HHS Public Access

Author manuscript

J Neurovirol. Author manuscript; available in PMC 2020 June 09.

Published in final edited form as:

J Neurovirol. 2018 August ; 24(4): 439–453. doi:10.1007/s13365-018-0633-5.

Ultradeep single-molecule real-time sequencing of HIV envelope reveals complete compartmentalization of highly macrophage-tropic R5 proviral variants in brain and CXCR4-using variants in immune and peripheral tissues

Robin L. Brese[1], **Maria Paz Gonzalez-Perez**[1], **Matthew Koch**[1], **Olivia O'Connell**[1], **Katherine Luzuriaga**[1], **Mohan Somasundaran**[1], **Paul R. Clapham**[1], **James Jarad Dollar**[2], **David J Nolan**[2], **Rebecca Rose**[2,*], **Susanna L. Lamers**[2]

[1]Program in Molecular Medicine, University of Massachusetts Medical School, Biotech 2, 373 Plantation Street, Worcester, Massachusetts 01605.

[2]718 Bayou Lane, Thibodaux, LA 70301.

Abstract

Despite combined anti-retroviral therapy (cART), HIV+ patients still develop neurological disorders, which may be due to persistent HIV infection and selective evolution in brain tissues. Single-molecule real-time (SMRT) sequencing technology offers an improved opportunity to study the relationship among HIV isolates in the brain and lymphoid tissues because it is capable of generating thousands of long sequence reads in a single run. Here, we used SMRT sequencing to generate ~50,000 high-quality full-length HIV envelope sequences (>2200bp) from 7 autopsy tissues from an HIV+/cART+ subject, including three brain and four non-brain sites. Sanger sequencing was used for comparison with SMRT data and to clone functional pseudoviruses for *in vitro* tropism assays. Phylogenetic analysis demonstrated that brain-derived HIV was compartmentalized from HIV outside the brain and that the variants from each of the three brain tissues grouped independently. Variants from all peripheral tissues were intermixed on the tree, but independent of the brain clades. Due to the large number of sequences, a clustering analysis at three similarity thresholds (99%, 99.5% and 99.9%) was also performed. All brain sequences clustered exclusive of any non-brain sequences at all thresholds; however frontal lobe sequences clustered independently of occipital and parietal lobes. Translated sequences revealed potentially functional differences between brain and non-brain sequences in the location of putative N-linked glycosylation sites (N-sites), V1 length, V3 charge and the number of V4 N-sites. All brain sequences were predicted to use the CCR5 co-receptor, while most non-brain sequences were predicted to use CXCR4 co-receptor. Tropism results were confirmed by *in vitro* infection assays. The study is the first to use a SMRT sequencing approach to study HIV compartmentalization in tissues and supports other reports of limited trafficking between brain and non-brain sequences during cART. Due to the long sequence length, we could observe changes along the entire

*Corresponding author: Rebecca Rose, Bioinfoexperts, LLC, 718 Bayou Ln, Thibodaux LA, 70301, rebecca.rose@bioinfox.com.

STATEMENT OF CONFLICT OF INTEREST

The authors declare they have no conflict of interest.

envelope gene, likely caused by differential selective pressure in the brain, that may contribute to neurological disease.

Keywords

evolution; phylogenetics; clustering; tropism; combined anti-retroviral therapy (cART); next-generation sequencing (NGS)

INTRODUCTION

Before the advent of combined antiretroviral therapy (cART), HIV-infected subjects frequently suffered from severe HIV-associated dementia (HAD). After cART was introduced in the mid-1990s, the incidence of HAD dramatically decreased; however, milder forms of HIV-associated neurocognitive disorders (HAND) became prevalent. HAND includes a spectrum of symptoms from asymptomatic neurocognitive impairment, to minor neurocognitive disorder, as well as HAD (Antinori *et al.*, 2007; Clifford, 2017; Liner *et al.*, 2010; McArthur *et al.*, 2010). HIV-1 infects the brain early after infection (Davis *et al.*, 1992); however, viral DNA is difficult to detect in the brain during the asymptomatic phase (Bell *et al.*, 1993; Donaldson *et al.*, 1994; McCrossan *et al.*, 2006) and it is unclear how HIV in the brain responds to therapy. HAND may result from the replication or reactivation of HIV already present in brain tissue or from newly introduced virus brought in by HIV infected monocytes/macrophages (Buckner *et al.*, 2011; Burdo *et al.*, 2010; Gartner, 2000; Liu *et al.*, 2000; Nottet and Gendelman, 1995) or by T-cells (Schnell *et al.*, 2011) migrating across the blood brain barrier.

HIV-1 sequences in brain are usually highly compartmentalized from those outside the CNS, including immune tissue (Burkala *et al.*, 2005; Caragounis *et al.*, 2008; Chang *et al.*, 1998; Chen *et al.*, 2000; Gatanaga *et al.*, 1999; Gonzalez-Perez *et al.*, 2012; Haggerty and Stevenson, 1991; Hughes *et al.*, 1997; Korber *et al.*, 1994; Lamers *et al.*, 2011; Ohagen *et al.*, 2003; Pillai *et al.*, 2006; Ritola *et al.*, 2005; Salemi *et al.*, 2005; Shapshak *et al.*, 1999; Smit *et al.*, 2004; Strain *et al.*, 2005; van't Wout *et al.*, 1998; Wong *et al.*, 1997). This compartmentalization is consistent with the evolution of a brain-specific viral population. In the brain, HIV is mainly detected in the predominant CD4+ immune cells found in that tissue, perivascular macrophages and microglia (Cosenza *et al.*, 2002; Fischer-Smith *et al.*, 2004; Fischer-Smith *et al.*, 2001; Glass *et al.*, 1995; Gonzalez-Scarano and Martin-Garcia, 2005; Lane *et al.*, 1996; Takahashi *et al.*, 1996; Williams *et al.*, 2001). This HIV population consists of highly macrophage-tropic viruses, which utilize the CCR5 co-receptor for viral entry (Dunfee *et al.*, 2009; Gonzalez-Perez *et al.*, 2012; Peters *et al.*, 2004; Peters *et al.*, 2006; Thomas *et al.*, 2007) and has evolved to exploit the low levels of CD4 on myeloid lineage cells for infection (Gonzalez-Perez *et al.*, 2012; Peters *et al.*, 2004; Peters *et al.*, 2006; Thomas *et al.*, 2007). In contrast, the vast majority of HIV sequences amplified from blood or immune tissue of most subjects also utilize CCR5, but are typically T-cell tropic and do not efficiently infect macrophages (Gonzalez-Perez *et al.*, 2012; Gonzalez-Perez *et al.*, 2017; Peters *et al.*, 2004; Peters *et al.*, 2006; Schnell *et al.*, 2011).

Single genome analysis (SGA) represents the principal approach used so far to investigate HIV envelope (*env)* because it reduces the chance of experimentally re-sampling the same specific HIV sequences amplified during PCR, as well as eliminating PCR-mediated recombination (Keele *et al.*, 2008; Nolan *et al.*, 2017; Simmonds *et al.*, 1990). However, although SGA provides an opportunity to investigate *env* variability, it is restrictive because only a small number of variants are feasibly generated from each sample. Next generation sequencing (NGS) approaches, such as Illumina, can be used to investigate large numbers of sequences generated from a single sample; however, this approach is problematic with HIV because millions of short sequences are generated, which subsequently require assembly, a near impossible feat with HIV *env* due to its high sequence variability combined with the error rate of NGS. A newer technology, Single Molecule, Real-Time (SMRT) DNA sequencing generates a much longer read-length and can produce thousands of high-quality consensus sequences through repeated sequencing of circularized templates (Travers *et al.*, 2010). These SMRT data also come with post-processing difficulties, such as insertions, deletions, reamplification, and sequence errors; however, these hurdles can be overcome using appropriate data filtering approaches (Laird Smith *et al.*, 2016).

In this study, we used SMRT, SGA and *in vitro* co-receptor and macrophage assays to assess the selective effects of different anatomical tissue sites on the variation of HIV *env* proviral sequences for a cART+ HIV+ subject who died from AIDS. Tissues examined included three brain (frontal, occipital and parietal lobes), immune (lymph node), lung and colon tissues as well as blood cells. The study enabled the investigation of tens of thousands of full-length *env* sequences amplified from different tissues. To our knowledge, this is the first study to apply the SMRT approach to anatomical tissue-based HIV. This study provides novel and in-depth insights into (1) the comparative variation of *env* sequences in brain and lymphoid tissues, (2) compartmentalization of HIV variants inside and outside of brain tissue and (3) the receptor and tropism properties of Envs obtained from different sites.

MATERIALS AND METHODS

Subject and tissues

Post-mortem tissues from subject 123, a 43-year-old Caucasian male, were provided by the National Disease Research Interchange (NDRI), without patient identifying information. The University of Massachusetts Medical School IRB considered that this research was not human subjects research as defined by DHHS and FDA regulations (IRB ID: H00014098). Subject 123 was on drug therapy (Tenofovir, Ritonavir, Abacavir, Tipranavir, Raltegravir, Etravirine) and had a well-controlled plasma viral load (<75 copies/ml^3) at the last test, 7 months before death. The cause of death was cardio-pulmonary arrest and was diagnosed with HAD. Post-mortem plasma and serum samples were found to have undetectable viral copy number by the Infectious Disease Laboratory at the Children's Hospital of Chicago using Abbott RealTime HIV-1 RNA PCR. Tissues collected at autopsy included frontal lobe, occipital lobe, parietal lobe, lymph node, lung, colon and blood cells.

DNA isolation

Genomic DNA was harvested from tissues (25 mg of tissue per preparation) and cells pelleted from approximately 1 ml of plasma using the QIAamp DNA Mini Kit (QIAGEN) as described by the manufacturer's protocol. DNA was eluted in 300 μl nuclease free, PCR grade water and stored at −80°C.

PCR of env from tissue DNA, library preparation and SMRT runs

Full length *envs* were amplified for SMRT NGS. A nested PCR approach was used to amplify a 2.6kb product from tissue DNA. Outer primers RevenvA (bp 5853–5877 relative to HXB2; TAG AGC CCT GGA AGC ATC CAG GAA G) and EnvN (bp 9171–9145; CTG CCA ATC AGG GAA GTA GCC TTG TGT) and inner primers PacBioF-long (bp 6207–6234; GAG CAG AAG ACA GTG GCA ATG AGA GTG A) and PacBioR-long (bp 8815–8788; TTG ACC ACT TGC CAC CCA TCT TAT AGC A) were used. The PCR protocol was as follows; 95°C for 1 min, then 12 cycles of 95°C for 30s, 68°C for 3 mins, 70°C for 10 min. A second PCR round of 35 cycles was then undertaken using the same conditions. The Advantage 2 PCR Kit (Clontech Inc.) was used as described in the manufacturer's protocol for 50 μl reactions. PCR products were electrophoresed on a 2% E-Gel EX Agarose Gel (ThermoFisher Scientific Inc.), bands excised and subsequently purified with a QIAquick Gel Extraction Kit (QIAGEN). PCR products were eluted in 50 μl nuclease free, PCR grade water. Samples were sent to the Deep Sequencing Core Labs at the University of Massachusetts Medical School for sequencing using the PacBio RS II instrument. One SMRT Cell was utilized for each tissue with 6-hour collection times.

SMRT sequence processing

SMRT raw reads for each tissue were quality filtered at 99% quality (using SMRT Pipeline) to remove low-quality sequences. The quality filtered reads were further selected to include only circular consensus sequences >2000 base pairs with at least 10 complete passes using PacBioSmartPipe tools (v.2.3.0). For each tissue-specific data file, reads were assembled to a single end-point dilution tissue-specific reference sequence using BLASR (v.2.3.0). This alignment was gap stripped to remove spurious inserts to retain columns with ≥95% coverage using Geneious software version 9.0 (http://www.geneious.com/). Short sequences (<2000bp) were removed from the alignment. The final alignment contained 52,725 reads and was designated as "Aligned" (Table 1). As a large number of single nucleotide deletions remained in the resulting assembly, different filtering methods for gap-handling were used for phylogenetic analysis and for the analysis of translated data as described below.

Generation of High Quality Consensus Reads (HQCRs) for phylogenetic and clustering analysis

The generation of HQCRs corrects for sequencing errors that may have resulted due to the SMRT sequencing technology, and especially template reamplification. For the generation of HQCRs, we generally followed the methods described in Laird Smith *et al* (Laird Smith *et al.*, 2016). First, ambiguous and difficult to align regions incorporating the V1 and V4 hypervariable domains were removed because phylogenetic analysis assumes homology at each position. Next, an alignment of *Env* reads for each subject data set was generated in

MAFFT (Katoh and Standley, 2013) and manually optimized using AliView (1.17.1). USEARCH v10.0.24 was used to generate HQCSs on each tissue data set and to filter out duplicate reads. Non-identical reads were subsequently combined into USEARCH clusters at 99% genetic identity. Each cluster was then assigned a centroid and consensus sequence (50% majority rule). The number of centroid sequences in each tissue/patient is noted in Table 1 as “HQCS”. We then combined the HQCS variants from each tissue back into a single alignment for the subject (n=11,773). To further reduce the data set for phylogenetic studies, we condensed the HQCS and only included variants that represented >10 sequences from the original alignment (HQCS10, Table 1).

Phylogenetic analysis

Both the HQCS and the HQCS10 alignments were used for phylogenetic analysis (Figure 1). Trees from the HQCS variant alignments were generated using the Fast-Tree maximum-likelihood approach with the Generalized Time Reversible (GTR) model in Geneious v9.0. Trees from the HQCS10 alignments were inferred using PhyMLv3.0 with bootstrap support.

Clustering

Distance-based clustering groups related sequences based on genetic similarity and (in some cases) phylogenetic branch support. Here we used HIV-TRACE (Wertheim *et al.*, 2014), which relies only on genetic distance, because it detects larger and fewer clusters and it can cluster large sequence files in a reasonable time period (Rose *et al.*, 2017). For this analysis, we used the HQCS for each tissue, for a total of n=11,772 sequences (Table 1). Pairwise distances for each HQCS variant alignment were generated using the tn93 model. Clustering was performed at three difference thresholds: 99%, 99.5%, and 99.9% similarity. Tableau (v10) was used to visualize the cluster data.

Translated sequence alignments

Amino acid variation within functional envelope variable domains (position of putative N-linked glycosylation sites (N-sites), length, number of putative N-Sites, co-receptor prediction and charge) for each tissue were calculated for all sequences for which all five full-length variable regions were available (Table 1). For this analysis, we started with the “Aligned” data. Next, the alignment was reduced to include only the variable domains. These sequences were translated using GeneCutter (www.lanl.gov) and any sequences that contained stop codons or deletions were removed. The final alignment contained 27,100 sequences. Co-receptor usage for the V3 was predicted using WebPSSM x4r5 algorithm (Jensen *et al.*, 2006). The positions and number of predicted N-linked glycosylation sites (according to HXB2 numbering) were calculated with the N-Glycosite tool at the HIV Database at Los Alamos (hiv.lanl.gov). Variable region length, was calculated using the Variable Region Characteristics tool at the HIV Database at Los Alamos (hiv.lanl.gov). Results were viewed in Tableau (v10).

Nested PCR endpoint cloning of *env* genes for SGA-Sanger investigation

Full length *env* genes were amplified by nested PCR using the endpoint dilution approach from proviral DNA from frontal lobe, occipital lobe, parietal lobe, colon, lung, lymph node

and blood cells. Nested PCRs were performed using limiting dilution amplification of HIV-1 proviral DNA. For the first PCR, outer primers RevenvA and EnvN were used (described above). Inner primers RevenvB_TOPO (bp 5954–5984; CAC CTA GGC ATC TCC TAT GGC AGG AAG AAG) and Env-lo (bp 9096–9063; GTT TCT TCC AGT CCC CCC TTT TCT TTT AAA AAG) were then used for the second round PCR. Conditions for both rounds of PCR were as follows: 98°C for 30s, then 40 cycles of 98°C for 10s, 70°C for 30s, 72°1.5 min, followed by 72° for 7 min. The Phusion High-Fidelity DNA Polymerase Kit (New England BioLabs) was used according to the manufacturer's protocol. DNA was serially diluted to determine the concentration at which less than 30% of the reactions were positive. At this concentration, there is a >80% chance of PCR products being derived from single genomes so that the generation of PCR generated recombinants during the amplification can be avoided (Salazar-Gonzalez *et al.*, 2008).

Cloning of SGA PCR products

The PCRs were run out on a 1.0% agarose gel with 0.8% crystal violet in 1XTAE buffer to determine which dilutions resulted in fewer than 30% of the reactions as positive. DNA was purified using the QIAquick Gel Extraction Kit (QIAGEN). Eluted PCR products were ligated into pcDNA 3.1D/V5-His-TOPO using TA Expression Kit (Invitrogen) and transformed into TOP10 cells (Invitrogen Inc.). Each positive clone was sequenced in both directions (Genewiz Inc.).

Pseudovirus production

Pseudoviruses were constructed using *env*+ pcDNA 3.1D/V5-His-TOPO plasmids, carrying patient derived *env* genes, and an *env*- pNL4.3Δenv plasmid, an HIV-1 clone with a premature stop codon in the *env* gene as described previously (Gonzalez-Perez *et al.*, 2012; Peters *et al.*, 2004).

Env functionality studies

HeLa TZM-bl were used to determine Env functionality using Env+ pseudoviruses. For each infectivity assay, cells were diluted to 4.0×10^4 cells/ml and 500 μl/well were added to 48-well plates the day prior to infection. Mac-tropic R5 B33 and JRFL, and non-mac-tropic R5 JRCSF Env+ pseudoviruses were used as controls. Cells were infected by removing media and adding 100 μl of pseudovirus dilution. Plates were incubated for 3 hours at 37°C and then 400 μl of media were added to each well before incubating at 37°C for 48 hours. HeLa TZM-bl cells carry a β-galactosidase reporter gene for HIV infection. They were fixed with 0.5% glutaraldehyde in phosphate-buffered saline (PBS) rinsing twice in 1 × PBS and stained by adding 500 μl X-gal substrate (0.5 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, 3 mM potassium ferrycyanide, 3 mM potassium ferrocyanide, 1 mM magnesium chloride). Plates were incubated at room temperature overnight and stained cells counted to determine the focus forming units (FFU)/ml of each pseudovirus stock.

To evaluate macrophage infection, macrophages were prepared from blood monocytes as described previously (Gonzalez-Perez *et al.*, 2012; Peters *et al.*, 2004). 500 μl of macrophages at 2.5×10^5 cells/ml were plated in 48-well plates the day prior to infection. On the day of infection, media was removed from cells and 100 μl of 10 μg/ml DEAE

dextran were added to each well and plates were incubated at 37°C for 30 minutes. Following incubation, 100 μl of serially diluted pseudovirus stocks, described above, were added to each well. B33 and JRFL Env+ pseudoviruses were used as mac-tropic controls and JRCSF, as a non-mac-tropic control. Macrophages were spinoculated for 45 minutes at 1200 rpm in a benchtop centrifuge, incubated for 3 hours at 37°C, and then 300 μl media were added to each well. Plates were incubated for one week, adding additional media as needed. Macrophages were fixed in cold 1:1 methanol:acetone, washed, and immunostained for p24 using monoclonal antibodies 365 and 366 (NIBSC Centre for AIDS Reagents). Stained, infected cells were counted using phase contrast microscopy

NP2/CD4/CCR5 and NP2/CD4/CXCR4 cells (Soda *et al.*, 1999) were used to test co-receptor specificity of Envs using Env+ pseudoviruses carrying a GFP reporter gene. These Env+ pseudoviruses were prepared by cotransfection of 293T cells with an *Env+* vector, *env* − minus pNL4.3 and a GFP reporter vector, pHIVec2/GFP (Hofmann *et al.*, 1999) at a ratio of 1:1:1.5. The cell supernatants containing Env+ pseudovirions were harvested 48 h post-transfection. NP2/CD4/CCR5 and NP2/CD4/CXCR4 cells were plated the day before use in a 48-well tray at 3×10^4 cells/ml in 500 μl of growth medium. Infected cells were estimated by GFP expression using fluorescent microscopy two days post-infection.

RESULTS

Phylogenetic analysis

First, a phylogenetic tree was inferred using the HQCS variants, including the SGA-Sanger *env* sequences that were amplified from serially diluted DNA prepared from each tissue (Figure 1a). However, since the vast majority of HQCS sequences represented <10 clones, for computational reduction and clarity of the tree, we used a “reduced” dataset that included only HQCS representing >=10 aligned reads (HQCS10 Table 1, Figure 1b). Both trees clearly show that brain-derived viruses were compartmentalized from virus in tissues outside the brain with high branch support. The longest internal branch length in the tree lead to the brain variants, indicating the greatest divergence. Furthermore, the variants from each of the three brain tissues also grouped exclusively. Variants from all peripheral tissues were dispersed on the tree, exclusive of the brain sequence clades. This pattern largely held in both trees, although in the HQCS tree, the occipital reads were descendant from the parietal clade. Lymph node, blood, lung and colon sequences were interspersed among the sequences from these tissues. Importantly, the sequences of the SGA-Sanger, brain-derived clones segregated with SMRT sequences from each of the same brain tissues, demonstrating that the SGA clones are genuinely representative of *env*s from each tissue for this subject. Note that due to the size of the dataset, the low number of colon sequences are obscured in the full tree (Figure 1a).

Prediction of env co-receptor use

Co-receptor usage for each *env* based on the V3 loop amino acid sequence was predicted using WebPSSM x4r5 algorithm (Jensen *et al.*, 2006). All *envs* derived from brain tissue were predicted to be R5, while the vast majority of *envs* from outside brain were classified as X4 (Figure 1c). However, a segregated group of R5 *envs* were present in the colon (27%

of all colon sequences, although note the low number of reads), while a small number of R5 *envs* present in colon, lung and lymph node formed a group close to the branch leading to *envs* present in brain tissue. This is most clearly seen in the HQCS10 tree (Figure 1d), since the majority X4 reads in the full tree obscure these branches.

Distance-based clustering

The composition of the clusters at different clustering thresholds assist in understanding the genetic relatedness of many sequences (>10,000) from different tissues that can be difficult to interpret in a phylogenetic format. As one would expect, fewer HQCS sequences were in clusters as the specified thresholds of 99%, 99.5%, and 99.9% similarity score was increased (Figure 2a–d). We report the following results based on the number of non-identical reads that each HQCS represented. At the 99% threshold, 18 clusters were found, of which 7 had >100 sequences. The largest cluster contained all of the occipital and parietal sequences. The second largest cluster contained sequences from all four peripheral tissues: lymph node/lung/colon/blood cells. The third largest contained only frontal lobe sequences. Lung sequences comprised the other clusters with >100 variants, with one containing blood sequences. At the 99.5% threshold, 73 clusters were found, of which 15 contained >100 reads. The largest three clusters resembled those at the 99% level. A small number (n=293) occipital sequences were in their own cluster. The rest of the clusters contained either lymph or lung exclusively, except for one which contained both lung and colon sequences. At the 99.9% threshold, 213 clusters were detected, of which 29 contained >100 reads. The largest two clusters were now comprised of exclusively either parietal or occipital lobe sequences. Frontal lobe sequences comprised two clusters, again exclusive of other peripheral sequences. Only 2 of the 29 clusters contained sequences from multiple peripheral tissues. While sequences from brain and periphery were highly segregated from each other, sequences from different brain compartments were also segregated, although less so between the occipital and parietal lobes. Sequences from the non-brain tissues still showed some degree of segregation, although far less so than in the brain.

Amino acid characteristics of variable domains

A total of 27,100 reads translated into amino acids and were analyzed for amino acid variation (Table 1). We investigated the position of N-Sites along Env (Figure 3). In the V1 domain, most sequences contained 4–5 putative N-Sites within HXB2 coordinates 130–190; however, the position of these varied among the different non-brain tissues, whereas all positions were identical among the three brain tissues, suggesting a selective process or founder effect for V1 glycosylation in brain. Outside of V1, the positions of the N-Sites were similar with few exceptions: at HXB2 coordinate 362, lymph lode envs shared this site in most sequences with occipital and parietal lobes, although it was absent in other tissues; at HXB2 coordinate 448, nearly all brain sequences contained a site that was absent in all non-brain tissues, at position 461, the colon was missing a glycosylation motif that was present in nearly all other sequences; and at position 816 a glycosylation motif occurred in 100% of the frontal lobe sequences that was absent in all other tissues. In terms of the number of glycosylation motifs and the length of the variable regions, overall there were few differences between the brain and peripheral tissues (Figure 4). The majority of brain sequences had fewer glycosylation sites in V4 domain (sites=4) than the majority of non-

brain sequences (sites=5). The majority of brain sequences also had longer V1 domain sequences (length=31) than non-brain (length=26).

The pie charts (Figure 4) demonstrated some interesting patterns between variable domains. For example, we were interested in whether minority variants in either the number of N-Sites or the length correlated across variable regions. To test this hypothesis, we used a chi-squared test to determine significance in cases where at least 4 pies contained unique combinations within a single tissue and occurred in at least 10 sequences (Table 2). There was significant correlation of minority variants in the number of glycosylation sites in lung (wherein V2 minor variants were associated with V4 minor variants) and frontal lobe (wherein V1 minor variants were associated with V4 minor variants). Interestingly, while the number of minor variants in parietal lobe glycosylation sites in V1 and V2 were nearly identical, the minority variants were *not* significantly associated with each other; i.e. the majority variant for V1 segregated with the minority variant for V2, and vice versa. A similar pattern was observed in the colon for length, where in two cases, the minority variant in one region was *only* found with a majority variant of the other (V2:V4, and V4:V5). This suggests that particular combinations of variable regions may be functionally selected.

Functionality and tropism of SGA-Sanger Envs expressed on pseudovirions

Similar to SMRT sequences, all cloned *envs* from brain tissues were predicted CCR5-tropic, while all of the *envs* derived from tissues outside brain were CXCR4-tropic. Env+ pseudoviruses were prepared to confirm CXCR4-use, and to investigate macrophage-tropism. We first tested whether Envs were functional by evaluating whether they conferred efficient infection of HeLa TZM-bl cells (Figure 5). Overall, 24 Envs yielded Env+ pseudoviruses that conferred infectivity titers of $>10^3$ FFU/ml on HeLa TZM-bl cells and were considered sufficiently functional for further study (Table 3, Suppl. Table 1). Pseudoviruses carrying functional Envs were then tested for macrophage infection using cells prepared from at least two different donors. Pseudoviruses carrying Envs derived from the brain tissues conferred mainly high levels of infectivity for macrophages (Figure 5). In contrast, pseudoviruses carrying Envs from lymph node, colon or lung mediated modest infection of macrophages at best but were mostly non-mac-tropic.

We next sought to confirm the use of CXCR4 for infection for the peripheral-tissue derived Envs identified as X4 by using WebPSSM. Our results showed most Envs identified as X4 by PSSM infected both CCR5+ and CXCR4+ cells, indicating that they were actually R5X4 (Figure 6). Curiously, three Envs (two from lung, one from colon) conferred modest titers on NP2/CD4/CCR5 cells but only background infectivity on NP2/CD4/CXCR4. These Envs were therefore R5 despite scoring as X4 by WebPSSM (Figure 6). For controls, we tested the occipital lobe-derived Env (19–34-2) from this subject as well as R5 B33, JRFL and JRCSF Envs. These Envs only infected CCR5+ NP2/CD4 cells, as expected. Control CXCR4-using NL4.3 mediated high levels of infection via CXCR4 as well as low infection via CCR5. R5X4 89.6 conferred high infection for both CCR5 and CXCR4-expressing NP2/CD4 cells. Together, these data confirm the predominance of CXCR4-using Envs outside brain.

DISCUSSION

Various selection pressures in different tissues and environments influence the properties and characteristics of HIV-1 Env glycoproteins. Previously, we and others, reported that HIV-1 R5 Envs from brain tissue of subjects with neuroAIDS were highly mac-tropic (Dunfee *et al.*, 2009; Gonzalez-Perez *et al.*, 2012; Peters *et al.*, 2004; Peters *et al.*, 2006; Thomas *et al.*, 2007), while R5 Envs from outside brain were generally not (Gonzalez-Perez *et al.*, 2012; Gonzalez-Perez *et al.*, 2017; Peters *et al.*, 2004; Peters *et al.*, 2006; Schnell *et al.*, 2011). In addition, we reported that highly macrophage-tropic R5 virus from brain tissue carried gp120s with a lower positive charge compared to those in immune tissue (Gonzalez-Perez *et al.*, 2012). Others have reported that CCR5-using Envs with shorter variable loops or fewer N-linked glycosylation sites are preferentially transmitted (Chohan *et al.*, 2005; Derdeyn *et al.*, 2004; Derdeyn and Hunter, 2008; Gnanakaran *et al.*, 2011; Gray *et al.*, 2007; Li *et al.*, 2006; Liu *et al.*, 2008; Sagar *et al.*, 2009; Wu *et al.*, 2006; Zhang *et al.*, 2010). These differences in Env characteristics may result from varying selection pressures exerted in different tissues during viral replication and transmission, as well as reduced immune pressures in the immune privileged environment of the brain.

In this study, a SMRT sequencing approach generated tens of thousands of full-length *env* sequences from a multisite autopsy of an HIV+ person who was cART suppressed at least seven months prior to death. The handling of the large number of SMRT sequences largely followed a previous study (Laird Smith *et al.*, 2016), which filtered data based on quality, alignment, and % identity to ultimately reduce millions of reads to a high-quality data set for subsequent analysis. While some resampling can occur using this approach, the important characteristic of the technology is that it can potentially identify all the diversity in each tissue. SMRT uses a circular consensus sequencing approach, whereby the same gene region is repeatedly sequenced. While the error rate for a single pass is quite high (>10%), the consensus of 10 or more passes removes most of the spurious base-calls (Korlach, 2017). Furthermore, there is no bias as to where errors occur, and indels are the major type of error in single-pass reads (Weirather *et al.*, 2017), Therefore, removing indels and taking the consensus sequence can bring down error rates to as low as 0.001% (Ferrarini *et al.*, 2013). Here, we further increased the robustness of the approach by using HQCS sequences, which are grouped at 99% identity, thus eliminating the vast majority of sequencing errors. A major advantage of SMRT is the long reads, which allows for accurate analysis of highly variable regions in HIV (Huang *et al.*, 2016). Other platforms, e.g. Illumina, produce very short (<300bp) reads, which must then be assembled into contigs. However, the high variation substantially increases the uncertainty in the reconstruction, and therefore linkage within long stretches cannot be known with confidence (McElroy *et al.*, 2014). Since accurate reconstruction is critical for phylogenetics, we opted to use the SMRT approach, permitting an in-depth computational genetic analysis that confirmed brain compartmentalization at several levels, including phylogeny, clustering, tropism, preservation or selection for glycosylation and length in hypervariable *env* domains. SGA *env* sequences were incorporated into phylogenetic analyses to confirm the reliability of SMRT sequences. SGA sequences were also used in *in vitro* functional tropism assays in order to validate

computational tropism predictions showing that brain-derived virus utilized CCR5 and non-brain virus mainly utilized CXCR4 (with the majority R5X4 or dual tropic) for viral entry.

Compartmentalization of HIV in brain tissue has been reported for most (but not all) subjects with neuro-disease (Burkala *et al.*, 2005; Caragounis *et al.*, 2008; Chang *et al.*, 1998; Chen *et al.*, 2000; Gatanaga *et al.*, 1999; Gonzalez-Perez *et al.*, 2012; Haggerty and Stevenson, 1991; Hughes *et al.*, 1997; Korber *et al.*, 1994; Lamers *et al.*, 2011; Ohagen *et al.*, 2003; Pillai *et al.*, 2006; Ritola *et al.*, 2005; Salemi *et al.*, 2005; Shapshak *et al.*, 1999; Smit *et al.*, 2004; Strain *et al.*, 2005; van't Wout *et al.*, 1998; Wong *et al.*, 1997). In contrast to the compartmentalization between *envs* present in brain and in non-brain tissues, the lack of clear compartmentalization of *env* sequences observed between lymph node, lung and colon is consistent with the traffic of virions or infected cells between these tissues. In our study, relatedness among *env* sequences was first assessed using phylogenetic and clustering approaches. HIV compartmentalization within the brain and among brain tissues was undoubtedly seen in the tree containing all HQCS; however, with so much data in a single figure, it remained difficult to view the degree of variation within brain sequences or the location of SGA clones used as controls (Figure 1a). Once the data was reduced to HQCS10 sequences, the subtle variation in the brain sequence populations was clear, confirming that SMRT did not simply re-amplify thousands of sequences that were derived from a single or very few proviruses, but likely reflects the true diversity in the tissue (Figure 1b). Conversely, much more variation was observed in non-brain sequences, which supports that the SMRT sequencing approach could efficiently capture the sequence population in each tissue sample. When the HQCS were clustered using HIV-TRACE, another interesting phenomenon was observed, that is, despite frontal lobe sequences significantly branching with parietal and occipital sequences, they remained somewhat unique. And, in fact, exclusive features of frontal lobe sequences were observed in the location of glycosylation sites. For example, exclusive of sequences from occipital and parietal lobe, frontal lobe sequences were lacking an N-Site at position 362 and had an additional N-site at position 816; the two differences occurred in *env* conserved domains. Brain HIV compartmentalization has been observed in other studies; however, these analyses typically incorporate much smaller data sets. Here we showed that SMRT deep sequencing enabled the observation of complete compartmentalization of brain HIV, and even compartmentalization between different brain locations. Remarkably, CXCR4-using variants (which form the majority of variants outside the brain) were completely excluded from brain tissue.

The non-brain sequences from our subject were, for the most part, mixed in the phylogenetic tree. A small subset of *env* sequences from colon, lung and lymph node were the most intermediate to brain HIV populations, suggesting that there may be some adaptive qualities in sequences from these tissues that are shared with brain, such as the evolution an enhanced Env affinity for CD4 or other adaptations that predispose HIV for infection of macrophages. This study highlights the need for therapy that targets tissue-based HIV.

Env glycosylation plays a vital role in viral evasion from the host immune response by masking key neutralization epitopes (Pikora, 2004). In V1, all putative N-sites were conserved in brain but differed from non-brain sequences. A similar pattern was observed in

V4; however, the overall brain N-site pattern in V4 matched those from lung sequences. In general, the number of N-sites across Env were stable in brain tissues; however, very small populations of virus with different number of N-sites occurred in frontal lobe V1 and V4 and parietal lobe V1 and V2 domains. We assessed the possibility of co-evolution occurring between variable domains and observed that minor variants from the frontal lobe coevolved but the opposite was true for parietal lobe (Table 2). Interestingly, there was almost no N-site variation in either the location or number in all 9906 occipital-derived sequences, suggesting a fully adjusted population of brain sequences or reduced HIV replication. Overall, brain N-site patterns, in both number and locations, were much more conserved than lymphoid-derived tissues, where every tissue had a minor variant population in at least one variable domain. This again suggests that sequence adaptation in brain is non-random, associated with functional Env elements and likely dependent on structural, biochemical and immune characteristics of brain tissue.

Length variation was completely conserved in brain variable domains V1-V4, whereas a moderate amount of length variation was observed in all non-brain variable domains, especially in lung and colon sequences. Blood sequences were the most conserved in terms of length and N-sites, with N-site variation only observed in V1 and no length variation occurring in any variable domain. Since the patient was virally suppressed until at least 7 months pre-mortem, we may have observed only relatively newer populations of HIV from blood reservoirs (i.e. circulating resting T-cells), whereas in tissues, we may have observed recently integrated HIV combined with residual historical virus that was never fully cleared by cART. The finding emphasizes that HIV variation in blood does not necessarily reflect persistent HIV in tissues.

One of the most striking observations in this study was that using WebPSSM to investigate all SMRT sequences revealed that R5 HIV was ubiquitous in brain, whereas nearly all non-brain HIV were CXCR4-using. Because sequence based co-receptor determinations can over estimate CXCR4-using variants (Mulinge *et al.*, 2013; Tsai *et al.*, 2015), we also performed in vitro tropism assays on select SGA tissue envs. We noted that several Envs classified as CXCR4 by PSSM failed to confer infection via CXCR4 even though modest infection via CCR5 was observed. While CXCR4-using variants were widespread in tissues outside the brain, *in vitro* assays identified these as dual tropic isolates (R5X4), so therefore these could exploit CCR5 (the predominant co-receptor on macrophages) in addition to CXCR4. The complete exclusion of the CXCR4-using variants from brain tissue for this subject is a significant observation. Although CXCR4-using HIV-1 variants are infrequently detected in brain tissue, their presence there has been reported. For example, Gorry *et al.* reported the presence of mac-tropic R5X4 variants in a HAD subject (Gorry *et al.*, 2001; Mefford *et al.*, 2008), while Yi *et al.* described a mac-tropic variant from CNS that exclusively used CXCR4 to infect macrophages (Yi *et al.*, 2003). The CXCR4-using Envs described here did not confer macrophage infection. However, this is usual for Envs (including R5 Envs) derived from HIV outside brain tissue (Gonzalez-Perez *et al.*, 2012; Gonzalez-Perez *et al.*, 2017; Peters *et al.*, 2006; Sturdevant *et al.*, 2015). The selective pressures that lead to the preferential exclusion of CXCR4-using variants (over R5) from brain tissue are not known. It is of course possible that CXCR4-using Envs present at a low frequency in the tissues were not picked up in this experiment. However, we saw the same pattern in three different

tissues, which lends robustness to our observation. Furthermore, variants present at below-detection frequencies are unlikely to have an impact on the population as a whole in the absence of an introduction of a new selective pressure (e.g. a new drug regimen which could select for drug-resistant variants).

SMRT sequencing detected some R5 *envs* in the colon of this subject. These R5 variants formed a cluster more closely related to CXCR4-using variants compared to R5 *envs* in the brain and may be derived from precursor R5 populations that seeded the generation of CXCR4-using variants. However, a subset of CXCR4-using variants in colon were more closely related to *envs* in brain and this is consistent with the emergence of that CXCR4-using variants in multiple different lineages (Archer *et al.*, 2010; Poon *et al.*, 2012; Rozera *et al.*, 2012; Sede *et al.*, 2014; Zhou *et al.*, 2016).

The environmental pressures *in vivo* that select for macrophage-tropic R5 variants in brain tissue and non-macrophage-tropic variants in the periphery have not been clearly defined. The observed compartmentalization and genomic conservation of HIV-1 in brain tissue are likely due to adaptive mechanisms for replication in macrophage lineage cells (Fischer-Smith *et al.*, 2004; Fischer-Smith *et al.*, 2001; Glass *et al.*, 1995; Gonzalez-Scarano and Martin-Garcia, 2005; Lane *et al.*, 1996; Takahashi *et al.*, 1996; Williams *et al.*, 2001) and microglia (Cosenza *et al.*, 2002; Fischer-Smith *et al.*, 2004; Fischer-Smith *et al.*, 2001), which represent the main reservoirs of HIV in brain tissue. The main genetic modification in brain appears to be an increased Env:CD4 affinity that allows mac-tropic variants to use low levels of CD4 in macrophages to gain entry (Dunfee *et al.*, 2006; O'Connell *et al.*, 2013; Peters *et al.*, 2008) and their enhanced sensitivity to sCD4 inhibition (O'Connell *et al.*, 2013; Peters *et al.*, 2008). Here, we may have observed these modifications as specific location of N-sites and constraints on length of variable domains; however, further testing of other brain tissues would assist in confirming this result. The enhanced Env:CD4 interaction can result from amino acid substitutions within (Dunfee *et al.*, 2006) or outside (Duenas-Decamp *et al.*, 2008; Duenas-Decamp *et al.*, 2009; Dunfee *et al.*, 2007) the CD4bs and could be due to a better Env:CD4 contact, increased exposure of the CD4bs on Env and/or more efficient triggering of the trimer to open and induce fusion and entry. It is likely that low levels of antibodies present in brain tissue play a role in the evolution of Envs with a more exposed CD4bs (Beauparlant *et al.*, 2017).

The selective pressures in tissues outside the brain that result in the predominance of non-mac-tropic Envs are also not clear. These Envs require high levels of CD4 to trigger entry into cells and are almost certainly restricted to replication in CD4+ T-cells that express substantially higher levels of CD4 compared to macrophages. It is possible that non-mac-tropic Envs that carry a protected CD4bs evolved in the presence of neutralizing antibodies. However, the possibility that such Envs evolve because they carry an advantage for entry and replication in CD4+ T-cells *in vivo* has not yet been ruled out (Beauparlant *et al.*, 2017).

Our findings are based on a single individual with HAD and may not be generalizable to patients without neuropathogies. To further confirm these results, additional subjects will need to be studied. However, although resources such as the AIDS/Cancer Specimen Resource (ACSR) can provide some such datasets, tissue collections such as this one with

well-documented clinical information and carefully preserved tissue samples are not common.

In summary, our data show tight compartmentalization of mac-tropism and *env* sequences in brain tissues compared to diverse tissues outside the brain. This compartmentalization included the complete exclusion of CXCR4-using variants from brain tissue. In addition, we did not observe clear segregation of sequences between the different tissues outside brain. Together, our data highlight how combined use of SGA-Sanger and SMRT NGS approaches can be used to provide in depth analyses of HIV-1 *env* sequences and tropism or other properties in different tissues of HIV+ subjects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Benjamin Murrell (University of California, San Diego) for early discussions on this project.

We thank Yvonne Edwards for help in managing PacBio sequencing data at UMASS Med.Sch.

REFERENCES

Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, Clifford DB, Cinque P, Epstein LG, Goodkin K, Gisslen M, Grant I, Heaton RK, Joseph J, Marder K, Marra CM, McArthur JC, Nunn M, Price RW, Pulliam L, Robertson KR, Sacktor N, Valcour V, Wojna VE (2007). Updated research nosology for HIV-associated neurocognitive disorders. Neurology 69: 1789–99. [PubMed: 17914061]

Archer J, Rambaut A, Taillon BE, Harrigan PR, Lewis M, Robertson DL (2010). The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time--an ultra-deep approach. PLoS Comput Biol 6: e1001022. [PubMed: 21187908]

Beauparlant D, Rusert P, Magnus C, Kadelka C, Weber J, Uhr T, Zagordi O, Oberle C, Duenas-Decamp MJ, Clapham PR, Metzner KJ, Gunthard HF, Trkola A (2017). Delineating CD4 dependency of HIV-1: Adaptation to infect low level CD4 expressing target cells widens cellular tropism but severely impacts on envelope functionality. PLoS Pathog 13: e1006255. [PubMed: 28264054]

Bell JE, Busuttil A, Ironside JW, Rebus S, Donaldson YK, Simmonds P, Peutherer JF (1993). Human immunodeficiency virus and the brain: investigation of virus load and neuropathologic changes in pre-AIDS subjects. J Infect Dis 168: 818–24. [PubMed: 8376829]

Buckner CM, Calderon TM, Willams DW, Belbin TJ, Berman JW (2011). Characterization of monocyte maturation/differentiation that facilitates their transmigration across the blood-brain barrier and infection by HIV: implications for NeuroAIDS. Cell Immunol 267: 109–23. [PubMed: 21292246]

Burdo TH, Soulas C, Orzechowski K, Button J, Krishnan A, Sugimoto C, Alvarez X, Kuroda MJ, Williams KC (2010). Increased monocyte turnover from bone marrow correlates with severity of SIV encephalitis and CD163 levels in plasma. PLoS Pathog 6: e1000842. [PubMed: 20419144]

Burkala EJ, He J, West JT, Wood C, Petito CK (2005). Compartmentalization of HIV-1 in the central nervous system: role of the choroid plexus. AIDS 19: 675–84. [PubMed: 15821393]

Caragounis EC, Gisslen M, Lindh M, Nordborg C, Westergren S, Hagberg L, Svennerholm B (2008). Comparison of HIV-1 pol and env sequences of blood, CSF, brain and spleen isolates collected ante-mortem and post-mortem. Acta Neurol Scand 117: 108–16. [PubMed: 18184346]

Chang J, Jozwiak R, Wang B, Ng T, Ge YC, Bolton W, Dwyer DE, Randle C, Osborn R, Cunningham AL, Saksena NK (1998). Unique HIV type 1 V3 region sequences derived from six different regions

of brain: region-specific evolution within host-determined quasispecies. AIDS Res Hum Retroviruses 14: 25–30. [PubMed: 9453248]

Chen H, Wood C, Petito CK (2000). Comparisons of HIV-1 viral sequences in brain, choroid plexus and spleen: potential role of choroid plexus in the pathogenesis of HIV encephalitis. J Neurovirol 6: 498–506. [PubMed: 11175322]

Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, Overbaugh J (2005). Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol 79: 6528–31. [PubMed: 15858037]

Clifford DB (2017). HIV-associated neurocognitive disorder. Curr Opin Infect Dis 30: 117–122. [PubMed: 27798498]

Cosenza MA, Zhao ML, Si Q, Lee SC (2002). Human brain parenchymal microglia express CD14 and CD45 and are productively infected by HIV-1 in HIV-1 encephalitis. Brain Pathol 12: 442–55. [PubMed: 12408230]

Davis LE, Hjelle BL, Miller VE, Palmer DL, Llewellyn AL, Merlin TL, Young SA, Mills RG, Wachsman W, Wiley CA (1992). Early viral brain invasion in iatrogenic human immunodeficiency virus infection. Neurology 42: 1736–1739. [PubMed: 1513462]

Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH, Shaw GM, Korber BT, Allen S, Hunter E (2004). Envelope-Constrained Neutralization-Sensitive HIV-1 After Heterosexual Transmission. Science 303: 2019–22. [PubMed: 15044802]

Derdeyn CA, Hunter E (2008). Viral characteristics of transmitted HIV. Curr Opin HIV AIDS 3: 16–21. [PubMed: 19372939]

Donaldson YK, Bell JE, Ironside JW, Brettle RP, Robertson JR, Busuttil A, Simmonds P (1994). Redistribution of HIV outside the lymphoid system with onset of AIDS. Lancet 343: 383–5. [PubMed: 7905551]

Duenas-Decamp MJ, Peters P, Burton D, Clapham PR (2008). Natural resistance of human immunodeficiency virus type 1 to the CD4bs antibody b12 conferred by a glycan and an arginine residue close to the CD4 binding loop. J Virol 82: 5807–14. [PubMed: 18385254]

Duenas-Decamp MJ, Peters PJ, Burton D, Clapham PR (2009). Determinants flanking the CD4 binding loop modulate macrophage tropism of human immunodeficiency virus type 1 R5 envelopes. J Virol 83: 2575–83. [PubMed: 19129457]

Dunfee RL, Thomas ER, Gabuzda D (2009). Enhanced macrophage tropism of HIV in brain and lymphoid tissues is associated with sensitivity to the broadly neutralizing CD4 binding site antibody b12. Retrovirology 6: 69. [PubMed: 19619305]

Dunfee RL, Thomas ER, Gorry PR, Wang J, Taylor J, Kunstman K, Wolinsky SM, Gabuzda D (2006). The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. Proc Natl Acad Sci U S A 103: 15160–5. [PubMed: 17015824]

Dunfee RL, Thomas ER, Wang J, Kunstman K, Wolinsky SM, Gabuzda D (2007). Loss of the N-linked glycosylation site at position 386 in the HIV envelope V4 region enhances macrophage tropism and is associated with dementia. Virology.

Ferrarini M, Moretto M, Ward JA, Šurbanovski N, Stevanović V, Giongo L, Viola R, Cavalieri D, Velasco R, Cestaro A, Sargent DJ (2013). BMC Genomics 14:670. [PubMed: 24083400]

Fischer-Smith T, Croul S, Adeniyi A, Rybicka K, Morgello S, Khalili K, Rappaport J (2004). Macrophage/microglial accumulation and proliferating cell nuclear antigen expression in the central nervous system in human immunodeficiency virus encephalopathy. Am J Pathol 164: 2089–99. [PubMed: 15161643]

Fischer-Smith T, Croul S, Sverstiuk AE, Capini C, L'Heureux D, Regulier EG, Richardson MW, Amini S, Morgello, Khalili K, Rappaport J (2001). CNS invasion by CD14+/CD16+ peripheral blood-derived monocytes in HIV dementia: perivascular accumulation and reservoir of HIV infection. J Neurovirol 7: 528–41. [PubMed: 11704885]

Gartner S (2000). HIV infection and dementia. Science 287: 602–4. [PubMed: 10691542]

Gatanaga H, Oka S, Ida S, Wakabayashi T, Shioda T, Iwamoto A (1999). Active HIV-1 redistribution and replication in the brain with HIV encephalitis. Arch Virol 144: 29–43. [PubMed: 10076507]

Glass JD, Fedor H, Wesselingh SL, McArthur JC (1995). Immunocytochemical quantitation of human immunodeficiency virus in the brain: correlations with dementia. Ann Neurol 38: 755–62. [PubMed: 7486867]

Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, Shen T, Gaschen B, Krishnamoorthy M, Li H, Decker JM, Salazar-Gonzalez JF, Wang S, Jiang C, Gao F, Swanstrom R, Anderson JA, Ping LH, Cohen MS, Markowitz M, Goepfert PA, Saag MS, Eron JJ, Hicks CB, Blattner WA, Tomaras GD, Asmal M, Letvin NL, Gilbert PB, Decamp AC, Magaret CA, Schief WR, Ban YE, Zhang M, Soderberg KA, Sodroski JG, Haynes BF, Shaw GM, Hahn BH, Korber B (2011). Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. PLoS Pathog 7: e1002209. [PubMed: 21980282]

Gonzalez-Perez MP, O'Connell O, Lin R, Sullivan WM, Bell J, Simmonds P, Clapham PR (2012). Independent evolution of macrophage-tropism and increased charge between HIV-1 R5 envelopes present in brain and immune tissue. Retrovirology 9: 20. [PubMed: 22420378]

Gonzalez-Perez MP, Peters PJ, O'Connell O, Silva N, Harbison C, Cummings Macri S, Kaliyaperumal S, Luzuriaga K, Clapham PR (2017). Identification of emerging mac-tropic HIV-1 R5 variants in brain tissue of patients without severe neurological complications. J Virol 91: In press.

Gonzalez-Scarano F, Martin-Garcia J (2005). The neuropathogenesis of AIDS. Nat Rev Immunol 5: 69–81. [PubMed: 15630430]

Gorry PR, Bristol G, Zack JA, Ritola K, Swanstrom R, Birch CJ, Bell JE, Bannert N, Crawford K, Wang H, Schols D, De Clercq E, Kunstman K, Wolinsky SM, Gabuzda D (2001). Macrophage tropism of human immunodeficiency virus type 1 isolates from brain and lymphoid tissues predicts neurotropism independent of coreceptor specificity. J Virol 75: 10073–89. [PubMed: 11581376]

Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, Treurnicht F, Mlisana K, Shaw GM, Karim SS, Williamson C, Morris L (2007). Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. J Virol 81: 6187–96. [PubMed: 17409164]

Haggerty S, Stevenson M (1991). Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-1-infected individuals. Viral Immunol 4: 123–31. [PubMed: 1684709]

Hofmann W, Schubert D, LaBonte J, Munson L, Gibson S, Scammell J, Ferrigno P, Sodroski J (1999). Species-specific, postentry barriers to primate immunodeficiency virus infection. J Virol 73: 10020–8. [PubMed: 10559316]

Huang DW, Raley C, Jiang MK, Zheng X, Liang D, Rehman MT, Highbarger HC, Jiao X, Sherman B, Ma L, Chen X, Skelly T, Troyer J, Stephens R, Imamichi T, Pau A, Lempicki RA, Tran B, Nissley D, Lane HC, Dewar RL (2016). Towards better precision medicine: PacBio single-molecule long reads resolve the interpretation of HIV drug resistant mutation profiles at explicit quasispecies (haplotype) level. Journal of Data Mining in Genomics & Proteomics 7:182. [PubMed: 26949565]

Hughes ES, Bell JE, Simmonds P (1997). Investigation of the dynamics of the spread of human immunodeficiency virus to brain and other tissues by evolutionary analysis of sequences from the p17gag and env genes. J Virol 71: 1272–80. [PubMed: 8995651]

Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI (2006). A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. J Virol 80: 4698–704. [PubMed: 16641263]

Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–80. [PubMed: 23329690]

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105: 7552–7. [PubMed: 18490657]

Korber BT, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R, Wolinsky SM (1994). Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. J Virol 68: 7467–7481. [PubMed: 7933130]

Korlach J (2017). Understanding Accuracy in SMRT® Sequencing. Pacific Biosciences http://www.pacb.com/wp-content/uploads/2015/09/Perspective_UnderstandingAccuracySMRTSequencing.pdf. Accessed December 1, 2017.

Laird Smith M, Murrell B, Eren K, Ignacio C, Landais E, Weaver S, Phung P, Ludka C, Hepler L, Caballero G, Pollner T, Guo Y, Richman D, The IAVI Protocol C investigators & The IAVI African HIV Research Network, Poignard P, Paxinos EE, Kosakovsky Pond SL, Smith DM (2016). Rapid Sequencing of Complete env Genes from Primary HIV-1 Samples. Virus Evolution 2: vew018. [PubMed: 29492273]

Lamers SL, Gray RR, Salemi M, Huysentruyt LC, McGrath MS (2011). HIV-1 phylogenetic analysis shows HIV-1 transits through the meninges to brain and peripheral tissues. Infect Genet Evol 11: 31–7. [PubMed: 21055482]

Lane JH, Sasseville VG, Smith MO, Vogel P, Pauley DR, Heyes MP, Lackner AA (1996). Neuroinvasion by simian immunodeficiency virus coincides with increased numbers of perivascular macrophages/microglia and intrathecal immune activation. J Neurovirol 2: 423–32. [PubMed: 8972425]

Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, Mulenga J, Allen S, Hunter E, Hahn BH, Shaw GM, Blackwell JL, Derdeyn CA (2006). Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. J Virol 80: 5211–8. [PubMed: 16699001]

Liner KJ 2nd, Ro MJ, Robertson KR (2010). HIV, antiretroviral therapies, and the brain. Curr HIV/AIDS Rep 7: 85–91. [PubMed: 20425562]

Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, Zhu H, Woodward AS, Maenza J, Stevens CE, Stekler J, Collier AC, Genowati I, Deng W, Zioni R, Corey L, Zhu T, Mullins JI (2008). Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. Virology 374: 229–33. [PubMed: 18314154]

Liu Y, Tang XP, McArthur JC, Scott J, Gartner S (2000). Analysis of human immunodeficiency virus type 1 gp160 sequences from a patient with HIV dementia: evidence for monocyte trafficking into brain. J Neurovirol 6 Suppl 1: S70–81. [PubMed: 10871768]

McArthur JC, Steiner J, Sacktor N, Nath A (2010). Human immunodeficiency virus-associated neurocognitive disorders: Mind the gap. Ann Neurol 67: 699–714. [PubMed: 20517932]

McCrossan M, Marsden M, Carnie FW, Minnis S, Hansoti B, Anthony IC, Brettle RP, Bell JE, Simmonds P (2006). An immune control model for viral replication in the CNS during presymptomatic HIV infection. Brain 129: 503–16. [PubMed: 16317019]

McElroy K, Thomas T, Luciani F. (2015). Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. Microb Inform Exp 4:1

Mefford ME, Gorry PR, Kunstman K, Wolinsky SM, Gabuzda D (2008). Bioinformatic prediction programs underestimate the frequency of CXCR4 usage by R5X4 HIV type 1 in brain and other tissues. AIDS Res Hum Retroviruses 24: 1215–20. [PubMed: 18788913]

Mulinge M, Lemaire M, Servais JY, Rybicki A, Struck D, da Silva ES, Verhofstede C, Lie Y, Seguin-Devaux C, Schmit JC, Bercoff DP (2013). HIV-1 tropism determination using a phenotypic Env recombinant viral assay highlights overestimation of CXCR4-usage by genotypic prediction algorithms for CRF01_AE and CRF02_AG [corrected]. PLoS One 8: e60566. [PubMed: 23667426]

Nolan DJ, Lamers SL, Rose R, Dollar JJ, Salemi M, McGrath MS (2017). Single Genome Sequencing of Expressed and Proviral HIV-1 Envelope Glycoprotein 120 (gp120) and nef Genes. Bio-protocol 7: e2334.

Nottet HS, Gendelman HE (1995). Unraveling the neuroimmune mechanisms for the HIV-1-associated cognitive/motor complex. Immunol Today 16: 441–448. [PubMed: 7546209]

O'Connell O, Repik A, Reeves JD, Gonzalez-Perez MP, Quitadamo B, Anton ED, Duenas-Decamp M, Peters P, Lin R, Zolla-Pazner S, Corti D, Wallace A, Wang S, Kong XP, Lu S, Clapham PR (2013). Efficiency of bridging-sheet recruitment explains HIV-1 R5 envelope glycoprotein sensitivity to soluble CD4 and macrophage tropism. J Virol 87: 187–98. [PubMed: 23055568]

Ohagen A, Devitt A, Kunstman KJ, Gorry PR, Rose PP, Korber B, Taylor J, Levy R, Murphy RL, Wolinsky SM, Gabuzda D (2003). Genetic and functional analysis of full-length human

immunodeficiency virus type 1 env genes derived from brain and blood of patients with AIDS. J Virol 77: 12336–45. [PubMed: 14581570]

Peters PJ, Bhattacharya J, Hibbitts S, Dittmar MT, Simmons G, Bell J, Simmonds P, Clapham PR (2004). Biological analysis of human immunodeficiency virus type 1 R5 envelopes amplified from brain and lymph node tissues of AIDS patients with neuropathology reveals two distinct tropism phenotypes and identifies envelopes in the brain that confer an enhanced tropism and fusigenicity for macrophages. J Virol 78: 6915–6926. [PubMed: 15194768]

Peters PJ, Duenas-Decamp MJ, Sullivan WM, Brown R, Ankghuambom C, Luzuriaga K, Robinson J, Burton DR, Bell J, Simmonds P, Ball J, Clapham P (2008). Variation in HIV-1 R5 macrophage-tropism correlates with sensitivity to reagents that block envelope: CD4 interactions but not with sensitivity to other entry inhibitors. Retrovirology 5: 5. [PubMed: 18205925]

Peters PJ, Sullivan WM, Duenas-Decamp MJ, Bhattacharya J, Ankghuambom C, Brown R, Luzuriaga K, Bell J, Simmonds P, Ball J, Clapham PR (2006). Non-macrophage-tropic human immunodeficiency virus type 1 R5 envelopes predominate in blood, lymph nodes, and semen: implications for transmission and pathogenesis. J Virol 80: 6324–32. [PubMed: 16775320]

Pikora CA (2004). Glycosylation of the ENV spike of primate immunodeficiency viruses and antibody neutralization. Curr HIV Res 2: 243–54. [PubMed: 15279588]

Pillai SK, Pond SL, Liu Y, Good BM, Strain MC, Ellis RJ, Letendre S, Smith DM, Gunthard HF, Grant I, Marcotte TD, McCutchan JA, Richman DD, Wong JK (2006). Genetic attributes of cerebrospinal fluidderived HIV-1 env. Brain 129: 1872–83. [PubMed: 16735456]

Poon AF, Swenson LC, Bunnik EM, Edo-Matas D, Schuitemaker H, van 't Wout AB, Harrigan PR (2012). Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. PLoS Comput Biol 8: e1002753. [PubMed: 23133358]

Ritola K, Robertson K, Fiscus SA, Hall C, Swanstrom R (2005). Increased human immunodeficiency virus type 1 (HIV-1) env compartmentalization in the presence of HIV-1-associated dementia. J Virol 79: 10830–4. [PubMed: 16051875]

Rose R, Lamers SL, Dollar JJ, Grabowski MK, Hodcroft EB, Ragonnet-Cronin M, Wertheim JO, Redd AD, German D, Laeyendecker O (2017). Identifying Transmission Clusters with Cluster Picker and HIV-TRACE. AIDS Res Hum Retroviruses 33: 211–218. [PubMed: 27824249]

Rozera G, Abbate I, Ciccozzi M, Lo Presti A, Bruselles A, Vlassi C, D'Offizi G, Narciso P, Giombini E, Bartolini B, Ippolito G, Capobianchi MR (2012). Ultra-deep sequencing reveals hidden HIV-1 minority lineages and shifts of viral population between the main cellular reservoirs of the infection after therapy interruption. J Med Virol 84: 839–44. [PubMed: 22996031]

Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, Serwadda D, Sewankambo NK, Shepherd JC, Toma J, Huang W, Quinn TC (2009). Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J Infect Dis 199: 580–9. [PubMed: 19143562]

Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BT, Sharp PM, Shaw GM, Hahn BH (2008). Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 82: 3952–70. [PubMed: 18256145]

Salemi M, Lamers SL, Yu S, de Oliveira T, Fitch WM, McGrath MS (2005). Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS. J Virol 79: 11343–52. [PubMed: 16103186]

Schnell G, Joseph S, Spudich S, Price RW, Swanstrom R (2011). HIV-1 replication in the central nervous system occurs in two distinct cell types. PLoS Pathog 7: e1002286. [PubMed: 22007152]

Sede MM, Moretti FA, Laufer NL, Jones LR, Quarleri JF (2014). HIV-1 tropism dynamics and phylogenetic analysis from longitudinal ultra-deep sequencing data of CCR5- and CXCR4-using variants. PLoS One 9: e102857. [PubMed: 25032817]

Shapshak P, Segal DM, Crandall KA, Fujimura RK, Zhang BT, Xin KQ, Okuda K, Petito CK, Eisdorfer C, Goodkin K (1999). Independent evolution of HIV type 1 in different brain regions. AIDS Res Hum Retroviruses 15: 811–20. [PubMed: 10381169]

Simmonds P, Balfe P, Peutherer JF, Ludlam CA, Bishop JO, Brown AJ (1990). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J Virol 64: 864–72. [PubMed: 2296085]

Smit TK, Brew BJ, Tourtellotte W, Morgello S, Gelman BB, Saksena NK (2004). Independent evolution of human immunodeficiency virus (HIV) drug resistance mutations in diverse areas of the brain in HIV-infected patients, with and without dementia, on antiretroviral treatment. J Virol 78: 10133–48. [PubMed: 15331746]

Soda Y, Shimizu N, Jinno A, Liu HY, Kanbe K, Kitamura T, Hoshino H (1999). Establishment of a new system for determination of coreceptor usages of HIV based on the human glioma NP-2 cell line. Biochem Biophys Res Commun 258: 313–321. [PubMed: 10329384]

Strain MC, Letendre S, Pillai SK, Russell T, Ignacio CC, Gunthard HF, Good B, Smith DM, Wolinsky SM, Furtado M, Marquie-Beck J, Durelle J, Grant I, Richman DD, Marcotte T, McCutchan JA, Ellis RJ, Wong JK (2005). Genetic composition of human immunodeficiency virus type 1 in cerebrospinal fluid and blood without treatment and during failing antiretroviral therapy. J Virol 79: 1772–88. [PubMed: 15650202]

Sturdevant CB, Joseph SB, Schnell G, Price RW, Swanstrom R, Spudich S (2015). Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection. PLoS Pathog 11: e1004720. [PubMed: 25811757]

Takahashi K, Wesselingh SL, Griffin DE, McArthur JC, Johnson RT, Glass JD (1996). Localization of HIV-1 in human brain using polymerase chain reaction/in situ hybridization and immunocytochemistry. Ann Neurol 39: 705–711. [PubMed: 8651642]

Thomas ER, Dunfee RL, Stanton J, Bogdan D, Taylor J, Kunstman K, Bell JE, Wolinsky SM, Gabuzda D (2007). Macrophage entry mediated by HIV Envs from brain and lymphoid tissues is determined by the capacity to use low CD4 levels and overall efficiency of fusion. Virology 360: 105–19. [PubMed: 17084877]

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38: e159. [PubMed: 20571086]

Tsai HC, Chou PY, Wann SR, Lee SS, Chen YS (2015). Chemokine co-receptor usage in HIV-1-infected treatment-naive voluntary counselling and testing clients in Southern Taiwan. BMJ Open 5: e007334.

van't Wout AB, Ran LJ, Kuiken CL, Kootstra NA, Pals ST, Schuitemaker H (1998). Analysis of the temporal relationship between human immunodeficiency virus type 1 quasispecies in sequential blood samples and various organs obtained at autopsy. J Virol 72: 488–96. [PubMed: 9420250]

Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL (2014). The global transmission network of HIV-1. J Infect Dis 209: 304–13. [PubMed: 24151309]

Williams KC, Corey S, Westmoreland SV, Pauley D, Knight H, deBakker C, Alvarez X, Lackner AA (2001). Perivascular macrophages are the primary cell type productively infected by simian immunodeficiency virus in the brains of macaques: implications for the neuropathogenesis of AIDS. J Exp Med 193: 905–15. [PubMed: 11304551]

Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, Buck D, Au KF (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Res 6:100 [PubMed: 28868132]

Wong JK, Ignacio CC, Torriani F, Havlir D, Fitch NJ, Richman DD (1997). In vivo compartmentalization of human immunodeficiency virus: evidence from the examination of pol sequences from autopsy tissues. J Virol 71: 2059–2071. [PubMed: 9032338]

Wu X, Parast AB, Richardson BA, Nduati R, John-Stewart G, Mbori-Ngacha D, Rainwater SM, Overbaugh J (2006). Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. J Virol 80: 835–44. [PubMed: 16378985]

Yi Y, Chen W, Frank I, Cutilli J, Singh A, Starr-Spires L, Sulcove J, Kolson DL, Collman RG (2003). An unusual syncytia-inducing human immunodeficiency virus type 1 primary isolate from the central nervous system that is restricted to CXCR4, replicates efficiently in macrophages, and induces neuronal apoptosis. J Neurovirol 9: 432–41. [PubMed: 12907388]

Zhang H, Tully DC, Hoffmann FG, He J, Kankasa C, Wood C (2010). Restricted genetic diversity of HIV-1 subtype C envelope glycoprotein from perinatally infected Zambian infants. PLoS One 5: e9294. [PubMed: 20174636]

Zhou S, Bednar MM, Sturdevant CB, Hauser BM, Swanstrom R (2016). Deep Sequencing of the HIV-1 env Gene Reveals Discrete X4 Lineages and Linkage Disequilibrium between X4 and R5 Viruses in the V1/V2 and V3 Variable Regions. J Virol 90: 7142–58. [PubMed: 27226378]

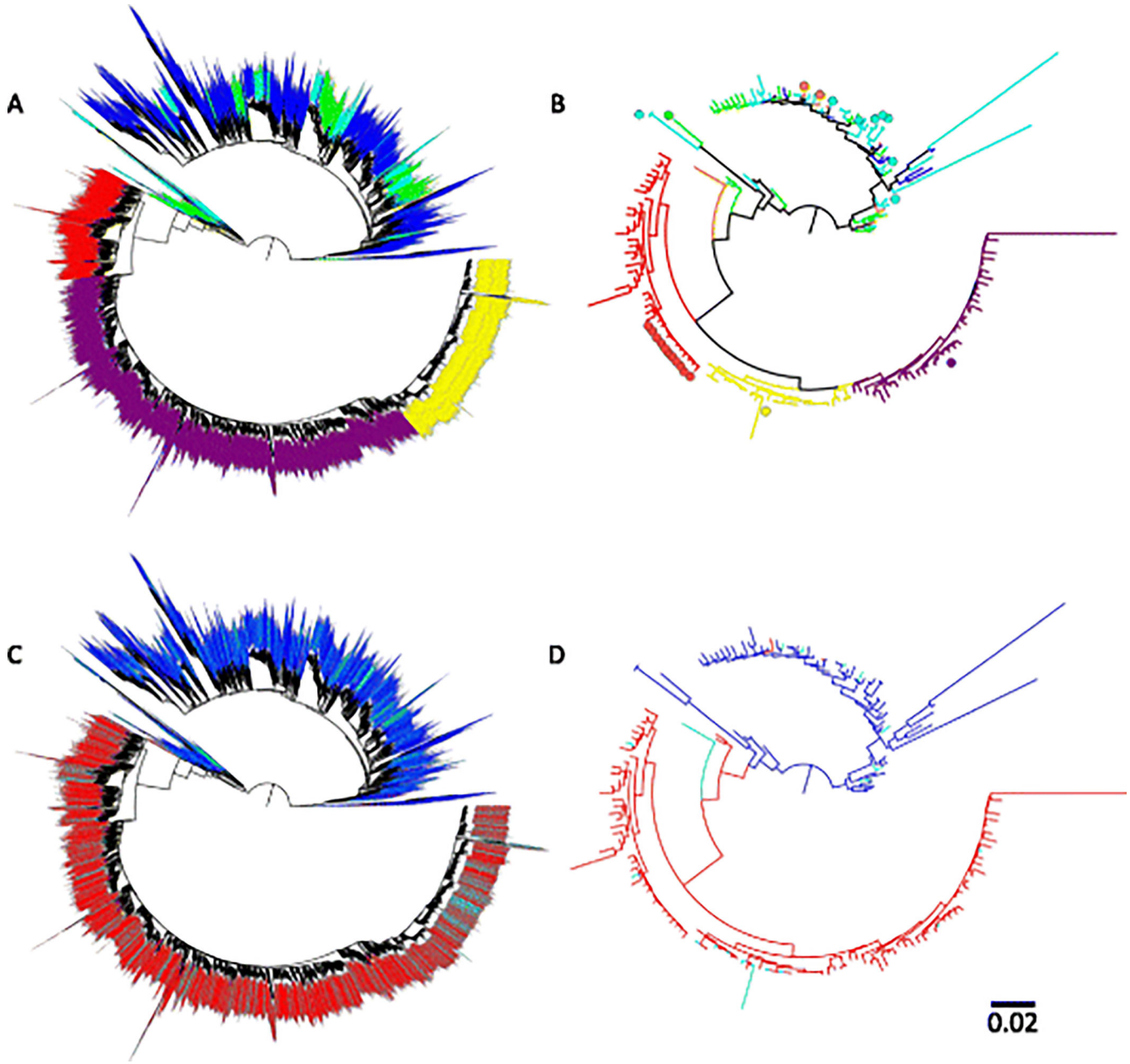

Fig. 1. Maximum-likelihood tree using the HQCS (a,c) and HQCS10 (b) datasets.

Branches are scaled in substitutions/site according to the bar at the bottom of each tree. (a-b) Branches are colored to indicate the tissue of origin as follows: dark blue=blood; orange=colon; aqua=lung; green=lymph node; red=frontal lobe; yellow=occipital lobe; purple=parietal lobe. (b) Molecular clones are indicated with circles. (c) R5 and X4 variants estimated using WebPSSM x4r5 algorithm. Branches are colored as follows: blue=R5; red=X4; aqua=undetermined

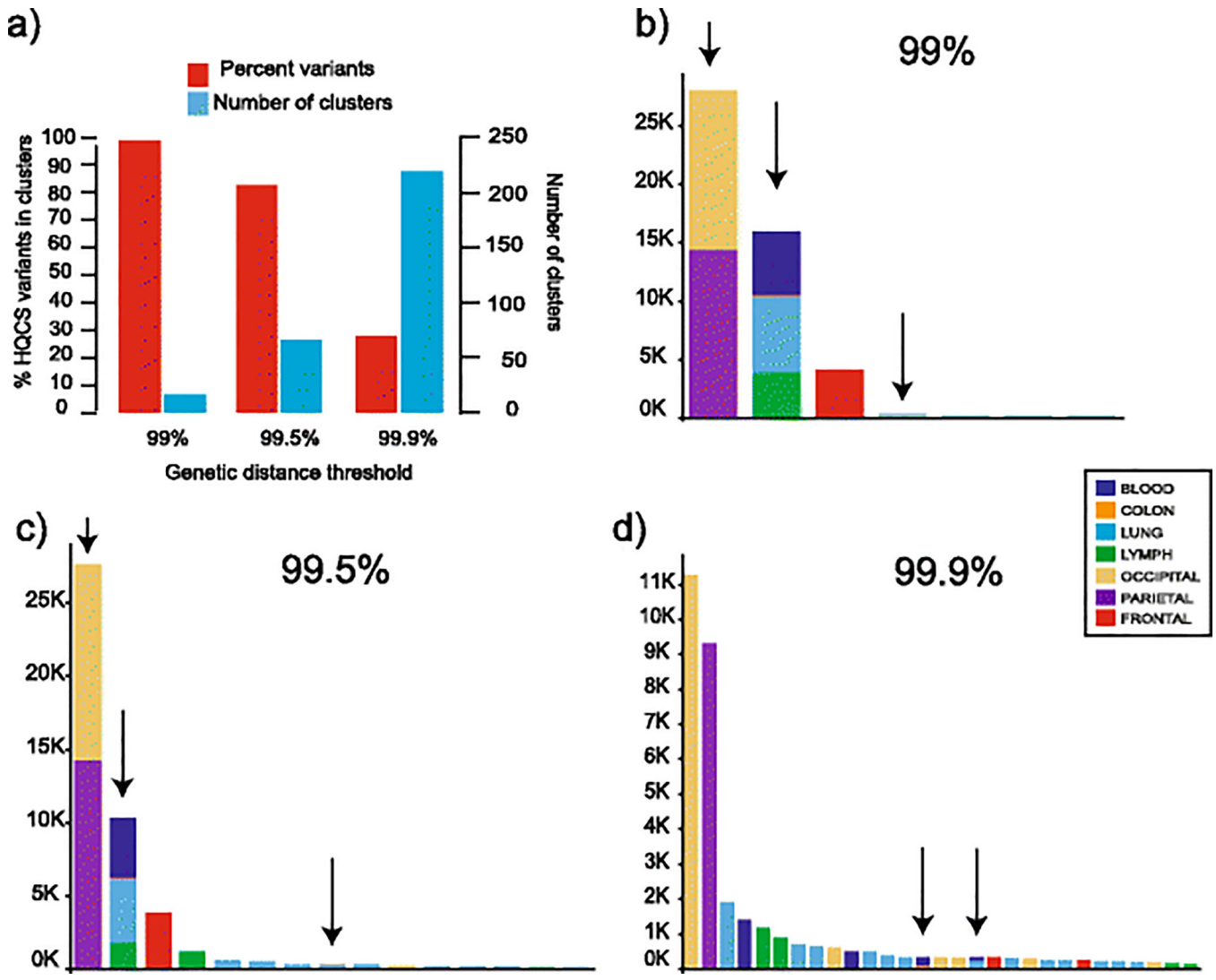

Fig. 2. Distance-based clustering for SMRT HQCS at 3 genetic distance thresholds.

(a) Percent of all variants placed in any cluster (red bars, left axis) and the total number of clusters (blue bars, right axis) at three thresholds (x-axis). (b-d) The number of total Aligned reads (y-axis) in identified clusters (x-axis) at three thresholds. The tissue of origin is represented by colors as shown in the legend. Arrows above the bars indicate clusters with sequences from >1 tissue. For clarity only clusters with >=100 reads are shown

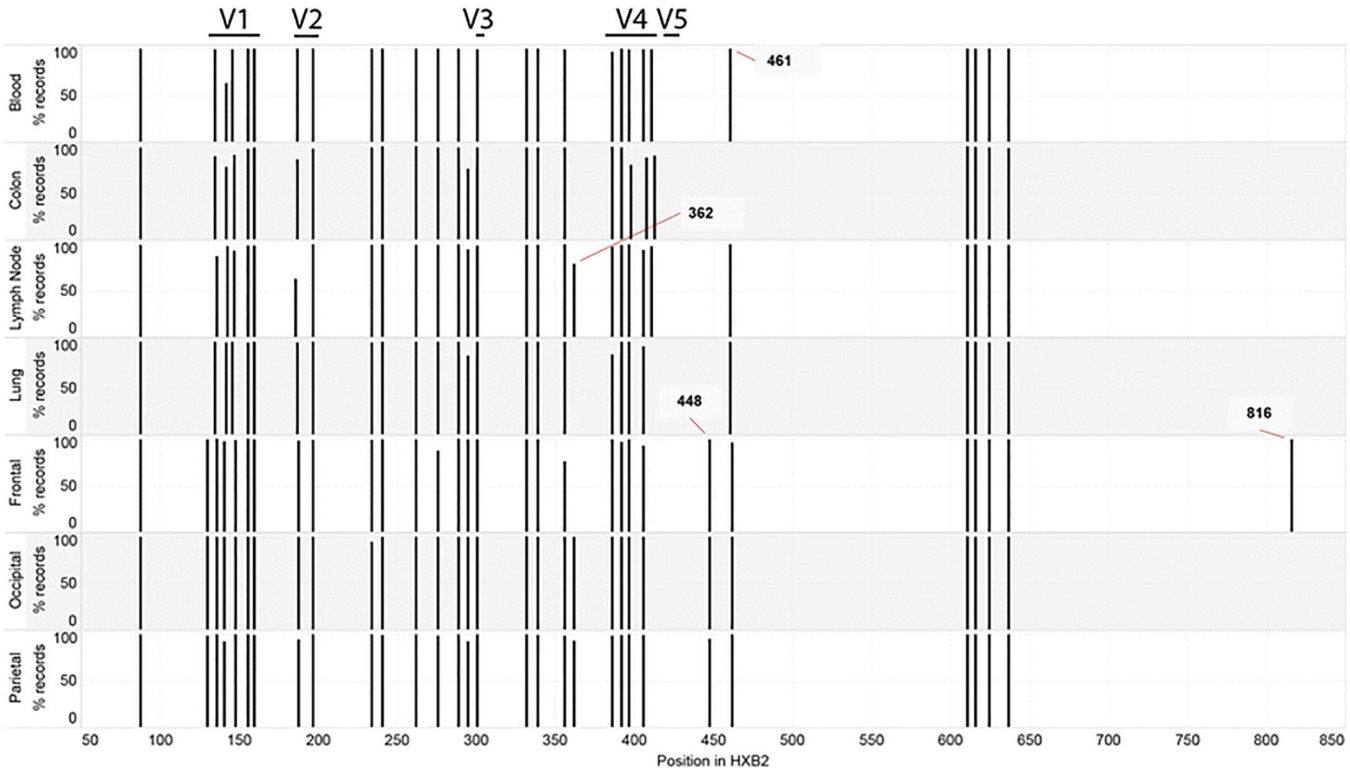

Fig. 3. Positions of N-sites along the HIV env.

The position in the protein numbered according to HXB2 is on the *x-axis*. The location of variable domains V1-V5 are shown above the graph. The tissue names are on the *y-axis*, with bars indicating the percent records from Table 1 AA sequences that had a glycosylation site at any position. Sites where there was a dramatic absence or presence of a N-site are labeled for clarity

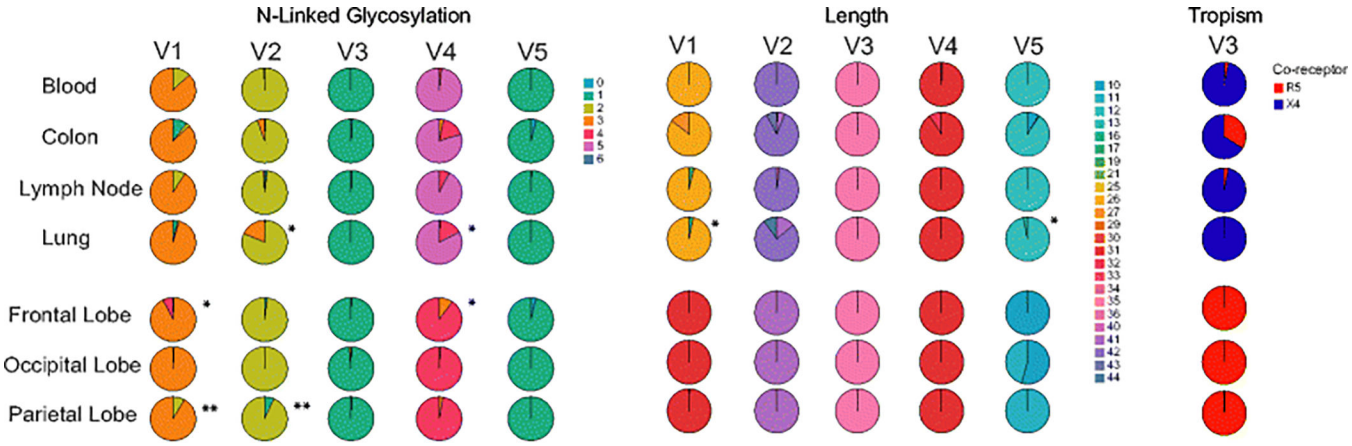

Fig. 4. Variable region characteristics.

For each of the 7 tissues (on left), the number of glycosylation sites (left), length (middle) and tropism prediction using WebPSSM (right) are shown as pie charts. The sections of the pie charts are colored according to the legend for each analysis. The size of the pie slices is proportional to the number of sequences for each tissue

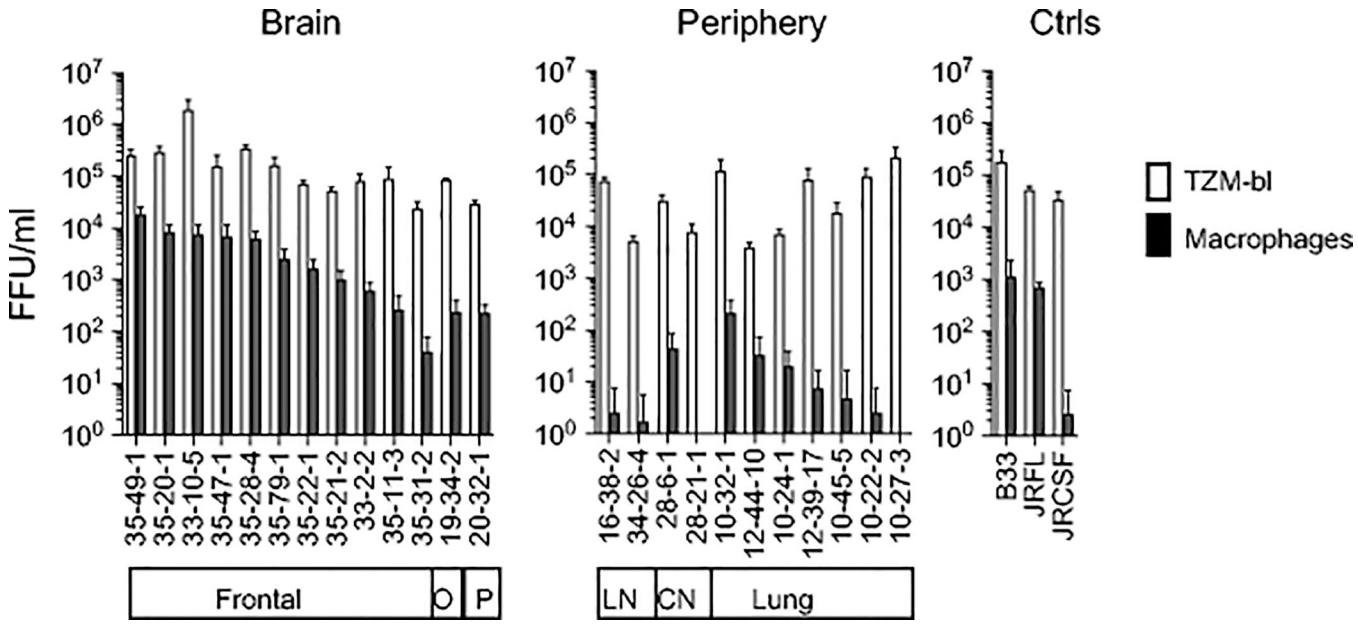

Fig. 5. Functionality and macrophage infectivity of Env+ pseudoviruses.

Infectivity of Env+ pseudoviruses for HeLa TZM-bl cells and primary macrophages. O, occipital lobe; P, parietal lobe; LN, lymph node; CN, colon

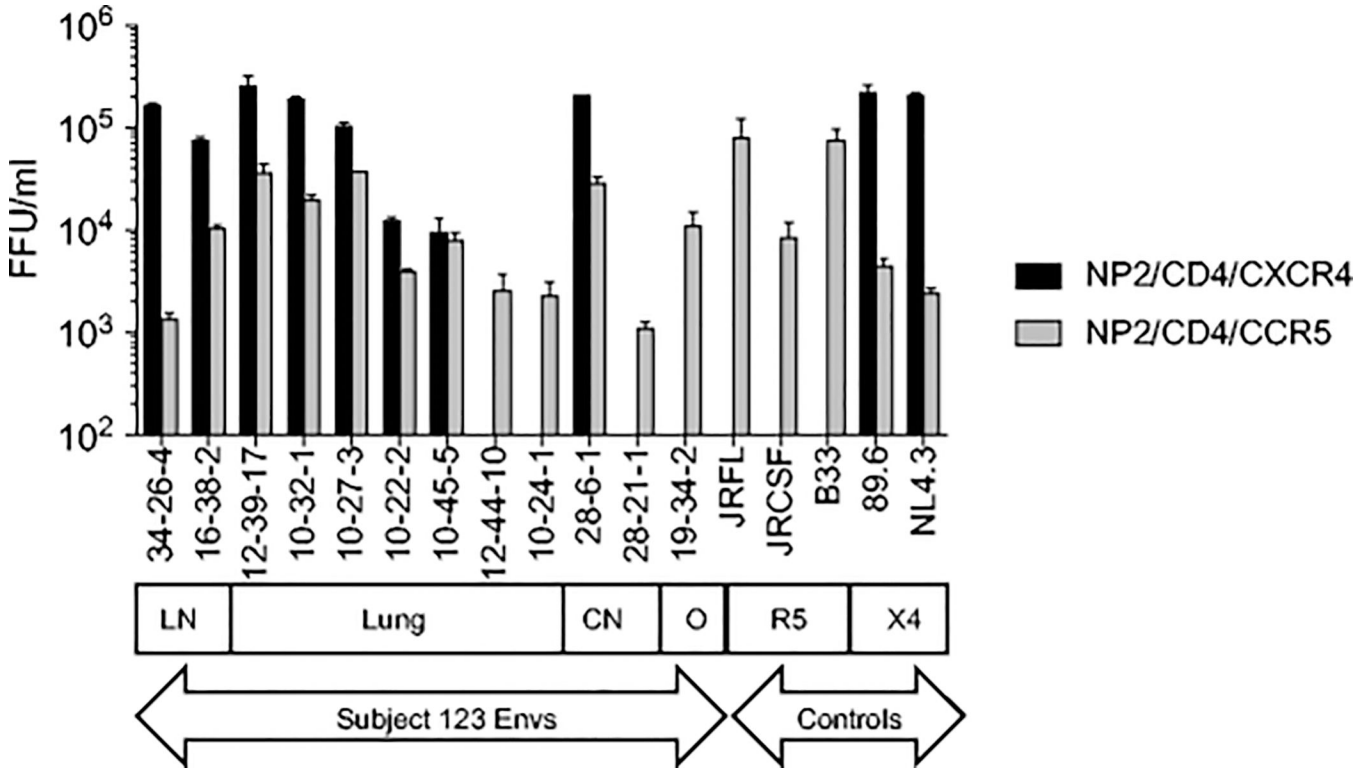

Fig. 6. Co-receptor-use of 11 Env+ pseudoviruses.

Infectivity of pseudoviruses carrying subject 123 Envs for NP2/CD4/CCR5 and NP2/CD4/CXCR4 cells to establish CXCR4-use. LN, lymph node; CN, colon. O, occipital lobe

Table 1.

SMRT full length *env* sequences and clustering by USEARCH

Sequences	Frontal lobe	Occipital lobe	Parietal lobe	LN	Colon	Lung	Blood cells	Total
Raw	4725	17030	16360	5341	292	10024	7198	60970
Assembled	4704	17026	15899	4959	268	10011	7163	60030
Aligned	4184	15445	14404	4454	249	8369	5620	52725
Unique	4047	13581	14385	4124	249	7410	5516	49312
HQCS	1115	1600	4115	1030	7	769	3137	11772
HQCS10	29	26	37	21	6	27	8	154
AA	2849	9906	7981	2803	249	1512	1800	27100

Table 2.

Comparisons of segregating glycosylation and length variants between variable regions for each tissue.

Analysis	Regions Compared	Blood	Colon	Frontal	Lymph	Lung	Occipital	Parietal
Glyc	V1:V2	NA	NA	No	No	No	NA[#]	p<0.00001
Glyc	V1:V4	No	No	p<0.00001	No	No	NA	No
Glyc	V2:V4	NA	No	No	No	p<0.00001	NA	No
Length	V1:V4	NA	No	NA	NA	NA	NA	NA
Length	V1:V5	NA	No	NA	NA	p<0.00001	NA	NA
Length	V2:V4	NA	NA[#]	NA	NA	NA	NA	NA
Length	V4:V5	NA	NA[#]	NA	NA	NA	NA	NA

Significance assessed at the p=0.01 level. No=p>0.01.

Glyc = number of glycosylation sites.

NA= The chi-squared test was not performed because there were cells with value=0 and/or no minority variants occurring in at least 10 sequences.

[#]Due to cells with value=0 the chi-squared test could not be performed; however, the minority variant for each region only segregated with the majority variant of the other region.

Table 3.

Env genes cloned from nested end-point PC

Tissue	No. of env clones
Frontal lobe	11
Occipital lobe	1
Parietal lobe	1
LN	2
Colon	2
Lung	7