Persistence of intact HIV-1 proviruses in the brain during antiretroviral therapy

Weiwei Sun[1], Yelizaveta Rassadkina[1], Ce Gao[1], Sarah Isabel Collens[2], Xiaodong Lian[1], Isaac H. Solomon[3], Shibani Mukerji[2], Xu G. Yu[1,4], Mathias Lichterfeld[1,4]

[1]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA

[2]Department of Neurology, Massachusetts General Hospital, Boston, MA

[3]Department of Pathology, Brigham and Women's Hospital, Boston, MA

[4]Infectious Disease Division, Brigham and Women's Hospital, Boston, MA

Corresponding author:
Mathias Lichterfeld, M. D., Ph. D.
Professor of Medicine, Harvard Medical School
Infectious Disease Division, Brigham and Women's Hospital
65 Landsdowne Street
Cambridge, MA 02139, USA
Email: mlichterfeld@partners.org

Abstract

HIV-1 reservoir cells that circulate in peripheral blood during suppressive antiretroviral therapy (ART) have been well characterized, but little is known about the dissemination of HIV-1-infected cells across multiple anatomical tissues, especially the central nervous system (CNS). Here, we performed single-genome, near full-length HIV-1 next-generation sequencing to evaluate the proviral landscape in distinct anatomical compartments, including multiple CNS tissues, from 3 ART-treated participants at autopsy. While lymph nodes and, to a lesser extent, gastrointestinal and genitourinary tissues represented tissue hotspots for the persistence of intact proviruses, we also observed intact proviruses in CNS tissue sections, particularly in the basal ganglia. Multi-compartment dissemination of clonal intact and defective proviral sequences occurred across multiple anatomical tissues, including the CNS, and evidence for the clonal proliferation of HIV-1-infected cells was found in the basal ganglia, in the frontal lobe, in the thalamus and in periventricular white matter. Deep analysis of HIV-1 reservoirs in distinct tissues will be informative for advancing HIV-1 cure strategies.

Introduction

Combination antiretroviral therapy has extended the life expectancy of people living with HIV (PLWH) to near-normal levels, but it is clear that antiretroviral therapy (ART), in its present form, does not lead to a sustained, drug-free remission of HIV-1 infection; instead, a long-lived reservoir of virally-infected cells persists despite ART (1, 2) and can effectively fuel rebound viremia in case of treatment interruptions. Hence, antiretroviral therapy needs to be taken life-long; finding a way to eliminate HIV-1 reservoir cells remains critical to finding a cure for HIV-1 infection.

HIV-1 reservoir cells that circulate in the peripheral blood have been well-described in recent years (3). These circulating reservoir cells mostly consist of memory CD4 T cells that persist life-long and harbor chromosomally-integrated viral DNA, also referred to as a “provirus”. The pool size of HIV-1-infected cells during suppressive antiretroviral therapy is frequently maintained or replenished by clonal proliferation of viral reservoir cells, during which an identical viral DNA copy is passed on to daughter cells (4-7). Important advances in recent technology development, including single-genome near full-length proviral sequencing, have demonstrated that the vast majority of HIV-1 DNA species encountered in ART-treated persons are defective (8-10), mostly due to errors occurring during reverse transcription of viral RNA into DNA. Such errors can lead to large deletions, premature stop codons, or defects in the viral packaging signal region; moreover, specific host proteins such as APOBEC-3G/3F can induce lethal hypermutations into the proviral sequence. The ability of viral reservoir cells to persist indefinitely is frequently attributed to very limited or absent proviral gene transcription; this viral latency protects infected cells from host immune responses and reduces possible cytopathic effects associated with viral gene expression. However, recent studies have emphasized that proviral gene expression can be ongoing in some HIV-1-infected cells during antiretroviral therapy (11), typically when proviruses are integrated in immediate proximity to activating chromatin marks (12).

Much less is currently known about HIV-1 reservoir cells in other body compartments that are more difficult to access for analytic purposes. Lymph nodes and lymphoid tissues in the gastrointestinal tract harbor the vast majority of all lymphocytes and are likely to represent a prime location for the persistence of virally-infected cells (13-16); however, studies that interrogate viral reservoir cells in these tissues remain limited. Even less is known about HIV-1 persistence in other organs, and, in particular, in the CNS, although recent studies have started to explore viral reservoir cells in such specific body compartments (17). In the present study, we used single-genome proviral sequencing to conduct a detailed analysis of HIV-1 proviral sequences in autopsy samples from up to n=18 different organ systems from three study participants.

Methods

Study Participants

HIV-1-infected study participants were recruited at the Massachusetts General Hospital (MGH) and the Brigham and Women's Hospital in Boston, MA. Fresh tissues were sampled during routine autopsy according to protocols approved by the Institutional Review Board and cryopreserved for future study according to standard protocols. The clinical characteristics of study participants are summarized in **Figure S1.**

HIV-1 DNA quantification by IPDA

Tissue samples were dissected and subjected to genomic DNA extraction using the DNeasy Blood and Tissue Kit (QIAGEN DNeasy, #69504). HIV-1 DNA was analyzed by the Intact Proviral DNA Assay (IPDA), using primers and probes described previously (18). PCR was performed using the following program: 95 °C for 10 min, 45 cycles of 94 °C for 30 s and 59 °C for 1 min, 98 °C for 10 min. The droplets were subsequently read by the QX200 droplet reader (Bio-Rad), and data were analyzed using QuantaSoft software (Bio-Rad).

Near full-length HIV proviral sequencing

Genomic DNA diluted to single HIV-1 genome levels based on Poisson distribution statistics and IPDA results was subjected to near full-genome HIV-1 amplification using a one-amplicon approach (9, 19). PCR products were visualized by agarose gel electrophoresis. Amplification products were individually subjected to Illumina MiSeq sequencing at the MGH DNA Core facility. The resulting short reads were de novo assembled using Ultracycler v1.0 and aligned to HXB2 to identify large deleterious deletions (<8000 bp of the amplicon aligned to HXB2), out-of-frame indels, premature/lethal stop codons, internal inversions, or 5′-LTR defect (≥15 bp insertions and/or deletions relative to HXB2), using an automated in-house pipeline written in Python scripting language (https:// github.com/BWH-Lichterfeld-Lab/Intactness-

Pipeline). The presence/absence of APOBEC-3G/3F–associated hypermutations were determined using the Los Alamos HIV Sequence Database Hypermut 2.0 program. The sequences of individual genes were extracted by GeneCutter, and the start codons of Gag, Pol, and Env were examined and considered. Viral sequences that lacked all defects listed above were classified as “genome-intact”. Multiple sequence alignments were performed using MUSCLE (20). Phylogenetic analyses were conducted using MEGA X, applying maximum likelihood approaches. Viral sequences were considered clonal if they had completely identical consensus sequences; single-nucleotide variations in primer binding sites were excluded for clonality analysis. When viral DNA sequences were undetectable, data were reported as LOD (limit of detection), calculated as 0.5 copies per maximum number of cells tested without target identification. The sensitivity of proviral species to broadly-neutralizing antibodies (bnAb) was estimated by calculating the number of amino acid signature sites associated with sensitivity to four bnAb classes within the env amino acid sequence from each intact provirus, as previously described (21).

Data Analysis and Statistics

Data are summarized as bar graphs. Phylogenetic relationships were evaluated using maximum-likelihood phylogenetic trees. Images were prepared using Adobe Illustrator. HIV-1 tropism was computationally inferred using Geno2pheno (https://coreceptor.geno2pheno.org/). HIV-1 tropism was classified as “CCR5” if the false-positive rate (FPR) predicted by Geno2pheno was >2%, however, 92% of our proviral sequences meeting this definition had a FPR score >10%; proviruses were considered “CXCR4-tropic” if FPR was <2%. Due to confidentiality concerns, proviral sequencing data cannot be publicly released, but will be made available upon reasonable request and after signing an institutional data-sharing agreement.

Results

Frequency of intact and defective proviruses in tissue compartments

To investigate the proviral landscape across multiple anatomical tissues, including the central nervous system (CNS), we focused on 3 participants from whom post-mortem tissue samples were available for HIV-1 research. Tissue samples from the CNS and other organs were collected by a rapid (<24 hour) autopsy after death. The clinical and demographic characteristics of these study participants are shown in **Figure S1A**. All study participants adhered to antiretroviral treatment until death; plasma viral loads were undetectable by commercial assays in study participants 1 and 2, in whom 14 and 15 different tissue sections were sampled, respectively (**Figure S1B**). Organ-specific tissues analyzed in these two study participants included lymph node, spleen, colon, liver, pancreas, kidney, thyroid gland, and adrenal gland; in the female study participant 1, ovarian and uterus tissues were analyzed, while in the male study participant 2, prostate and testicular tissues were studied. In both of these study participants, four different CNS tissue sections (basal ganglia, thalamus, frontal lobe, and occipital lobe) were collected for investigation. In study participant 3, plasma viral load was 136 copies/ml at the time of death; 5 different tissue sections from the CNS (basal ganglia, thalamus, occipital lobe, frontal lobe, periventricular white matter) were analyzed in this person.

Near full-length single-template next-generation HIV-1 proviral sequencing was performed to profile the proviral reservoir landscape at single-molecule resolution in tissue samples. The number of cells analyzed from each organ in each participant is listed in **Figure S1B**. In total, 846.53, 425.54, and 199.66 million cells were assayed in study participants 1, 2, and 3, respectively, resulting in 1471.73 million cells analyzed in all study participants combined. A total of 1497 individual proviral sequences were amplified, of which n=497 were selected for next-generation sequencing based on their amplicon sizes on gel electrophoresis; the remaining sequences were classified as proviruses with large deletions. All amplicons (n=74) from CNS tissues were

sequenced, regardless of their length. Using a previously-described computational pipeline to identify lethal defects in proviral sequences, we identified 48 proviruses (3.21% of all proviruses) that met our criteria for genome-intactness (**Figure 1A-C**); this number is consistent with the small number of genome-intact proviruses detected in previous studies. Many sequences, both genome-intact and defective, were identified multiple times, consistent with clonal proliferation of infected cells (**Figure 1B**), as reported in prior work (7, 9, 10, 22).

To evaluate HIV-1 persistence in selected tissue compartments, the frequencies of total, intact, and defective HIV-1 proviruses in each tissue were analyzed (**Figure 2A-C**). Intact proviral sequences were only detected in 8 tissue sites, including basal ganglia, periventricular white matter, lymph node, spleen, colon, kidney, prostate, and thyroid (**Figure 2B**). The numbers of intact HIV-1 sequences in these 8 tissues varied from 0.01 to 0.6 copies per million cells. Consistent with previous studies, the frequency of intact proviral sequences was highest in the lymph node in participant 1 (0.52 intact proviruses/million cells) and participant 2 (0.58 intact proviruses/million cells), followed by kidney, spleen, colon, and basal ganglia in participant 1 and by prostate, spleen, thyroid in participant 2. Intact proviruses were detected in the basal ganglia in study participant 1 (frequency of 0.015/million cells) and in study participant 3 (0.030 intact proviruses/million cells). Moreover, one intact provirus was also detected in periventricular white matter in participant 3 (0.027 intact proviruses/million cells), in whom analysis was limited to brain tissues. No intact proviruses were detected in the CNS tissues of study participant 2, despite analyzing 66.29 million cells. Together, these results indicate that intact HIV-1 proviruses are preferentially detected in lymphoid and gastrointestinal (GI) tissues. The frequency of intact proviruses in the CNS is comparatively low; however, this study is the first one to document the presence of genome-intact proviral sequences in CNS tissues using near full-length proviral sequencing.

Defective proviral sequences were detected in all analyzed tissue samples except for

the thalamus from participant 1 (**Figure 2C**). In participant 1, the frequency of defective proviral sequences was highest in lymph nodes (5.7 defective proviruses/million cells), followed by colon, spleen, and kidney. In participant 2, the frequency of defective proviral sequences was highest in lymph nodes (13.0 defective proviruses/million cells), followed by prostate, colon, and spleen. Notably, the prostate had a very high frequency of virally-infected cells. The ratio of intact to defective proviral species was relatively high among sequences isolated from the basal ganglia of participants 1 and 3 (**Figure S2**). Taken together, these results demonstrate the highest frequencies of defective proviruses in lymph nodes, in the colon and in the prostate. CNS tissues contained relatively low frequencies of proviral sequences, compared to other tissue sites; however, defective proviruses were isolated in all but one of the analyzed 13 different CNS samples.

Phylogenetic associations and clonality

A series of prior studies demonstrated large sequence-identical clones of intact and defective proviruses in the peripheral blood of ART-treated study participants. In our subsequent analysis, we studied the dissemination of clonal proviral sequences across different tissues. In participant 1, a total of 24 intact proviral sequences were detected. Two large clones of intact proviruses were observed, one of which included sequences detected in kidney, lymph node, and spleen samples (**Figure 3A-B**). The other clone involved intact proviral sequences from basal ganglia and lymph node tissues. Among 218 defective proviruses sequenced in participant 1, 14 clones were observed across multiple tissues (**Figure 3C**). In participant 2, among 22 intact proviral sequences, only one clone with two member sequences was identified; both of these clonal sequences were located in the prostate (**Figure 3A-B**). Fourteen clones of defective proviruses were observed across multiple tissues in participant 2 (**Figure 3C**).

Clonal proviral sequences in CNS tissues were detected in all three study participants. In study participant 1, members of a clone of defective proviruses were observed in the basal ganglia and the frontal lobe (**Figure 4A-B**). In participant 2, two clones of defective proviruses were noted in the thalamus, one clone was detected in cells from

the basal ganglia, and members of a fourth clone were detected in the thalamus and the occipital lobe (**Figure 4C-D**), demonstrating rather extensive evidence for clonal proliferation of virally-infected cells in the brain. Two of the 44 proviral sequences detected in the central nervous system of participant 3 met the criteria for genome-intactness; one of those was isolated from the basal ganglia and one from periventricular white matter. Defective proviruses, most of them harboring large deletions, were detected in all 5 CNS tissues from participant 3 (**Figure 4E-F**). One large clone involving 9 defective proviral sequences in participant 3 was broadly distributed across different brain tissues, encompassing sequences in the occipital lobe, basal ganglia, thalamus, and periventricular white matter. The other clone of defective proviruses was only detected in the periventricular white matter. Again, these data suggest that clonal proliferation is a rather common feature of HIV-1 reservoir cells in the CNS.

Viral tropism and immune selection footprints

Viral tropism was evaluated based on the env V3 region of the proviruses using the Geno2pheno algorithm. Notably, all proviral sequences containing the env V3 region of participant 1 (n=130, 53.7%) and 3 (n=6, 13.6%) were predicted to be CCR5-tropic (**Figure 5A**)**.** In participant 2, 62.1% of proviral sequences (n=131) were predicted to be likely CXCR4-tropic, while 18.5% (n=39) were classified as CCR5-tropic; the remaining 19.4% (n=41) were classified as undetermined due to the lack of env V3 regions in this study participant (**Figure 5A**). Notably, approximately half of the 131 proviruses with predicted CXCR4 tropism from study participant 2 (n=66, 50.4%) were isolated from the prostate, followed by the spleen (n=38, 29.0%), lymph node (n=20, 15.3%) and thyroid gland (n=3, 2.3%) (**Figure 5B-C**). Among all proviruses (n=16) from CNS tissues of participant 2, only one provirus with a large deletion, isolated from the occipital lobe, had predicted CXCR4-tropism; the tropism of other proviruses from the CNS was unknown due to the lack of env V3 regions. As an additional analysis step, we evaluated footprints of immune selection pressure and mutations resulting in resistance to antiretroviral agents in intact proviral sequences from our study subjects.

We noted that the frequency of viral amino acid residues associated with resistance to broadly-neutralizing antibodies did not notably differ among sequences isolated from different tissue compartments (**Figure S3**). We did not observe sequence variations consistent with escape from antiretroviral agents in any of the intact proviral sequences analyzed here.

Discussion

The lifelong persistence of viral reservoir cells makes HIV-1 infection an incurable disease that necessitates indefinite antiretroviral suppression therapy. However, the location of HIV-1 viral reservoir cells across different tissues has been difficult to assess in the past, due to the limited availability of tissue samples. Recent studies, pioneered by investigators of the “Last Gift Cohort”, have catalyzed investigations to characterize HIV-1 sequences in diverse organ systems, specifically in the CNS (23, 24). In our study, we used single-genome near full-length proviral sequencing to evaluate the distribution of HIV-1 reservoir cells in multiple anatomical compartments from autopsy samples of three individuals living with HIV-1 and receiving antiretroviral therapy until the time of their decease. Consistent with previous studies (25, 26), intact proviruses were readily detected in lymph nodes; moreover, we detected relatively high frequencies of intact proviruses in the colon, likely reflecting viral infection of CD4 T cells residing in gut-associated lymphoid tissues (GALT). Of note, our study is among the first investigations to identify near full-length proviral sequences from the central nervous system in two study participants, supporting the hypothesis that CNS cells can serve as reservoirs for long-term HIV-1 persistence despite antiretroviral therapy. Importantly, we noted that large clones of virally infected cells were broadly disseminated across multiple tissues, and, in selected cases, involved cells from the CNS; this suggests that HIV-1 reservoir cells seeded to the brain via hematogenous spread can proliferate in the local tissue microenvironment of the central nervous system. Together, our work suggests that HIV-1 reservoir cells harboring intact proviruses are broadly distributed across multiple anatomical locations, involving lymphoid tissues, gastrointestinal tissues, genitourinary tissues, and, importantly, central nervous system tissues.

HIV-1 can invade the central nervous system (CNS) within days after infection, as demonstrated in animal (27) and in human studies (28). Most likely, infection of CNS cells occurs as a result of transmigration of infected CD4 T cells and, possibly, macrophages across the blood-brain-barrier (BBB) (29, 30), a process that may be

facilitated by the increased permeability of the BBB during the initial, highly-replicative stage of HIV-1 infection. Infected cells that successfully enter the CNS may frequently be short-lived, however some infected CD4 T cells in the brain may persist long-term. Moreover, invading infected cells can transmit the virus to resident CNS cells via effective cell-to-cell transmission (30). At least three different CNS cell types seem to be susceptible to HIV-1 infection: perivascular macrophages, microglial cells, and astrocytes, although the role of the latter as HIV-1 target cells is more controversial (31-33). Yet, due to the difficulties in accessing brain tissues for analytic purposes, the role of the central nervous system in HIV-1 persistence during antiretroviral therapy remained largely unknown for a long time. Using the intact proviral DNA assay (IPDA), a ddPCR-based technique allowing to identify proviruses with a high probability of being genome-intact, previous investigators identified intact proviruses in 6 out of 9 ART-treated persons (17), although the precise proviral DNA sequence and their possible clonality was not assessed with this technology. In our study, a total of 13 CNS tissue samples from three study participants were analyzed, including specimens from the basal ganglia, thalamus, occipital lobe, frontal lobe, and periventricular white matter. In two study persons (participants 1 and 3), intact proviral sequences were detected in basal ganglia, suggesting that HIV-1 may preferentially persist in this anatomical compartment in the CNS; an additional intact provirus was detected in periventricular white matter. Notably, the intact provirus from basal ganglia in one of our study persons (participant 1) was clonal with 4 intact proviruses from the lymph node, indicating, to our knowledge for the first time, that CNS tissue can be involved in the multi-compartment dissemination of large clones of HIV-1 proviruses in ART-treated persons. Moreover, in multiple instances, we observed clones of HIV-1-infected cells that were distributed across different autologous CNS tissues, specifically in study participant 3, suggesting local spread of virally infected cells through clonal proliferation within the immune microenvironment of the CNS.

Our study did not allow to determine which cell types were infected by HIV-1 and responsible for clonal expansion of viral reservoir cells in the CNS; however, it is

329 possible that infected microglia are involved. Microglial cells originate from
330 erythromyeloid progenitors in the yolk sac and colonize the developing CNS during
331 embryogenesis (34); these cells act as the main innate immune cell population of the
332 CNS. Due to their long half-life (typically several years), their ability to divide and self-
333 renew, and their high cell-intrinsic susceptibility to HIV-1 (35), these cells may
334 represent a primary cellular site for long-term HIV-1 persistence in the CNS. In
335 particular, self-renewal through homeostatic proliferation in microglia (36) may support
336 HIV-1 persistence through clonal expansion. A recent study indeed identified HIV-1
337 DNA and RNA in microglia cells from autopsies from ART-treated PLWH who did not
338 have specific (HIV or non-HIV associated) CNS pathology (37). Evidence for HIV-1
339 persistence in CD68+ myeloid cells, most likely microglia cells, was also described by
340 previous investigators (17). That said, the presence of clonal proviral sequences
341 shared between the CNS and lymphoid tissues suggests that migrating CD4 T cells
342 infected with R5-tropic viruses may infect the brain as “Trojan horses”, and then
343 potentially clonally expand in situ in the CNS; this hypothesis is consistent with recent
344 findings from Kincer et al (38). In the future, it may be possible to capture the
345 phenotypic characteristics of HIV-1 reservoir cells from the CNS directly with single-
346 cell assays that permit combined assessments of the phenotype and the proviral
347 sequence; an example for such an assay system was recently described (39).
348
349 In our study, CXCR4-tropic proviruses were exclusively detected in participant 2.
350 Compared to the other two participants who began ART shortly after HIV-1 diagnosis,
351 participant 2 was diagnosed with HIV-1 25 years prior to starting antiretroviral therapy
352 and died 10 months after ART commencement; therefore, viral CXCR4 tropism most
353 likely resulted from “coreceptor switch” frequently occurring during advanced stages of
354 immune deficiency (40). Whether the preferential persistence of CXCR4-tropic viruses
355 in study participant 2 was associated with our inability to detect intact proviruses in the
356 CNS in this person is unclear; however, prior studies suggested that CCR5 can act as
357 the principal co-receptor for HIV-1 isolates in the brain (41, 42). Moreover, rebound
358 viremia in cerebrospinal fluid after ART interruption is mostly fueled by CCR5-tropic

359 virus (38), further supporting the assumption that R5-tropic viruses are better adjusted
360 to persist in the brain. Another notable finding in our study was the high number of HIV-
361 1 proviruses isolated from the prostate, which had the second highest proviral
362 frequency among all analyzed tissues in study participant 2, second only to lymph node
363 samples. Other studies also reported that the prostate can represent a tissue reservoir
364 for HIV-1 (24). We were unable to identify the precise cell type harboring HIV-1 in the
365 prostate, but it is possible that myeloid cells may harbor HIV-1 in this location. A prior
366 study indeed demonstrated that intact, replication-competent HIV-1 can persist in
367 myeloid cells from the urethra, located in immediate anatomical proximity to the
368 prostate (43).
369
370 Our study has several limitations. Importantly, this work includes only 3 participants,
371 and brain tissues were the only samples available from participant 3. Moreover, due to
372 limited tissue sizes, very few cells were assayed from the terminal ileum and testes;
373 prior studies suggested high HIV-1 DNA levels in these 2 tissues during suppressive
374 antiretroviral therapy (44, 45). Another limitation was that peripheral blood samples
375 were not available from the 3 participants, which made it impossible to study
376 phylogenetic associations between tissue reservoirs of HIV-1 relative to viral species
377 circulating in peripheral blood. Moreover, we cannot fully exclude contamination of
378 tissue samples with cells from peripheral blood. However, after autopsy, the tissue
379 samples were washed thoroughly with PBS to eliminate blood as much as possible.
380 Notably, we failed to detect intact proviral sequences from over 100 million liver cells
381 of 2 participants, despite the fact that the liver contains about 13% of all human blood
382 supply at any given time point, arguing against contamination of tissues with blood
383 cells.
384
385 In sum, this study provides a deep analysis of tissue reservoirs for HIV-1 that includes
386 a detailed assessment of HIV-1 sequences in CNS tissues. Our work supports the
387 persistence of genome-intact HIV-1 in many tissues, including CNS tissues,
388 emphasizing the difficulties in finding strategies to effectively eliminate HIV-1 from the

389 human body in clinical settings.

References

1. 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(6629):183-8.
2. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science.* 1997;278(5341):1291-5.
3. Margolis DM, Archin NM, Cohen MS, Eron JJ, Ferrari G, Garcia JV, et al. Curing HIV: Seeking to Target and Clear Persistent Infection. *Cell.* 2020;181(1):189-206.
4. Einkauf K, Lee G, Gao C, Sharaf R, Sun X, Hua S, et al. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest.* 2019;in press.
5. Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, et al. HIV-1 Integration Landscape during Latent and Active Infection. *Cell.* 2015;160(3):420-32.
6. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DI, et al. Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *J Exp Med.* 2017;214(4):959-72.
7. Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog.* 2017;13(3):e1006283.
8. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell.* 2013;155(3):540-51.
9. Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. *J Clin Invest.* 2017;127(7):2689-96.
10. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell Rep.* 2017;21(3):813-22.
11. Yukl SA, Kaiser P, Kim P, Telwatte S, Joshi SK, Vu M, et al. HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med.* 2018;10(430).
12. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, et al. Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell.* 2022.
13. Estes JD, Kityo C, Ssali F, Swainson L, Makamdop KN, Del Prete GQ, et al. Defining total-body AIDS-virus burden with implications for curative strategies. *Nat Med.* 2017;23(11):1271-6.
14. Baiyegunhi OO, Mann J, Khaba T, Nkosi T, Mbatha A, Ogunshola F, et al. CD8 lymphocytes mitigate HIV-1 persistence in lymph node follicular helper T cells during hyperacute-treated infection. *Nat Commun.* 2022;13(1):4041.

15. Kroon E, Chottanapund S, Buranapraditkun S, Sacdalan C, Colby DJ, Chomchey N, et al. Paradoxically Greater Persistence of HIV RNA-Positive Cells in Lymphoid Tissue When ART Is Initiated in the Earliest Stage of Infection. *J Infect Dis.* 2022;225(12):2167-75.
16. Beckford-Vera DR, Flavell RR, Seo Y, Martinez-Ortiz E, Aslam M, Thanh C, et al. First-in-human immunoPET imaging of HIV-1 infection using (89)Zr-labeled VRC01 broadly neutralizing antibody. *Nat Commun.* 2022;13(1):1219.
17. Cochrane CR, Angelovich TA, Byrnes SJ, Waring E, Guanizo AC, Trollope GS, et al. Intact HIV Proviruses Persist in the Brain Despite Viral Suppression with ART. *Ann Neurol.* 2022;92(4):532-44.
18. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature.* 2019;566(7742):120-5.
19. Lee GQ, Reddy K, Einkauf KB, Gounder K, Chevalier JM, Dong KL, et al. HIV-1 DNA sequence diversity and evolution during acute subtype C infection. *Nat Commun.* 2019;10(1):2737.
20. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-7.
21. Bricault CA, Yusim K, Seaman MS, Yoon H, Theiler J, Giorgi EE, et al. HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design. *Cell Host Microbe.* 2019;26(2):296.
22. Pinzone MR, VanBelzen DJ, Weissman S, Bertuccio MP, Cannon L, Venanzi-Rullo E, et al. Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun.* 2019;10(1):728.
23. Tang Y, Chaillon A, Gianella S, Wong LM, Li D, Simermeyer TL, et al. Brain microglia serve as a persistent HIV reservoir despite durable antiretroviral therapy. *J Clin Invest.* 2023;133(12).
24. Chaillon A, Gianella S, Dellicour S, Rawlings SA, Schlub TE, De Oliveira MF, et al. HIV persists throughout deep tissues with repopulation from multiple anatomical sources. *J Clin Invest.* 2020;130(4):1699-712.
25. Kuo HH, Banga R, Lee GQ, Gao C, Cavassini M, Corpataux JM, et al. Blood and lymph node dissemination of clonal genome-intact HIV-1 DNA sequences during suppressive antiretroviral therapy. *J Infect Dis.* 2020.
26. Banga R, Procopio FA, Noto A, Pollakis G, Cavassini M, Ohmiti K, et al. PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nat Med.* 2016;22(7):754-61.
27. Chakrabarti L, Hurtrel M, Maire MA, Vazeux R, Dormont D, Montagnier L, et al. Early viral replication in the brain of SIV-infected rhesus monkeys. *Am J Pathol.* 1991;139(6):1273-80.
28. Valcour V, Chalermchai T, Sailasuta N, Marovich M, Lerdlum S, Suttichom D, et al. Central nervous system viral invasion and inflammation during acute HIV infection. *J Infect Dis.* 2012;206(2):275-82.
29. Spudich S, and Gonzalez-Scarano F. HIV-1-related central nervous system

disease: current issues in pathogenesis, diagnosis, and treatment. *Cold Spring Harb Perspect Med.* 2012;2(6):a007120.

30. Liu Y, Tang XP, McArthur JC, Scott J, and Gartner S. Analysis of human immunodeficiency virus type 1 gp160 sequences from a patient with HIV dementia: evidence for monocyte trafficking into brain. *J Neurovirol.* 2000;6 Suppl 1:S70-81.
31. Churchill MJ, Gorry PR, Cowley D, Lal L, Sonza S, Purcell DF, et al. Use of laser capture microdissection to detect integrated HIV-1 DNA in macrophages and astrocytes from autopsy brain tissues. *J Neurovirol.* 2006;12(2):146-52.
32. Wallet C, De Rovere M, Van Assche J, Daouad F, De Wit S, Gautier V, et al. Microglial Cells: The Main HIV-1 Reservoir in the Brain. *Front Cell Infect Microbiol.* 2019;9:362.
33. Woodburn BM, Kanchi K, Zhou S, Colaianni N, Joseph SB, and Swanstrom R. Characterization of Macrophage-Tropic HIV-1 Infection of Central Nervous System Cells and the Influence of Inflammation. *J Virol.* 2022;96(17):e0095722.
34. Kierdorf K, Erny D, Goldmann T, Sander V, Schulz C, Perdiguero EG, et al. Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat Neurosci.* 2013;16(3):273-80.
35. Cenker JJ, Stultz RD, and McDonald D. Brain Microglial Cells Are Highly Susceptible to HIV-1 Infection and Spread. *AIDS Res Hum Retroviruses.* 2017;33(11):1155-65.
36. Reu P, Khosravi A, Bernard S, Mold JE, Salehpour M, Alkass K, et al. The Lifespan and Turnover of Microglia in the Human Brain. *Cell Rep.* 2017;20(4):779-84.
37. Ko A, Kang G, Hattler JB, Galadima HI, Zhang J, Li Q, et al. Macrophages but not Astrocytes Harbor HIV DNA in the Brains of HIV-1-Infected Aviremic Individuals on Suppressive Antiretroviral Therapy. *J Neuroimmune Pharmacol.* 2019;14(1):110-9.
38. Kincer LP, Joseph SB, Gilleece MM, Hauser BM, Sizemore S, Zhou S, et al. Rebound HIV-1 in cerebrospinal fluid after antiviral therapy interruption is mainly clonally amplified R5 T cell-tropic virus. *Nat Microbiol.* 2023;8(2):260-71.
39. Sun W, Gao C, Hartana CA, Osborn MR, Einkauf KB, Lian X, et al. Phenotypic signatures of immune selection in HIV-1 reservoir cells. *Nature.* 2023;614(7947):309-17.
40. Connor RI, Sheridan KE, Ceradini D, Choe S, and Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *J Exp Med.* 1997;185(4):621-8.
41. Albright AV, Shieh JT, Itoh T, Lee B, Pleasure D, O'Connor MJ, et al. Microglia express CCR5, CXCR4, and CCR3, but of these, CCR5 is the principal coreceptor for human immunodeficiency virus type 1 dementia isolates. *J Virol.* 1999;73(1):205-13.
42. He J, Chen Y, Farzan M, Choe H, Ohagen A, Gartner S, et al. CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature.* 1997;385(6617):645-9.

522 43. Ganor Y, Real F, Sennepin A, Dutertre CA, Prevedel L, Xu L, et al. HIV-1
523 reservoirs in urethral macrophages of patients under suppressive antiretroviral
524 therapy. *Nat Microbiol.* 2019;4(4):633-44.
525 44. Horn C, Augustin M, Ercanoglu MS, Heger E, Knops E, Bondet V, et al. HIV
526 DNA reservoir and elevated PD-1 expression of CD4 T-cell subsets particularly
527 persist in the terminal ileum of HIV-positive patients despite cART. *HIV Med.*
528 2021;22(5):397-408.
529 45. Miller RL, Ponte R, Jones BR, Kinloch NN, Omondi FH, Jenabian MA, et al.
530 HIV Diversity and Genetic Compartmentalization in Blood and Testes during
531 Suppressive Antiretroviral Therapy. *J Virol.* 2019;93(17).
532
533

Acknowledgements

ML is supported by NIH grants AI117841, AI120008, AI130005, DK120387, AI152979, AI155233, AI135940 and by the American Foundation for AIDS Research (amfAR, #110181-69-RGCV). XGY is supported by NIH grants AI155171, AI116228, AI078799, MH134823, HL134539, DA047034, amfAR ARCHE Grant # 110393-72-RPRL and the Bill and Melinda Gates Foundation (INV-002703). ML and XGY are members of the DARE, ERASE, PAVE and BEAT-HIV Martin Delaney Collaboratories (UM1 AI164560, AI164562, AI164566, AI164570). IHS is supported by NIH grant R21NS119660.

Author contributions

Sample collection and preservation: SIC, IHS, SM
HIV-1 sequencing assays: WS, YR
Data analysis, interpretation, presentation: WS, CG, XL, XGY, ML
Study concept and supervision: IHS, SM, XGY and ML

Conflicts of Interest

The authors declare that conflicts of interest do not exist.

552 **Figure legends**
553
554 **Figure 1. Proviral sequence classification in all analyzed HIV-1-infected cells**
555 **from 3 study participants.** (A) Pie charts reflecting proportions of proviruses
556 classified as intact or defective. All proviruses identified by single-genome, near-full-
557 length, next-generation sequencing and by counting amplification products in agarose
558 gel electrophoresis were included. The total number of individual sequences included
559 is listed below each pie chart. (B) Pie charts indicating proportions of intact and
560 defective proviruses detected once (classified as non-clonal) and detected multiple
561 times (classified as clonal). The total number of proviral sequences identified by single-
562 genome, near-full-length, next-generation sequencing is listed below each pie chart.
563 (C) Virograms summarizing individual HIV-1 proviral sequences aligned to the HXB2
564 reference genome from each participant; color coding reflects the classification of
565 proviral sequences.
566
567 **Figure 2. Distribution of total, intact, and defective HIV-1 proviruses in individual**
568 **tissue compartments.** Bar diagrams reflect relative frequencies of total (A), intact (B),
569 and defective (C) proviruses in all analyzed tissues in study participants 1-3. The total
570 number of individual proviral sequences determined by single-genome, near-full-length,
571 next-generation sequencing and by counting amplification products in agarose gel
572 electrophoresis from each tissue site of each participant is listed aside each bar. The
573 red bars reflect samples with detectable proviral sequences; grey bars reflect samples
574 at limit of detection for proviral sequences, calculated as 0.5 (single genome near–full-
575 length PCR) copies per maximum number of cells tested without target identification
576 (see Materials and Methods for details). N.d. (not done) indicates that the samples
577 were not available from the indicated tissue sites.
578
579 **Figure 3. Dissemination of HIV-1-infected cells across multiple anatomical**
580 **tissues in participants 1 and 2.** (A and B): Circular maximum likelihood phylogenetic
581 trees (A) and circus plots (B) of intact proviral sequences from participant 1 (P1) and

participant 2 (P2). HXB2, reference HIV-1 sequence. Color coding reflects tissue origins. Each symbol reflects one intact provirus. Clonal intact sequences, defined by complete sequence identity, are indicated by blue arches in (A) and by internal connections in (B). (C) Circos plots reflecting the clonality of defective proviral sequences from participant 1 (left panel) and participant 2 (right panel). Each symbol reflects one defective provirus. Clonal sequences, defined by complete sequence identity, are highlighted. Color-coded arches around the plots indicate types of defects in HIV-1 genomes.

Figure 4. Dissemination of HIV-1-infected cells across CNS tissues. (A, C, and E) Circular maximum likelihood phylogenetic trees of all proviral sequences derived from CNS tissues of the 3 study participants (A, participant 1; C, participant 2; E, participant 3). Color coding reflects tissue origins. Clonal sequences, defined by complete sequence identity, are indicated by blue arches. (B, D, and F) Circos plot reflecting the clonality of all proviral sequences isolated from CNS tissues of 3 participants (B, participant1; D, participant 2; F, participant 3). Each symbol reflects one provirus. Clonal sequences, defined by complete sequence identity, are highlighted. Color-coded arches around the plots indicate types of proviral sequences.

Figure 5. HIV-1 tropism analysis of proviral sequences. (A) Pie charts indicating the proportions of all proviruses from each participant with CCR5-tropic or non–CCR5/CXCR4-tropic V3 envelope sequences are shown. The total number of proviral sequences included in this analysis is listed below each pie chart. (B) The proportions of proviruses with CCR5-tropic or non-CCR5/CXCR4-tropic V3 env sequences in each tissue from participant 2 are shown. (C) Pie charts indicating the proportions of all non–CCR5-tropic proviruses from participant 2 are shown. Color coding reflects tissue origins. The total number of analyzed proviral sequences is listed below the pie chart. HIV-1 tropism was computationally inferred using Geno2pheno (https://coreceptor.geno2pheno.org/). HIV-1 tropism was classified as “CCR5” if the

611 false-positive rate (FPR) predicted by Geno2pheno was >2% and “CXCR4” if FPR was

612 <2%.

Supplemental Material

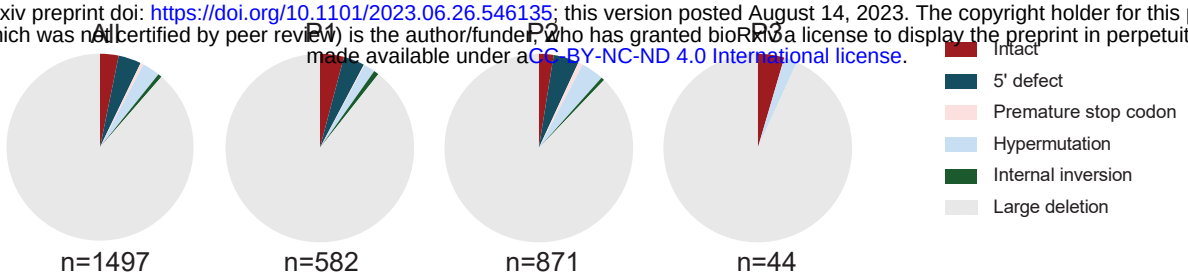
Supplemental Figure 1: (A) Clinical and demographical data of study participants. (B) Cell numbers analyzed from each tissue of each study participant.

Supplemental Figure 2: (A) Bar diagrams reflect the ratios of intact to defective HIV-1 proviruses in all analyzed tissues in study participants 1-3. The red bars reflect samples with detectable proviral sequences; grey bars reflect samples at limit of detection for proviral sequences, calculated as 0.5 proviral copies per maximum number of cells tested without target identification (see Materials and Methods for details). N.d. (not done) indicates that the samples were not available from the indicated tissue sites.

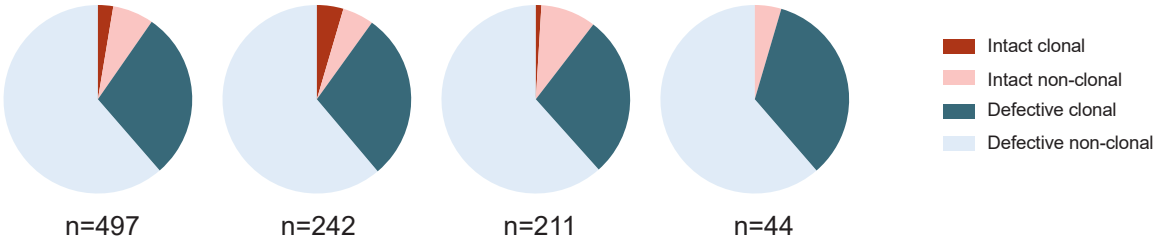
Supplemental Figure 3: (A-F) Numbers of broadly neutralizing antibody (bnAb) sensitivity (A, C, and E) and resistance (B, D, and F) signature sites in intact proviral sequences from indicated tissues from each participant are shown. Each dot represents one intact provirus. Amino acid residues associated with susceptibility or resistance to bnAbs were inferred based on the study by Bricault (21).

Figure 1

A



B



C

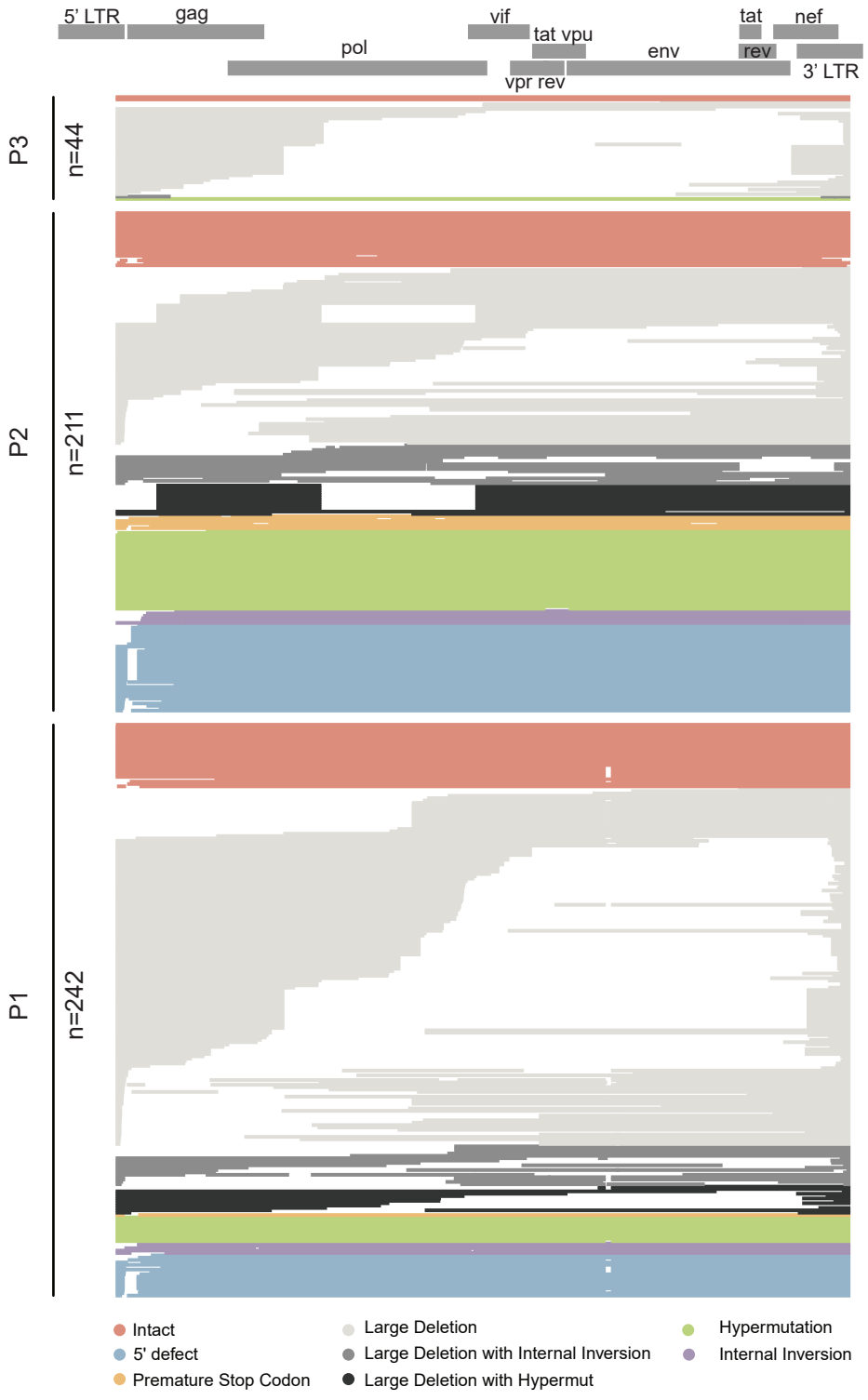


Figure 2

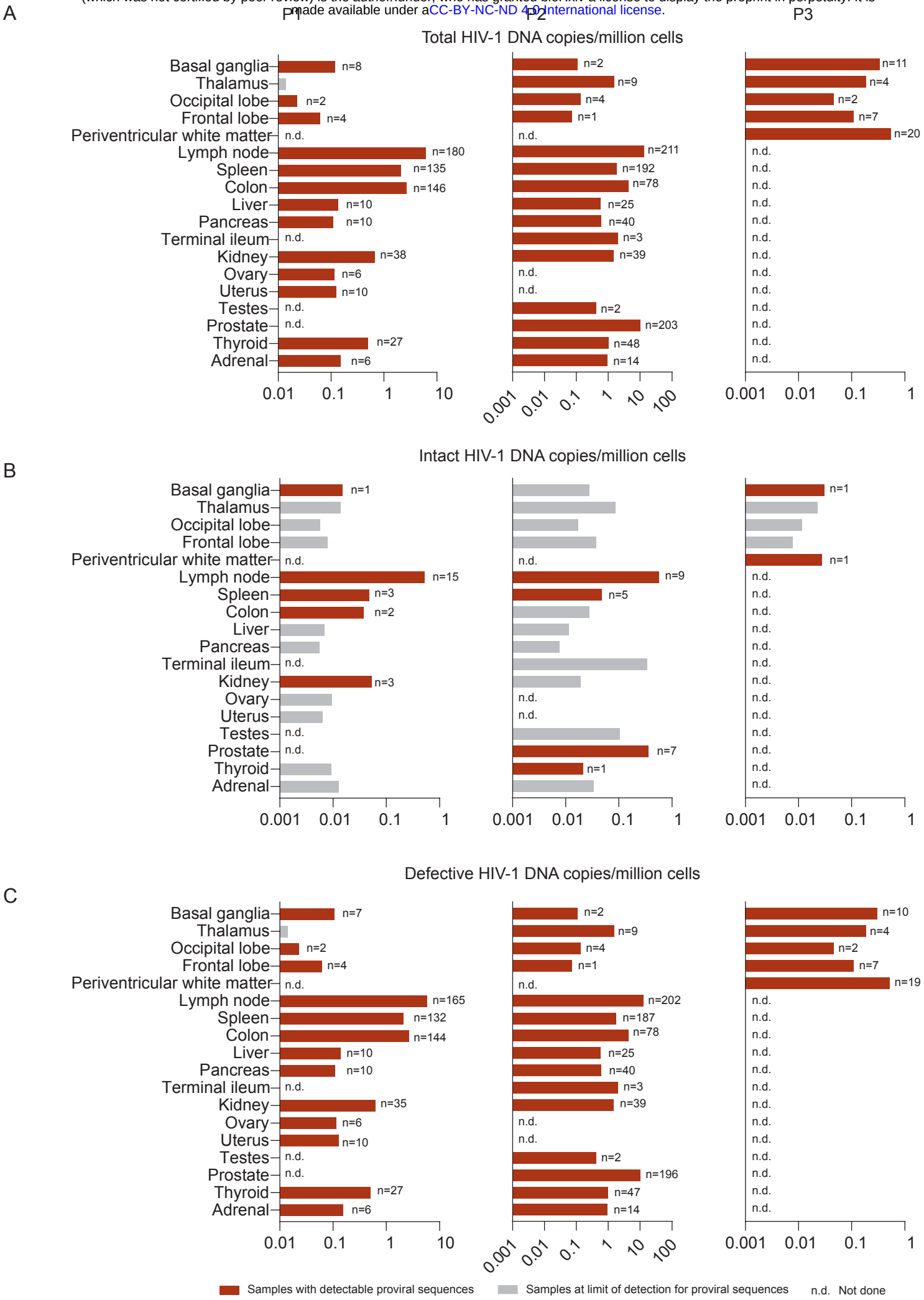


Figure 3

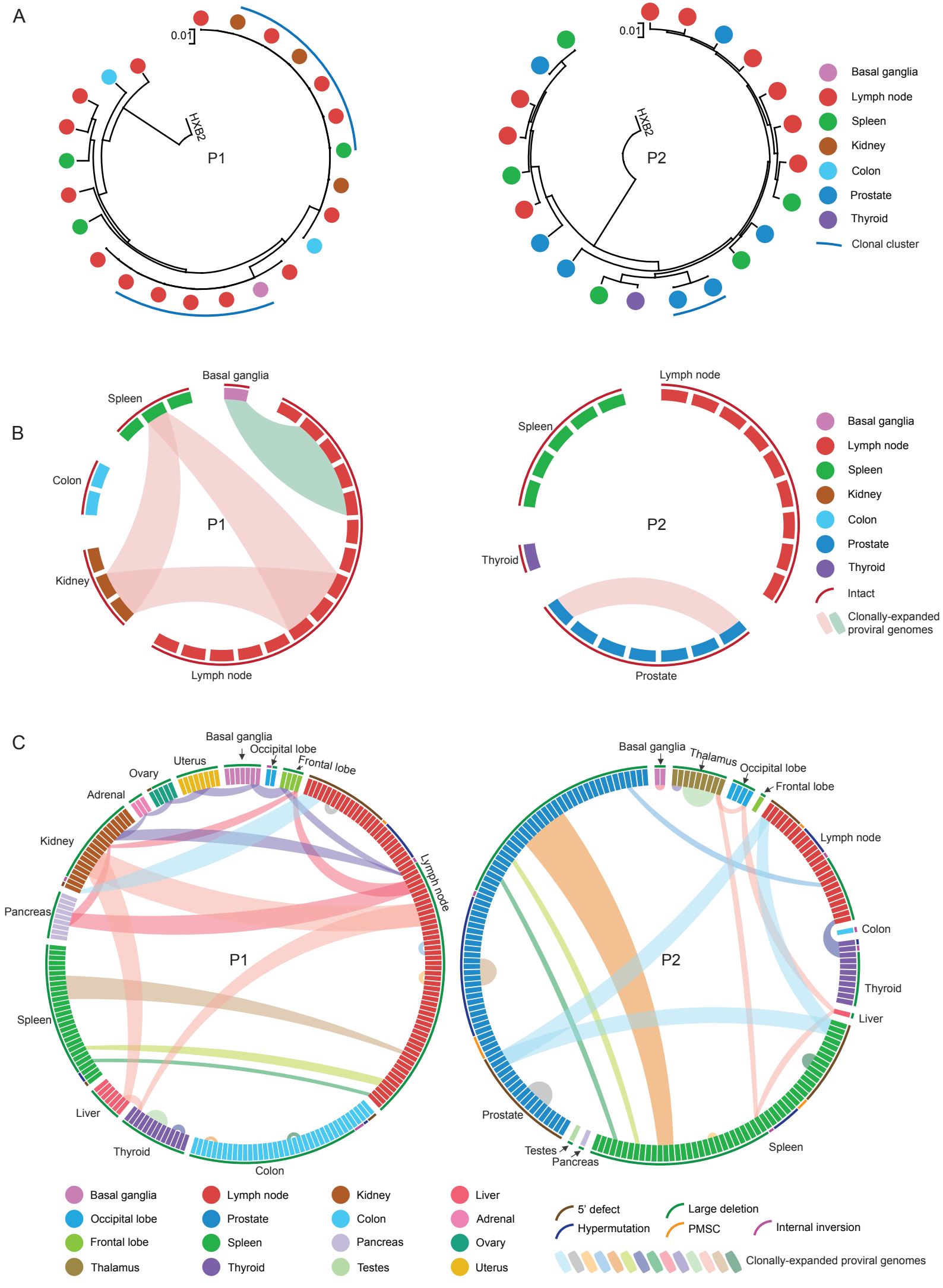


Figure 4

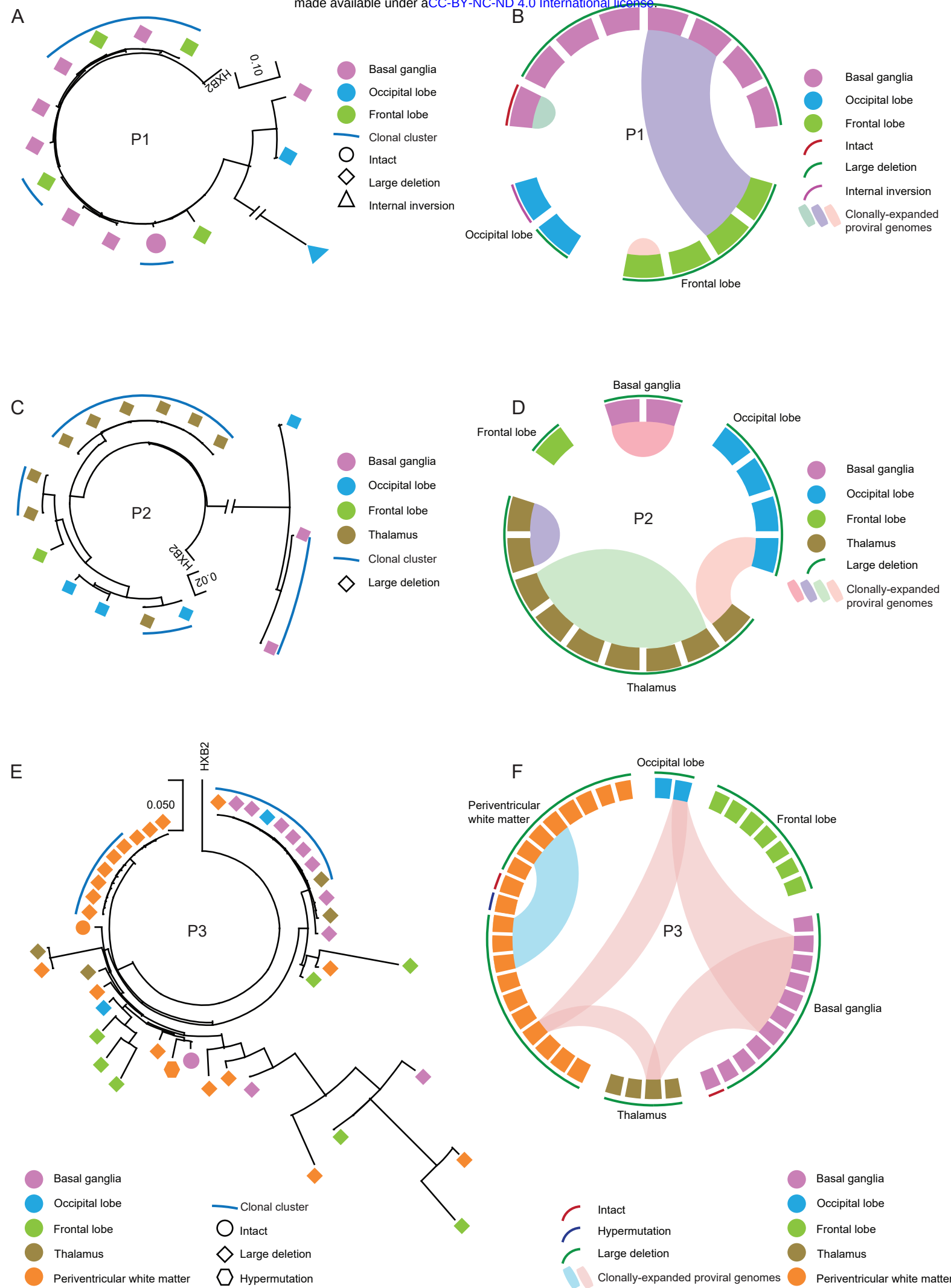


Figure 5