DIFFERENT REGIONS OF HIV-1 SUBTYPE C ENV ARE ASSOCIATED WITH PLACENTAL LOCALIZATION AND IN UTERO MOTHER-TO-CHILD TRANSMISSION

Surender B. Kumar§, Samuel K. Handelman§, Igor Voronkin§, Victor Mwapasa§, Daniel Janies§,
Stephen J. Rogerson§, Steven R. Meshnick§, and Jesse J. Kwiek†

1. Department of Veterinary Biosciences and The Center for Retrovirus Research, 2. Mathematical Biosciences Institute, 3. Department of Biomedical Informatics, 4. Division of Infectious Diseases, Department of Microbiology, The Center for Microbial Interface Biology, and The Center for Retrovirus Research, The Ohio State University, Columbus, Ohio. 5. Department of Community Health, Malawi College of Medicine, Blantyre, Malawi. 6. Department of Medicine, University of Melbourne, Parkville, Victoria, Australia. 7. Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina.

§ These Authors contributed equally to this manuscript.
†Corresponding Author.
Jesse J. Kwiek
1008 Biomedical Research Tower
460 West 12th Avenue
Columbus, OH 43210
Email: kwiek.2@osu.edu
T: (614) 292-3256
F: (614) 292-9616

Running Title: In utero HIV-1 mother-to-child transmission
ABSTRACT (212 words, 250 words maximum)

HIV-infections are initiated by a limited number of variants that diverge into a diverse quasispecies swarm. During in utero mother-to-child transmission (IU MTCT), transmitted viral variants must pass through multiple unique environments, and our previously published data suggest a non-stochastic model of transmission. As an alternative to a stochastic model of viral transmission, we hypothesize that viral selection in the placental environment influences the character of the viral quasispecies when HIV-1 is transmitted in utero. To test this hypothesis, we used single template amplification to isolate HIV-1 envelope gene (env) sequences from both peripheral plasma and the placentas of eight nontransmitting (NT) and nine IU-transmitting participants. Statistically-significant compartmentalization between peripheral and placental HIV-1 env was detected in one of the eight NT cases and six of the nine IU MTCT. In addition, viral sequences isolated from IU MTCT placental tissue showed variation in env V1 loop lengths compared to matched maternal-sequences, while NT placental env sequences did not. Finally, comparison of env sequences from NT and IU MTCT participants indicated statistically significant differences in Kyte-Doolittle hydropathy in the signal peptide, C2, V3, and C3 regions. Our working hypothesis is that the hydropathy differences in Env associated with IU MTCT alter viral cellular tropism or affinity, allowing HIV-1 to efficiently infect placentally-localized cells.
INTRODUCTION

According to the UNAIDS estimates, in 2007 there were approximately 1.4 million HIV-infected pregnant women in low- and middle-income countries in need of antiretroviral therapy to prevent HIV-1 mother-to-child transmission (MTCT). In populations where replacement feeding is not feasible, it has been estimated that 12% of MTCT occurs \textit{in utero} before 36 weeks gestation, 29% occurs \textit{in utero} between 36-weeks and delivery, 20% occurs during labor and delivery (peripartum), and the remaining 39% occurs during prolonged breastfeeding (post-partum)(39). Although low- and middle-income countries have made progress to increase the use of antiretrovirals to prevent peripartum and post-partum HIV-1 MTCT, in 2007 only 23% of the estimated 73,000 HIV-infected pregnant Malawians had access to the required antiretrovirals (WHO/UNAIDS/UNICEF 2008 progress report). In addition, access to other preventative measures such as elective cesarean sections and replacement formula feeding remain unrealized. As coverage of new intervention strategies to prevent breast feeding and intrapartum transmissions increases, the overall rate of transmission will decrease but a greater proportion of the transmissions will occur \textit{in utero} via largely uncharacterized biological mechanisms (21).

HIV-infections are initiated by a single or limited number of variants that diverge and diversify into a complex quasispecies swarm (1, 63, 68, 71, 80, 83, 85). In other words, although a chronically infected pregnant woman has a heterogeneous viral population, the complexity of the viral population is severely restricted during transmission, and a homogeneous virus population is often seen in vertically-infected infants. Three mechanisms have been proposed to explain this restriction: 1) transmission of the most abundant viral variant (stochastic model), 2) transmission of multiple variants followed by selective amplification of host-adapted variants (selective amplification model),
or, 3) limited variant transmission restricted by host and/or viral biology (selection model) (86). We have previously reported a transmission bottleneck of HIV-1 subtype C during in utero MTCT, based on a heteroduplex tracking assay of the V1/V2 region of env (42), and these data suggested that the bottleneck was inconsistent with a stochastic model of transmission. Building on these findings, we investigated the possibility that selection of viral lineages underlies transmission of HIV-1 variants in IU MTCT (model three, above).

There are several opportunities for selective pressure during IU MTCT. The transmitted variants must cross several cell layers in the placenta (17) and sequentially adapt to three unique immune environments: an immunologically developed or depleted host, an immunologically-blunted placenta, and finally to an immunologically naïve infant. We hypothesize that the placental environment provides the selective force for the phenotype of the viral quasispecies when HIV-1 is transmitted in utero. To address this hypothesis, we used single template amplification to characterize the molecular evolution of HIV-1 maternal quasispecies swarms and the individual lineages that transmit and those that do not transmit during in utero MTCT. We report quantitative and qualitative differences in the gp120 region of HIV-1 env in viral populations isolated from peripheral and placental compartments. Furthermore, we report sequence differences in the env signal peptide, C2, V3, and C3 region that are associated with MTCT status.

MATERIAL AND METHODS

Study Participants and Clinical Isolates. Clinical isolates of virus were obtained from HIV-infected pregnant women participants in the Malaria, HIV and Pregnancy cohort (MHP cohort) during 2001-2004 (53, 54) in Blantyre, Malawi. IRB approval was obtained from The Ohio State University, University of North Carolina at Chapel Hill and the Malawi College of Medicine.
Research Ethics Committee; all participants gave written, informed consent. HIV-infected mothers and their infants received nevirapine according to the HIVNET 012 protocol, which was the standard of care at the time of this study (27). In the parent MHP study, participants who were sero-reactive on two independent HIV-tests were deemed HIV-positive. Infant HIV-1 status was determined by real-time PCR according to the methods of Luo et al. (47) and classified as non-transmitting (NT) or intrauterine transmission (IU) according to the definitions outlined by Bryson et al. (10). HIV-1 exposed but uninfected infants, hereafter called non-transmitters (NT), were HIV-1 DNA negative both at birth and at six-weeks postpartum, while intrauterine transmission (IU) cases were HIV-1 DNA positive within 48 hours of birth. Peripheral blood was collected from the pregnant women by venipuncture and plasma was isolated according to standard protocols. After delivery, an incision was made in the middle of the maternal surface of the placenta, 2 cm long and through half the thickness of the placental tissue, and the placental blood that pooled in the incision was collected, with the presumption that the resulting blood likely represents an admixture of maternal and fetal blood (57). In addition to placental blood, a full-thickness biopsy was taken from the pericentric area of the placenta and snap-frozen in liquid nitrogen. For simplicity, HIV-1 env genes from placental blood and placental biopsies are aggregated into a single category termed “placenta.” Blood from the umbilical cord vein was collected and plasma was prepared; previously we have demonstrated that HIV-1 variants in umbilical cord blood represented the infant and not the maternal quasispecies (42). Infant blood was collected from heel-pricks at birth, 6- and 12-weeks. Constrained by the availability of matched placental tissue and maternal plasma, we identified 84 NT participants (of which 29 were randomly selected) and 15 IU MTCT cases. Only 9 of the 29 NT cases and 9 of the 15 IU cases supported gp160 nested PCR amplification from all tissues, which is described below.
Using the quality control tool from the Los Alamos National Laboratory (http://www.hiv.lanl.gov), we determined that all env amplicons were HIV-1 subtype C, there was no contamination with common laboratory isolates, and there was epidemiological linkage between all mother-placenta-infant pairs, with one exception. Sequences isolated from MHP sample 2466 were not monophyletic on the phylogenetic tree; all maternal samples formed on a single subclade, whereas the matched placental sequences formed another clade that was not a sister group to the maternal clade. There was no co-mingling of the 2466 sequences with other sequences, so we excluded sample cross contamination as the source of this result. This sample was eliminated from further analysis, resulting in a total of eight NT samples. Three env sequences from three different mother/placenta pairs were predicted to be inter-subtype recombinants by the LANL quality control tool. In each case, comparison of these predicted recombinant sequences with their nearest neighbor showed sequence differences that could be explained by a single, large deletion and therefore these sequences are unlikely to be true inter-subtype recombinants.

Viral RNA Quantification, Extraction, and cDNA Synthesis. HIV-1 RNA concentration in peripheral and placental plasma isolates was determined with the Roche Amplicor HIV-1 Monitor test (version 1.5) (54), which had a lower threshold of detection of 400 copies/mL; any samples with viral loads below the lower limit of detection were assigned a value of 200 copies/mL. Differences in matched viral loads (i.e. peripheral versus placental viral load) were tested with the Wilcoxon signed-rank test, and differences between unmatched viral loads (i.e. placental plasma of NT versus placental plasma of IU) were tested with the two-sample Wilcoxon rank-sum (Mann-Whitney) test; both tests were implemented with STATA version 10.1 (STATA Corporation, College Station, TX). RNA was extracted from 140 µL centrifuged plasma using the QIAamp viral RNA mini kit used for other plasma samples. If the plasma viral concentration was <10,000 RNA copies/mL then samples
were concentrated by ultra-centrifugation at 24,500 x g for 1.5 hrs. at 4°C. To extract RNA from placental biopsies, approximately 100 mg tissue was pre-incubated for 3-4 minutes in Qiazol lysis reagent, placed in a QIA shredder column, and centrifuged for 2 min at 21,130 x g. The eluent was processed with the RNeasy Lipid Tissue mini kit (Qiaagen) and the RNA was eluted into 50 µL of RNase-free water. RNA was reverse transcribed into cDNA with SuperScript III (Invitrogen Life Technologies, Carlsbad, CA) as follows: reactions containing 50 µL RNA, 0.5 mM dNTPs, and 0.25 µM HIV-1 subtype C specific primer (OFM19) were heated for 5 min at 65°C then quenched on ice. cDNA synthesis was performed in 100 µl with 1 × reverse transcriptase buffer, 5 mM dithiothreitol, 80 units RNaseOUT, and 1000 units SuperScript III (Invitrogen Life Technologies, Carlsbad, CA). The reaction mixture was sequentially incubated at 50°C and 55°C for one hour each, inactivated at 70°C for 15 minutes, incubated with 2 units RNase H (Invitrogen Life Technologies, Carlsbad, CA) and then digested at 37°C for 20 minutes. To ensure that only HIV-1 RNA was isolated from placental biopsies, placental tissue extracts underwent a mock cDNA synthesis without SuperScript III; this reaction never supported PCR amplification, indicating that HIV-1 RNA and not HIV-1 DNA in the placental biopsy served as the PCR template (data not shown).

**Single-template Amplification.** It has been demonstrated that PCR amplification of a related, yet heterogeneous mixture of templates is prone to PCR-induced template recombination, which can be mitigated by performing PCR on single templates (60, 78). Single template amplification was performed according to published protocols (35, 58, 66, 74). Briefly, serial dilutions of cDNA were made until less than 30% of eight to ten replicate PCR reactions generated amplicons. All PCR reactions used High Fidelity Platinum Taq (Invitrogen) according to the manufacturer’s instructions and the subtype C-optimized outer primers VIF and OFM19 and inner primers ENVA* and ENVN (sequences listed in Supplemental Table 1).
DNA Sequencing. Prior to sequencing, PCR amplicons were incubated with exonuclease I (New England Biolabs) and alkaline phosphatase (Roche) at 37°C for 30 minutes followed by heat inactivation at 95°C for 5 minutes (6). Full-length gp160 was sequenced using eight HIV-1 subtype C optimized primers (Genewiz, Inc, NJ) [Supplemental Table S1], which provided a minimum of two-fold coverage. Sequencing chromatograms were manually examined for multiple peaks, and those with multiple concurrent peaks were discarded. Env sequences were trimmed, assembled, and manually edited using Geneious software (http://www.geneious.com). All sequences were examined with the HIV Sequence Quality Control Tool on the Los Alamos National Laboratory HIV-database website (http://www.hiv.lanl.gov). Results from this tool allowed us to exclude potential contamination with laboratory isolates and to conclude that all env sequences are most closely related to HIV-1 subtype C (data not shown). Env sequences from the signal-peptide through gp120 (abbreviated sp-gp120) have been deposited into Genbank (http://ncbi.nlm.nih.gov) under accession numbers (submission in progress); features of gp41 will be described elsewhere.

Sequence Alignment and Phylogenetic Tree Construction. Cleaned gp160 sequences were first aligned using HMMER (19) to the Env protein model provided by the LANL. This alignment was manually adjusted in less than ten instances correct alignment of sequence landmarks and maintain reading frame. The adjusted alignment was used to train a new subtype-C specific Env protein model, followed by alignment of the sequences to this new model, using HMMER. Finally, this alignment was used for a tree search in RAxML (76) with a general time reversible substitution model with gamma distributed rate heterogeneity and a proportion of invariant sites (GTR-gamma-I) (determined using Findmodel, http://www.hiv.lanl.gov). The best scoring tree under the maximum likelihood criterion after 100 replicates is shown in Supplemental Figure 1.
Assessment of Recombination. Genetic Algorithm Recombination Detection [GARD] (38), a standard recombination-detection method included in HYpothesis testing using PHYlogenies [HYPHY] (59), was used to screen all epidemiologically-linked maternal and placental sequences for evidence of recombination. Significant breakpoints are reported at $p<0.05$, as reported by HYPHY.

The phylogenetic tree associated with each segment identified by GARD is used for the positive selection analyses described below.

Compartmentalization testing. Using the tree produced by RAxML, Slatkin-Maddison (75) $p$-values were calculated using the standard analysis packaged in HYPHY (59). A second test of compartmentalization, the Hudson nearest-neighbor method (31) was also used; this method is a robust measure of compartmentalization when there is an unequal number of sequences in the two compartments (84), as is the case for some of the pregnancies in these data. The analysis was performed both with and without duplicate sequences.

Sequence Characteristics. Evidence of APOBEC-induced sequence hypermutation was measured with Hypermur (http://www.hiv.lanl.gov). Hypervariable loops from the gp120 sequences were extracted with the gene cutter program (http://www.hiv.lanl.gov). Potential N-linked glycosylation sites (PNLGs) on both gp120 and the individual hypervariable loops were enumerated with the N-glycoside program (http://www.hiv.lanl.gov). Co-receptor usage of the subtype C gp120 genes was predicted from the extracted V3 sequences with the web position-specific scoring matrices tool trained for HIV-1 subtype C (PSSM-C), which is maintained by the University of Washington (32).
Alignment position quality controls and multiple hypothesis correction. Because multiple amino acid positions are considered simultaneously, it is necessary to perform a multiple hypothesis correction. An amino acid position was included in the analysis if it was consensus deletion in no more than two (of seventeen) participants and no more than twelve of the participants shared the same plurality amino acid, excluding those participants where the Shannon entropy at that amino acid position (among different variants) was greater than 0.4. This minimum diversity requirement excludes amino acid positions that are too conserved to ever rise above the multiple hypothesis correction (see next), based on estimates from Fisher’s exact test (p<0.001 with fewer than five mothers showing the rare variant amino acid; a Shannon entropy of 0.4 roughly corresponds to a rare-variant frequency of 22%, which should be detected in 95% of bootstrap replicates.) After filtering, 118 amino acid positions and their corresponding codons were included in the analyses described below. To correct for multiple hypothesis testing, the Benjamini-Hochberg method was used with a primary false-discovery rate less than 0.05; in order to reduce the probability of false negative results, we also report a secondary false discovery rate of 0.25 (7).

Positive and Purifying Selection. Positions under positive selection were identified by differences in relative synonymous and non-synonymous substitution rates (dN – dS) as reported by the standard analysis packaged with HYPHY (59), including an adjustment for any recombination detected by GARD. HIV-1 sequences derived from each volunteer were treated separately and the observed and expected values for synonymous and non-synonymous substitution rates were combined across volunteers after stratification by their transmission status. The p-value for the sum of these substitution rates was obtained using the same binomial-approximation calculation employed by HYPHY(59). This analysis was performed only for the 118 codons corresponding to amino acids used in other analyses, and the same multiple hypothesis correction was applied.
Bootstrapping method to identify sequence differences associated with in utero MTCT. For purposes of identifying sequence differences associated with IU MTCT, comparisons are made at individual sites in HIV-1 env sequences derived from IU MTCT pregnancies versus those derived from NT pregnancies. Statistically significant differences were identified using a nested bootstrapping procedure (16), with a given pregnancy sampled on the outer layer and individual sequences from each pregnancy sampled on the inner layer, for a total of twelve sequences per pregnancy per sample (the fewest for any pregnancy). Allowing this number to match the full number of available sequences did not render any of the results non-significant. At each alignment position, in each bootstrap replicate (100,000 replicates), the difference between the average Kyte-Doolittle hydropathy index (43) of the IU MTCT sequences in the sample and the NT sequences in the sample is calculated. At each alignment position, a given bootstrap replicate supports the null hypothesis (similar hydropathy) if the difference is zero, or if the difference has a sign opposite to the difference observed in the data set as a whole. The p-value at each alignment position is the fraction of bootstrap replicates in which the null hypothesis is supported (e.g. the p-value drops, becoming more significant, if the difference seen in the data set is also seen in a large fraction of the bootstrap replicates). The bootstrap p-values were only calculated for the 118 alignment positions where a significant association could be detected given the available data (see section entitled “Alignment position quality controls and multiple hypothesis correction” above).

Feature Mapping on Env Crystal Structure. Amino acid positions of interest were mapped onto the JRFL Env crystal structure (PDB ID: 2B4C) using Jmol: an open-source Java viewer for chemical structures in 3D (http://www.jmol.org/).
RESULTS

Comparison of peripheral and placental plasma associated HIV-1 viral load. Previous studies from the Malaria and HIV in Pregnancy (MHP) cohort have reported an association between malaria and HIV-1 viral load (54); an association between maternal syphilis seroreactivity and HIV-1 MTCT (53); an association between maternal-infant blood admixture during intrapartum HIV-1 MTCT (40, 41); and a reduction in HIV-1 quasispecies diversity during HIV-1 MTCT (42). In order to better understand the relationship between HIV-1 viral load in peripheral and placental plasma, we stratified viral load data from the MHP cohort by HIV-1 transmission status. Table 1 shows the clinical characteristics of the 266 non-transmitting (NT) and 44 in utero transmitting (IU) HIV-infected women whose peripheral viral loads were quantified; a subset of 130 NT and 28 IU MTCT cases had matched placental viral load data. As shown in Figure 1, the median placental HIV-1 concentration among NT women was significantly lower than the median HIV-1 concentration in matched peripheral plasma (p<0.0001), while in contrast, the median placental HIV-1 concentration in IU women was equivalent or slightly higher than the median HIV-1 concentration in matched peripheral plasma (p=0.2). In comparisons between the two groups, the median placental HIV-1 concentration in IU women was significantly higher than in NT women (5.2 log_{10} copies/mL versus 3.5 log_{10} copies/mL, respectively, p<0.0001). While IU MTCT cases were significantly more likely to be co-infected with syphilis (consistent with previous reports (53)), other clinical features did not differ significantly by transmission status. These results show two distinct differences in viral load that are associated with MTCT status. First, placental HIV-1 concentration is quantitatively higher in cases of IU MTCT than in cases of NT. Second, the relationship between peripheral and placental HIV-1 concentration differs by MTCT status: the vast majority of NT pregnancies had placental viral loads that were on average an order of magnitude lower than matched peripheral...
concentrations, while the placental viral loads were similar to matched peripheral viral loads among
the IU MTCT pregnancies.

env Sequence Characteristics. Previous studies from the same MHP cohort revealed a restriction
of HIV-1 env genetic diversity after IU MTCT, in a manner inconsistent with a stochastic mechanism
of variant transmission (42). Building on these data, we tested an alternative model, natural selection
of viral lineages, to explain the genetic restriction observed during IU MTCT. In order to test this
hypothesis, we used single template amplification to isolate HIV-1 env from peripheral plasma
(maternal), various placental tissues (placental plasma and placental biopsies), and umbilical cord
blood (infant). Table 2 shows the clinical characteristics of the maternal donor, the number of gp160
amplicons sequenced from the peripheral and placental compartment, and the biochemical
characteristics of the sequences of these matched samples. A total of 589 gp160 sequences from
maternal plasma, placental plasma, placental biopsy, cord plasma, neonatal plasma and 12-week
infant plasma were obtained. After alignment and in silico translation, 61 of the 589 sequences
(~11%) had multiple stop codons and were excluded from further analysis; none of the sequences
showed evidence of APOBEC hypermutation (data not shown). The majority of those sequences
with stop codons were isolated from infant tissues (p<0.05, Fischer’s exact test). In total, 528 env
sequences, spanning the signal peptide through env gp120 (sp-gp120), from nine IU and eight NT
participants are described.

CD4+ T-cell count and HIV-concentration have been correlated with viral population diversity and
IU MTCT, therefore it is plausible that the observed sequence differences between IU and NT
samples (see below) might reflect differences in progression or immunological status between the IU
and NT groups. To address this concern, we calculated the mean viral sequence diversity for each
pregnancy and measured its correlation with viral load or CD4+ T-cell counts. Maternal viral
population diversity increased in both IU and NT with declining CD4+ T-cell counts (p < 0.003,
Pearson’s R²), but not with rising viral load (Table 2). No significant relationship was observed
between transmission status and viral population diversity in either compartment nor between
maternal sequences stratified by transmission status (0.04±0.01 substitutions per nucleotide in the
IU maternal sequences and 0.03±0.03 substitutions per nucleotide in the NT maternal sequences).
Therefore, the sequence differences observed in the sequencing sub-cohort are unlikely to be a
proxy for disease progression or immunological status.

Regardless of transmission status, there was variability in the env length within the nine IU and eight
NT isolates (Table 2). First we compared the median-length of the hypervariable regions of gp120
(V1-V5) isolated from matched placental and maternal plasma. Within single participants, insertions
and deletions occurred most frequently in env V1. The median env V1-V5 length between
epidemiologically-linked sp-gp120 sequences were similar in six of eight NT isolates, but in seven of
the nine IU MTCT isolates, the env V1-V5 loop lengths were different (either longer or shorter).
These differences were most robust in the env V1 loop, where differences as great as twelve amino
acids between epidemiologically-linked samples were observed (Table 2). On average, env V1 loops
from matched IU MTCT maternal-peripheral and placental samples differ by four amino acids in
length, in spite of the epidemiological linkage. This is equal to the average difference between env V1
loop lengths from different mothers (not epidemiologically-linked) and much greater than the
average of one amino acid change in average env V1 loop lengths seen between NT maternal-
peripheral and placental sequences. Finally, this is in strong contrast to other regions of gp120; for
example, the entire region from the end of env V1 to the end of env C5 (“V2-C5 length” in Table 2),
which is over twenty times as long as env V1 on average, shows an average length-change of eight
amino acids between different mothers, but an average change of only three amino acids in length when comparing matched, IU MTCT maternal-periphery to placental sequences. The absolute change in average env V1 loop lengths is significantly greater in IU MTCT pregnancies than in NT pregnancies ($p < 0.02$, Wilcoxon’s rank sum test), but the absolute change in average env C2-V5 length did not differ significantly by transmission status ($p = 0.5$, Wilcoxon’s rank sum test). The contrast between env V1 length differences (high between tissues from the same participant, low between mothers) and env C2-V5 length differences (low between tissues from the same participant, high between mothers) indicates that the env V1 length differences observed are not simply an indirect consequence of overall sequence differences associated with a placental compartment. This suggests that over a short evolutionary time frame, the env V1 loop among IU MTCT isolates varies tremendously between the periphery and the placenta.

Consistent with the env V1 loop length data, differences in the number of potential N-linked glycosylation sites (PNLGs) in epidemiologically-linked plasma/placental isolates were observed in eight of nine IU MTCT samples, while NT pairs had a similar number of PNLGs in seven of eight samples (Table 2). The average number of PNLGs differed between the maternal-peripheral and placental compartments in each transmitting participant, but not in non-transmitting participants ($p < 0.008$ by Wilcoxon rank test). However, the change in PNLGS was closely correlated with the change in overall loop length (Pearson R of 0.81, significant to $p < 0.005$). In one extreme case, the placental variants from one IU MTCT pair (MHP1485) had nine additional PNLGs.

Overall, env V3 loop length was consistent between placenta and peripheral variants, and evidence of predicted CXCR4 (X4) tropism was identified in 100 of the 528 sequences (19%) and ten of the seventeen (59%) of the epidemiologically-linked isolates: four of eight NT and six of nine IU
mother-placenta pairs. In all cases, if predicted X4 tropism was found in the peripheral blood it was also found in the placental tissue. Infant samples were available for five cases of IU MTCT, and two infants had env sequences with a predicted X4-tropic phenotype.

Next, we used GARD to identify significant recombination breakpoints in the sp-gp120 sequences. The eight NT mothers had $3.0 \pm 2.0$ recombination breakpoints significant at $p<0.05$, while the nine IU mothers had $3.4 \pm 1.7$ recombination breakpoints at the same significance threshold. It is possible that recombination might interfere with the detection of a real compartment (discussed below). In this case, however, the IU mothers (in whom significantly more HIV-1 compartmentalization is detected) also show more recombination in sp-gp120, although the difference is not significant.

The phylogenetic trees produced by GARD were used to calculate dN-dS values and the dN and dS values were pooled among volunteers according to transmission status to identify amino acid positions / codons that are under statistically significant positive (dN-dS > 0) or purifying (dN-dS < 0) selection. Positions 279 and 230 in the IU transmitters were under significant purifying selection, with dN – dS values of -1.41 and -1.17 (HXB2 coordinates, false discovery rate < 0.05 and 0.25, respectively). In the NT group, positions 62 and 92 were also under purifying selection, with dN – dS values of -0.72 and -1.28 (HXB2 coordinates, false discovery rate < 0.25). Figure 2 shows that the IU positions under purifying selection are clustered in env C2 and the NT positions under purifying selection are in env C1.

**env Sequence Differences associated with transmission status.** We identified several characteristic sequence differences associated with MTCT, which are shared by the viral variants present in the IU MTCT participants regardless of their tissue of origin. In order to better
understand these differences, and determine the statistical significance of observed differences, we
converted the observed amino acids to a quantitative measurement and used a nested bootstrapping
method to measure differences in Kyte-Doolittle hydropathy at individual gp-120 amino acid
positions, stratified by mother-to-child-transmission status. Hydropathy was chosen over other
measures of amino acid physical properties because it is especially sensitive to amino acid changes
that alter the energetics of ligand-binding interactions, which are dominated by interactions between
the protein and solvent (23). Among the 120 amino acid positions with significant amino acid
variability, statistically-significant differences in hydropathy as a function of transmission status were
identified at gp-120 amino acid positions 6 and 365 (HXB2 coordinates, false discovery rate =0.05)
and amino acid positions 291, 305 and 335 (false discovery rate =0.25)[Figure 3]. The most frequent
amino acids at the individual positions, segregated by transmission status, are displayed in Table 3
using sequence logos (13, 69). For comparison, the most frequent amino acids among HIV-1
subtype C env genes currently deposited in the Los Alamos HIV-1 database are shown in the final
column (only one sequence per patient was included in the analysis). Note that although amino acids
at low frequency can produce a statistically significant result, the nested bootstrapping method
requires those rare amino acids to be present in all or most participants in order to produce a
significant p-value. Using the same false discovery thresholds, no significant differences in
hydropathy were found between sequences in the same transmission type but from different tissues;
this is in noted contrast to the V1 loop-length differences described above.

Placental Compartmentalization. The quantitative data presented in Figure 1 indicate that the
viral load in the placenta is not equivalent to that in the periphery. This result could indicate an
independent compartment of HIV-1 replication. In order to test the hypothesis that HIV-1 is
compartmentalized in the placenta during IU MTCT, both a tree-based (Slatkin-Maddison(75)) and a
distance-based (Hudson FST(31)) test of compartmentalization were performed on the sp-gp120 sequences isolated from the placenta and peripheral plasma (phylogenetic trees and highlighter plots shown as supplemental Figure S2). In seven of the eight NT pregnancies, both tests failed to support statistically significant compartmentalization (defined as p<0.05) between the placenta and the maternal periphery (Table 4). In contrast, six of the nine IU pregnancies showed significant compartmentalization (p<0.05) by either the tree-based Slatkin-Maddison test or the Hudson-FST nearest neighbor test. Discordant results between these tests of compartmentalization could be a result of the different number of sequences sampled in the two potential compartments, in which case the Hudson-FST is reportedly a more sensitive test of compartmentalization (84). Exclusion of duplicate sequences attenuated the significance of compartmentalization tests but it did not alter the general association between compartmentalization and IU MTCT. In addition, five of the nine groups of IU isolates (MHP ID# 1468, 1485, 1851, 2400, and 2797) and one of the NT isolates (MHP ID#2437) showed evidence of local viral replication in the placenta (supplemental Figure S2). Overall, the IU pregnancies showed more compartmentalization, as assessed by a Wilcoxon sum of ranks test on the p-values associated with each individual volunteer: p<0.0006 for Slatkin-Maddison with duplicates; p<0.02 for Slatkin-Maddison without duplicates; p<0.02 for Hudson-nearest-neighbor with duplicates; and, p<0.12 for Hudson-nearest-neighbor without duplicates.

DISCUSSION

This study was designed to test the hypothesis that selected HIV-1 variants cross the placental barrier and cause IU MTCT. After genotypically characterizing 528 env genes from matched peripheral and placental tissue of pregnant women with documented HIV-1 vertical transmission status, we observed the following: 1) tissue-specific differences in viral load, which varied by HIV-1 transmission status, 2) discordant env V1 lengths between peripheral and placental HIV-1 isolated
from cases of IU MTCT, 3) phylogenetic reconstruction consistent with the presence of a placental compartment, and 4) differences in env hydropathy within the env-CD4 binding interface that are associated with IU MTCT.

There are several limitations to these findings, including the following: a small sample size at the participant level, which limits the generalizability of our findings and a potential bias towards the characterization of the most abundant viral variants, which is an inherent limitation of single template amplification. In addition, HIV-1 isolated from the placenta can be a proxy for infant HIV present in the fetal vasculature within the placenta (57). Maternal-fetal blood admixture could confound the interpretation of the placental-specific observations, but the following observations argue against the placenta representing only infant blood. First, viral diversity in the placental biopsies did not differ by transmission status (p=0.8, students t-test). Second, infant sequences are generally homogenous and therefore unlikely to contribute significantly to viral diversity (55, 62). Third, compartmentalization tests comparing infant env sequences to placental env sequences indicated the presence of both a placental and an infant compartment in three of five mother-infant pairs. Nevertheless, even if these findings were to represent the characteristics of HIV-1 in the infant and not the placenta, these data are still important because they suggest features of env that could be associated with enhanced viral fitness in the immunologically-naïve newborn.

Although we observed genotypic differences between peripheral and placental HIV-1 isolates (varied env V1 loop lengths), and codon-specific differences in hydropathy between isolates from in utero transmitting versus nontransmitters, the environmental pressures that selected for these changes in env are unknown. Immunological pressures on viral replication are a likely selective force, and several of the observed differences in env in this report have previously been associated with
immune escape. For example, extension of V1 loop-length has been associated with neutralizing antibody escape, through a hypothesized “glycan shield” masking mechanism (65, 82). Our data do not show a consistent increase or decrease in env V1 loop length between anatomical compartments; rather, we observe that the env V1 loops are frequently different during IU MTCT. Amino acid 230, which is under purifying selection in the IU cases, is an established glycosylation site and amino acid 279 is within the N-linked glycosylation site pattern and could potentially influence the efficiency of glycosylation. In addition, one of the position with differential hydropathy value is located on the α2-helix of env C3 (HXB2 position no. 335), and this position is known target of autologous neutralizing antibodies (52) and likely relevant to neutralizing antibody escape (25, 64). Although the sensitivity of the env isolates from this study to antibody neutralization has not been tested, another study using samples from the MHP cohort failed to detect an altered sensitivity of in utero transmitted Env variants to autologous, heterologous, or the well-characterized monoclonal antibodies (E. Russell, J. Kwick, S. Meshnick, R. Swanstrom, et al., unpublished observation). It is currently unknown if alterations in these Env positions allows HIV-1 to escape other antibody-directed immune reactions, such as antibody-dependent cellular cytotoxicity (ADCC).

In addition to selective pressure from the immune system, the differences in env associated with either placental localization or in utero MTCT could also be a result of viral adaptation to the tissues of the placenta, which has a unique repertoire of cellular targets for viral infection. The materno-fetal placental barrier is comprised of several cell layers, including syncytiotrophoblasts and cytotrophoblasts, and in the absence of placental lesions, it is presumed that HIV-1 crosses through these cells during in utero MTCT. To date, there are no data published that support maternal-fetal blood admixture as a mechanism of in utero HIV-1 mother to child transmission (40, 41, 46). Thus, our working assumption is that HIV-1 must cross an intact maternal-fetal barrier during in utero
MTCT. Several studies have demonstrated HIV-1 infection of trophoblasts and syncytiotrophoblasts both in immortalized and primary trophoblast cells (4, 5, 15, 50, 72).

Paradoxically, it has also been reported that syncytiotrophoblast cells express either low-level or no CD4 or CCR5 molecules (18, 44), although this finding is inconsistent (14, 51). Trophoblast cells are known to express various C-type lectins like gelatins, DC SIGN, globoside, and syndecans on their surface (12, 33, 34), many of which can serve as an attachment receptor for HIV-1 in CD4-negative cell lines (8, 20, 28, 45, 81). Because $env\ V1$ length and glycosylation are important in several aspects of the viral life cycle, including the replication competence (77), tropism (9, 26) and cytopathology (37, 56, 77), it is possible that the observed differences in $env\ V1$ loops lengths between peripheral and placental compartments during cases of in utero-transmitting are a manifestation of placental adaptation, such as an increased adherence to C-type lectins.

After comparing sequences from participants stratified by transmission status, we found significant differences in hydropathy between IU and NT isolates at several amino acid positions. Hydrophobicity was chosen for this application in order to be maximally sensitive to mutations between polar and non-polar residues. However the mutations identified (see Table 3) include two relatively conservative K to R mutations. Although less conservative mutations are expected to have a stronger impact on the biochemistry of the protein product, this is not always the case -- even relatively conservative amino acid substitutions may have significant biological effects (36). Although they were not associated with placental localization, it is possible that the sp-gp120 $env$ positions associated with in utero MTCT are also a marker of placental adaptation. One of the five hydrophobicity changes (position 365) and one of the positions under purifying selection during IU transmission (position 279) are likely in direct contact with the CD4 molecule (Figure 4). In addition to its role in binding the CD4 molecule, it has been shown that the amino acid sequence of $env\ V3$
through C3 is similar to the chemokines that bind either CCR5 or CXCR4 (73). Perhaps the
observed changes in env associated with in utero mother-to-child transmission expand the repertoire
of co-receptors available to HIV-1 for entry into cells in the placenta or infant.

Finally, the observation that during in utero MTCT the placenta shows phylogenetic evidence of viral
compartmentalization is consistent with the hypothesis that HIV-1 has adapted to the placental
environment. HIV-1 compartmentalization has been described in semen (2, 11), brain (70), kidney
(49), and metastatic tissue (67) but limited, if any, HIV-1 compartmentalization has been observed in
breast milk (29, 30). Sites of HIV-1 infection in placenta that could produce a distinct compartment
include placental macrophages, trophoblasts, decidual mononuclear cells, and T-cell infiltrates in the
intervillous space and decidua (3, 4, 48). Placental macrophages (PMs) express lower levels of the
CD4, CXCR4, and CCR5 receptors than monocyte derived macrophages (MDMs) in the peripheral
blood compartment (22, 79). Though CD4 and coreceptor expression is low, viral integration does
not differ between PMs and MDMs, (24) suggesting that alternate coreceptors or enhanced affinity
may be involved in infection of placental macrophages.

In summary, this study supports viral compartmentalization in the placenta during IU MTCT and it
identified changes in the CD4-gp120 binding interface that correlate with IU MTCT. Our working
hypothesis is that these mutations alter viral tropism, allowing HIV-1 to efficiently infect cells in the
placenta. Whether this model is relevant for all HIV-1 subtypes, or just subtype C, which has been
shown to be transmitted in utero more frequently than HIV-1 subtypes A and D (61), remains to be
evaluated. Future studies addressing the phenotypic consequences of these findings could provide
insight into the mechanism of in utero HIV-1 mother-to-child transmission.
ACKNOWLEDGEMENTS

We are thankful for the participation of the Malawian women and their newborns; the Malaria and HIV in Pregnancy Cohort staff running the MHP study; Debbie Kamwendo for managing the MHP study; Paul Wilson for the viral load measurements; Debbie Knight for her logistical support; Joe Verducci of the OSU Department of Statistics, for assistance with the statistics; the anonymous reviewers for their helpful suggestions. This research was presented in part at the 16th Conference on Retroviruses and Opportunistic Infections, the 17th annual meeting on HIV Dynamics and Evolution, and the 2010 Ohio Collaborative Conference on Bioinformatics. This project was supported in part by NIH grants K99-HD056586 and R00-HD056586 to JJK; and R01-AI49084 and R21-AI065369 to SRM. SJR was supported by a Wellcome Trust Senior Research Fellowship, number 063215. We acknowledge that this material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and Office under grant number W911NF-05-1-0271 (DJ). This material is also based upon work partially supported by the OSU Mathematical Biosciences Institute (National Science Foundation grant 0635561) to SKH. We thank the Medical Center Information Services team of OSU and the Ohio Supercomputer Center for hosting computing clusters used in this study. The content of this article is solely the responsibility of the authors and it does not necessarily represent the official views of the funding agencies.

REFERENCES


26
27  


65. Sagar, M., X. Wu, S. Lee, and J. Overbaugh. 2006. Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. J. Virol. 80:9586-9598.


FIGURE LEGENDS.

Figure 1. Peripheral and placental HIV-1 RNA concentration varies according to both transmission status and tissue compartment. Each point represents a viral load measurement from an individual participant, with the horizontal line representing the median value. P-values represent results of Wilcoxon signed-rank test of log_{10} peripheral vs. log_{10} placental viral load, or a two-sample Wilcoxon rank-sum test of log_{10} placental viral loads of nontransmitters (NT) versus in utero transmitters (IU).

Figure 2. Several positions in Env show statistically-significant selection. Vertical axis shows the p values for each codon among all HIV-1 sp-gp120 sequences, stratified by transmission status (NT= O, IU= ▲); all significant positions are under purifying selection. P-values correspond to the probability of at-most the observed number of non-synonymous substitutions given the synonymous substitutions and amino acids observed (as estimated using HYPHY). These results account for any recombination breakpoints identified by GARD. The horizontal axis shows amino acid position within sp-gp120, which is numbered according to the HXB2 reference sequence. The dashed lines indicate the threshold for statistical significance at false discovery rates of 0.05 and 0.25. Shaded areas indicate the hypervariable (V) or conserved (C) domains of env, also according to the HXB2 reference sequence.

Figure 3. Several positions in Env are associated with in utero mother to child transmission. Amino acid positions where differences in Kyte-Doolittle hydropathy values at the individual positions of HIV-1 sp-gp120 are associated with transmission status. The p-values (vertical axis) shown correspond to the fraction of bootstrap replicates in which the sign of the observed
hydropathy difference is maintained. The dashed lines indicate the threshold for statistical
significance at false discovery rates of 0.05 and 0.25. Shaded areas indicate the hypervariable (V) or
conserved (C) domains of env, also according to the HXB2 reference sequence (horizontal axis).
Shaded areas indicate the hypervariable (V) or conserved (C) domains of env, also according to the
HXB2 reference sequence.

Figure 4. Location of the amino acids positions associated with in utero mother-to-child
transmission. HIV-1 JR-FL gp120 core (PDB 2B4C) is displayed in white, cartoon format
complexed with a fragment of the CD4-ligand, which is drawn in grey, spacefill format. Four of the
five amino acids correlated with in utero transmission are colored yellow (position 6, in the signal
peptide, is not included in the structure); the two positions with evidence of purifying selection
during in utero transmission are displayed in red. Amino acid positions are indicated in white text.
TABLE 1: Clinical features of HIV-infected women with peripheral and placental HIV-1 viral load data, stratified by HIV-1 mother-to-child transmission status

<table>
<thead>
<tr>
<th>Feature</th>
<th>Non-transmitter (n=266)</th>
<th>In utero transmitter (n=44)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Age (IQR)</td>
<td>24 (21,28)</td>
<td>24 (20,26)</td>
<td>0.2*</td>
</tr>
<tr>
<td>Median CD4+ T Cell Count (IQR)</td>
<td>378 (234,530)</td>
<td>293 (202,439)</td>
<td>0.2*</td>
</tr>
<tr>
<td>% Syphilis Seroreactivity (n)</td>
<td>5 (265)</td>
<td>18 (44)</td>
<td>0.002+</td>
</tr>
<tr>
<td>% with Placental Malaria (n)</td>
<td>8 (256)</td>
<td>1 (41)</td>
<td>0.09+</td>
</tr>
</tbody>
</table>

* Kruskal-Wallis equality-of-populations rank test
+ Fisher’s exact test

TABLE 2: Clinical and virological features of the participants contributing env sequences.

<table>
<thead>
<tr>
<th>MHP ID</th>
<th>MTC T Status</th>
<th>Peripheral VL (copies/mL)</th>
<th>Placental VL (copies/mL)</th>
<th>CD4+ T-cells (cells/mL)</th>
<th>no. env sequences</th>
<th>Average no. subs./position</th>
<th>median V1 length, (min., max.)</th>
<th>median V2-CS length</th>
<th>median predicted no. PNLGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1639</td>
<td>NT</td>
<td>160,259</td>
<td>NA</td>
<td>375</td>
<td>7</td>
<td>0.004</td>
<td>21 (20)</td>
<td>369</td>
<td>369</td>
</tr>
<tr>
<td>1669</td>
<td>NT</td>
<td>81,178</td>
<td>NA</td>
<td>485</td>
<td>5</td>
<td>0.015</td>
<td>14 (10,14)</td>
<td>356</td>
<td>356</td>
</tr>
<tr>
<td>1702</td>
<td>NT</td>
<td>NA</td>
<td>NA</td>
<td>273</td>
<td>12</td>
<td>0.017</td>
<td>19 (17,33)</td>
<td>357</td>
<td>357</td>
</tr>
<tr>
<td>2437</td>
<td>NT</td>
<td>13,529</td>
<td>NA</td>
<td>445</td>
<td>10</td>
<td>0.019</td>
<td>19 (17,19)</td>
<td>361</td>
<td>361</td>
</tr>
<tr>
<td>2502</td>
<td>NT</td>
<td>154,972</td>
<td>NA</td>
<td>38</td>
<td>12</td>
<td>0.070</td>
<td>20 (13,20)</td>
<td>372</td>
<td>372</td>
</tr>
<tr>
<td>2512</td>
<td>NT</td>
<td>23,801</td>
<td>NA</td>
<td>222</td>
<td>7</td>
<td>0.040</td>
<td>21 (20,23)</td>
<td>358</td>
<td>358</td>
</tr>
<tr>
<td>2544</td>
<td>NT</td>
<td>587,702</td>
<td>NA</td>
<td>199</td>
<td>19</td>
<td>0.018</td>
<td>17 (17)</td>
<td>360</td>
<td>360</td>
</tr>
<tr>
<td>3274</td>
<td>NT</td>
<td>NA</td>
<td>NA</td>
<td>105</td>
<td>13</td>
<td>0.073</td>
<td>15 (9,25)</td>
<td>361</td>
<td>361</td>
</tr>
<tr>
<td>1468</td>
<td>IU</td>
<td>36,453</td>
<td>87,276</td>
<td>274</td>
<td>17</td>
<td>0.035</td>
<td>12 (12,17)</td>
<td>356</td>
<td>356</td>
</tr>
<tr>
<td>1472</td>
<td>IU</td>
<td>51,551</td>
<td>24,116</td>
<td>397</td>
<td>16</td>
<td>0.028</td>
<td>17 (13,19)</td>
<td>370</td>
<td>370</td>
</tr>
<tr>
<td>1485</td>
<td>IU</td>
<td>125,813</td>
<td>708,115</td>
<td>329</td>
<td>7</td>
<td>0.051</td>
<td>21 (20,31)</td>
<td>364</td>
<td>364</td>
</tr>
<tr>
<td>1646</td>
<td>IU</td>
<td>902,778</td>
<td>NA</td>
<td>95</td>
<td>14</td>
<td>0.050</td>
<td>17 (16,24)</td>
<td>364</td>
<td>364</td>
</tr>
<tr>
<td>1851</td>
<td>IU</td>
<td>NA</td>
<td>NA</td>
<td>761</td>
<td>6</td>
<td>0.015</td>
<td>15 (8,11)</td>
<td>355</td>
<td>355</td>
</tr>
<tr>
<td>2080</td>
<td>IU</td>
<td>165,000</td>
<td>NA</td>
<td>91</td>
<td>14</td>
<td>0.052</td>
<td>15 (14,38)</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>2460</td>
<td>IU</td>
<td>98,793</td>
<td>NA</td>
<td>439</td>
<td>16</td>
<td>0.043</td>
<td>15 (14,14)</td>
<td>349</td>
<td>349</td>
</tr>
<tr>
<td>2797</td>
<td>IU</td>
<td>NA</td>
<td>NA</td>
<td>399</td>
<td>14</td>
<td>0.041</td>
<td>17 (17,22)</td>
<td>345</td>
<td>345</td>
</tr>
<tr>
<td>3321</td>
<td>IU</td>
<td>NA</td>
<td>NA</td>
<td>220</td>
<td>8</td>
<td>0.065</td>
<td>24 (11,38)</td>
<td>357</td>
<td>357</td>
</tr>
</tbody>
</table>
TABLE 3: Changes in hydropathy at several Env positions are correlated with *in utero* mother-to-child transmission. Shown are the individual amino acid positions, according to HXB2 numbering, and the corresponding subdomain of Env where that position is found (SP= signal peptide; V=variable; C=conserved). Change in mean hydropathy indicates the difference in the bootstrap-mean estimate of the Kyte-Doolittle hydropathy at this position, stratified by transmission status: a positive value indicates more hydrophobic residues in *in utero* mother to child transmission (IU), a negative value indicates more hydrophobic residues in the nontransmitters (NT). P-values represent the raw (not multiple-hypothesis corrected) correlation between hydropathy at this position and transmission status. Most frequent amino acids are displayed using sequence logo, whereby the size of the letter is proportional to its frequency. The final column indicates the most frequent amino acids in 757 HIV-1 subtype C sequences in the Los Alamos HIV database.

<table>
<thead>
<tr>
<th>HXB2 Env AA #</th>
<th>Env subdomain</th>
<th>difference in mean hydropathy (NT to IU)</th>
<th>Most frequent amino acid(s), as sequence logo</th>
<th>p-value for hydropathy correlated to transmission status</th>
<th>Most frequent amino acids in the Los Alamos HIV database (subtype C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>SP</td>
<td>+3.2</td>
<td>![Sequence logo for SP]</td>
<td>&lt; 0.0001</td>
<td>![Sequence logo for SP]</td>
</tr>
<tr>
<td>291</td>
<td>C2</td>
<td>-0.6</td>
<td>![Sequence logo for C2]</td>
<td>0.004</td>
<td>![Sequence logo for C2]</td>
</tr>
<tr>
<td>305</td>
<td>V3</td>
<td>-0.2</td>
<td>![Sequence logo for V3]</td>
<td>0.005</td>
<td>![Sequence logo for V3]</td>
</tr>
<tr>
<td>335</td>
<td>C3</td>
<td>-1.8</td>
<td>![Sequence logo for C3]</td>
<td>0.008</td>
<td>![Sequence logo for C3]</td>
</tr>
<tr>
<td>365</td>
<td>C3</td>
<td>-0.7</td>
<td>![Sequence logo for C3]</td>
<td>&lt; 0.0001</td>
<td>![Sequence logo for C3]</td>
</tr>
</tbody>
</table>
TABLE 4: Evidence for placental compartmentalization of HIV-1 env sequences. Results of a
tree-based (Slatkin-Maddison) and distance-based (Hudson Nearest Neighbor [NN]) test of
compartmentalization. Statistically-significant compartmentalization (p<0.05) is highlighted in gray.
MHP# indicates the participant’s unique number; NT, nontransmitter; IU, in utero mother-to-child
transmission; Dup. indicates whether or not duplicate sequences were included in the analysis. na=
not available.

<table>
<thead>
<tr>
<th>MHP #</th>
<th>MTCT Status</th>
<th>Compartmentalization Test, maternal vs. placental (p-value)</th>
<th>Compartmentalization Test, child vs. placental (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Maddison-Slatkin</td>
<td>Hudson-NN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dup.</td>
<td>No Dup.</td>
</tr>
<tr>
<td>1639</td>
<td>NT</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>1669</td>
<td>NT</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1702</td>
<td>NT</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>2437</td>
<td>NT</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>2502</td>
<td>NT</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>2512</td>
<td>NT</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>2544</td>
<td>NT</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>3274</td>
<td>NT</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1468</td>
<td>IU</td>
<td>0.009</td>
<td>0.02</td>
</tr>
<tr>
<td>1472</td>
<td>IU</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>1485</td>
<td>IU</td>
<td>&lt;0.001</td>
<td>0.04</td>
</tr>
<tr>
<td>1646</td>
<td>IU</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1851</td>
<td>IU</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>2080</td>
<td>IU</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>2400</td>
<td>IU</td>
<td>&lt;0.001</td>
<td>0.02</td>
</tr>
<tr>
<td>2797</td>
<td>IU</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3321</td>
<td>IU</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Log$_{10}$ HIV-1 copies/mL

- NT Peripheral, (n=266)
- NT Placenta, (n=130)
- IU Peripheral, (n=44)
- IU Placenta, (n=28)

- p<0.0001
- p=0.2
Figure 2
Figure 3

HXB2 position
log p-value
0 200 400
0.00001
0.0001
0.001
0.01
0.1
1
FDR < 0.05
FDR < 0.25

SP C1 V1-V2 C2 V3 C3 V4 C4 V5 C5

HXB2 position
log p-value
0.00001
0.0001
0.001
0.01
0.1
1
FDR < 0.05
FDR < 0.25