped therapy. B. ML phylogeny for Individual LG03 who remained virally suppressed. Tips are colored by compartment as in the legend. Size and distribution of nearly clonal FL env populations (99% identical, populations of at least 3 identical proviruses) for each participant are presented in the middle of each tree. Colors represent tissues described in Left Panel legend. For LG01, nearly identical populations including HIV RNA viruses sampled in blood plasma during viral rebound are marked with *. See also Supplementary Figure S4 for the 4 remaining participants.

Compartment:

- Blood Plasma
- Cardiac Serum
- Colon (Left)
- Colon (Right)
- Duodenum
- Esophagus
- Frontal Cortex Motor
- Gut (any)
- Ileum
- Jejunum
- Kidney
- Liver
- Lymph Nodes (Aortic)
- Lymph Nodes (Axillary)
- Lymph Nodes (Hilar)
- Lymph Nodes (Inguinal)
- Lymph Nodes (Mediastinal)
- Lymph Nodes (Mesentary)
- Lymph Nodes (Peritrach.)
- Occipital Cortex
- Pancreas
- PBMC
- Pericardial Adipose
- Prostate
- Rectum
- Seminal Vesicle
- Spleen
- Testis
**FIGURE 5. Lineage Dispersal Events between Compartments.** Circle size is proportional to HIV DNA level (or RNA for plasma) in each compartment. The thickness of the arrows corresponds to the average number of inferred migration events between compartments. Only transition events between locations for which the adjusted BF ≥ 3 (at least positive evidence) are shown.
Figure 6. Proportion of Supported Transition Events between Compartments for 2 Participants Who Stopped ART (A) or Remained Virally Suppressed on ART (B). Sankey plot showing the proportion of transition events between locations for which the adjusted BF ≥ 3 (at least positive evidence). The adjusted BF support for each transition type is given next to the corresponding color. The source locations are depicted on the left side of the plots while the destination locations are given at the right side. See also Supplementary Figure S6.

A. LG01
B. LG03