Single Genome Analysis Reveals Genetic Characteristics of Neuroadaptation Across HIV-1 Envelope

Teresa H. Evering, M.D. M.S.[1], Leslie St. Bernard[1], Charles Farmer[1] and Martin Markowitz, M.D.[1]

[1]Aaron Diamond AIDS Research Center, an affiliate of the Rockefeller University, New York, New York, United States of America

BACKGROUND:

The prevalence of HIV-1 associated neurocognitive disorders (HANDs) has been estimated to be 45-50% in the CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER) study (Heaton et al., 2010). In the cART era, this disease burden is largely represented by asymptomatic and mild neurocognitive disease (ANI and MND). We hypothesized that detailed analyses of full-length HIV-1 *env* (>2.5Kb) would allow for the identification of genetic characteristics associated with the presence of viral variants in the CNS.

METHODS :

Research Design: We used single genome amplification to generate full-length HIV-1 *env* clade B variants from the paired CSF and plasma samples of 15 chronically infected HIV+ subjects from the CHARTER study with normal neurocognitive performance (NCN) (N=6), ANI (N=7) and MND (N=2). Subjects were viremic and either treatment naïve (N=8) or experienced (N=7).

Experimental Methods: Confirmed single genome sequences (SGS) were then subjected to phylogenetic analysis to determine viral compartmentalization, potential N-linked glycosylation sites (PNLGS) and genetic diversity. We identified patterns of compartmentalization with the Viral Epidemiology Signature Pattern Analysis software (Korber et al., 1992) and the CorMut package (Li et al., 2012) was used to detect correlated mutations.

RESULTS:

Subj.	NCN1	NCN2	NCN3	NCN4	NCN5	NCN6	ANI1	ANI2	ANI3	ANI4	ANI5	ANI6	ANI7	MND 1	MND 2
Age	52	41	42	41	45	43	61	51	42	44	37	34	62	40	32
Sex	M	M	M	M	M	M	M	M	M	M	M	M	M	M	F
Race	White	Black	Black	White	Black	White	White	White	Black	White	Hispanic	White	White	Black	Black
Est. Dur. Inf	59.6	173.1	48.9	214.4	75.3	136.9	154.1	211.7	46.2	56.2	26.5	59.8	206.6	75.0	58.5
ARV Status	No ARVs	Naïve	No ARVs	Naïve	Naïve	No ARVs	Naïve	No ARVs	Naïve	Naïve	Naïve	No ARVs	No ARVs	Naïve	No ARVs
CD4 Count	591	802	687	790	420	418	743	220	819	237	244	329	466	398	320
CD4 Nadir	400	554	491	461	297	392	537	220	495	237	244	329	215	398	320
CDC Stage	A2	A1	A2	B2	B2	A2	A1	A2	A2	A2	A2	A2	A2	B2	A2
Plasma VL	65,600	749	8,620	13,900	19,700	18,900	5,320	29,800	1,890	38,600	9,210	13,000	50,100	6,250	44,900
CSF VL	120	120	6,600	14,700	18,300	2,910	3,210	5,930	129	1320	653	2430	4,330	454	2,690
RPR	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
HCV	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
Date	5/27/04	2/2/06	7/22/05	1/24/06	1/23/07	5/17/07	7/20/04	11/5/07	1/2/07	6/28/06	2/23/07	10/12/05	5/4/05	8/14/07	3/6/07

Table 1. Clinical and Demographic Profiles for Study Participants

Participant neurocognitive disease classifications are as follows: NCN=Normal neurocognitive performance (NCN); ANI=Asymptomatic neurocognitive impairment; MND=Mild neurocognitive disorder. M=male, F=female. Estimated Duration of Infection (Est. Dur. Inf) is shown in months. CD4+ T Cell Count (CD4 Count) is in cells/mm^3, plasma and CSF HIV-1 RNA levels (VL) are in copies/mL. RPR = rapid plasma reagin. HCV = hepatitis C virus.

Groups	NCN	ANI + MND	*p*-value
# Male Sex/total	6/6	8/9	N/A
# Men who have sex with men/total	5/6	7/9	N/A
Mean Age (years, range)	44 (41-52)	45 (32-62)	0.72
Mean Years of Education (years, range)	13 (10-16)	14 (9-18)	0.31
Mean Est. duration of infection (months, range)	118 (49-214)	99 (27-212)	0.46
Mean CD4+ T cell count (cells/mm^3 , range)	618 (418-802)	420 (220-819)	0.09
Mean CD4+ T cell nadir (cells/mm^3, range)	433 (297-554)	333 (215-537)	0.15
Mean plasma HIV-1 RNA (log copies/ml, range)	4.3 (2.9-4.8)	4.3 (3.3-4.7)	1.00
Mean CSF HIV-1 RNA (log copies/ml, range)	3.9 (2.1-4.3)	3.4 (2.1-3.8)	0.52

Table 2. Statistical Comparison of Key Clinical and Demographic Parameters Across Study Groups

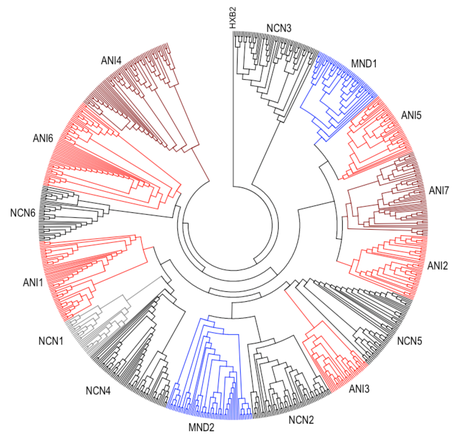


Figure 1. Intra-Patient Clustering of HIV-1 *env* quasi-species

Maximum Likelihood (ML) topology view tree depicting 717 full-length HIV-1 *env* sequences from fifteen experimental subjects are shown. For each subject, all sequences from both compartments (plasma and CSF) are shown. Each subject forms a tight cluster and is distinct from other experimental subjects with aLRT SH-like supports >95% for all inter-subject clusters. HXB2 was used as an outgroup.

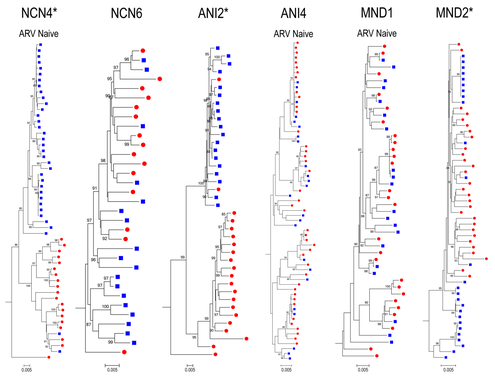


Figure 2. HIV-1 *env* phylogenies - Varying Degrees of Intra-Patient Sequence Diversity and Compartmentalized Virus Across Disease States

Representative ML trees of SGA sequences from select participants. For all panels, CSF (closed blue squares) and PBMC (closed red circles) are shown. aLRT SH-like supports were determined and values over 85% are shown. HIV-1 *env* sequences with statistical evidence of hypermutation were excluded. All scale bars represent 0.005 nucleotide substitutions per site. HXB2 was used as an outgroup. *= Phylogram with statistical evidence of viral compartmentalization.

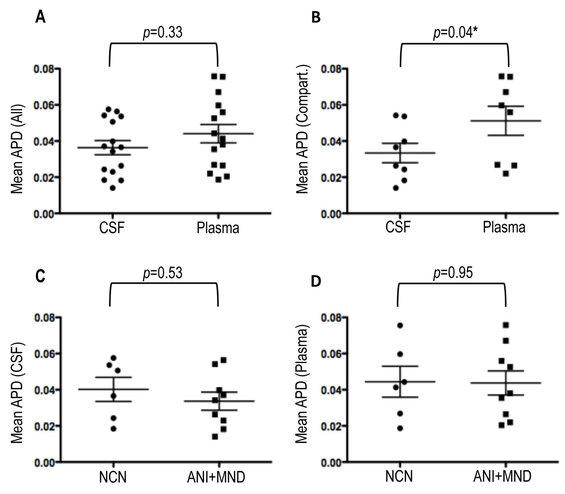


Figure 3. Amino-acid diversity is lower across compartmentalized CSF-derived full-length HIV-1 *env*

Mean Average Pairwise Distance determinations between the (A) paired CSF and plasma HIV-1 *env* variants of all participants (n=15) and (B) CSF and plasma HIV-1 *env* variants of participants with statistical evidence of CSF viral compartmentalization (n=8) (C) CSF of individuals with NCN (n=6) and those with neurocognitive disease (ANI+MND) (n=9) and (D) plasma of individuals with NCN (n=6) and those with ANI or MND (n=9). All calculations were performed after the exclusion of sequences with statistical evidence of hypermutation.

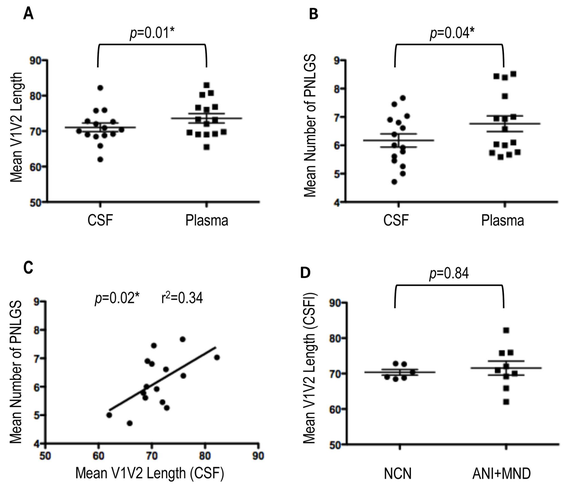


Figure 4. Mean V1V2 loop length is shorter and the mean number of PNLGS is lower in CSF-derived variants compared to paired plasma counterparts

The mean (A) amino-acid lengths and (B) number of potential N-linked glyosylation sites (PNLGS) of the HIV-1 *env* V1V2 region from paired CSF and plasma quasispecies are shown for all subjects (n=15). (C) Linear regression of the mean amino-acid lengths of the HIV-1 *env* V1V2 region from CSF quasispecies with the corresponding mean number of PNLGS in the same quasispecies. (D) The mean amino-acid lengths of the HIV-1 *env* V1V2 region in the CSF quasispecies of individuals with NCN and those with ANI+MND.

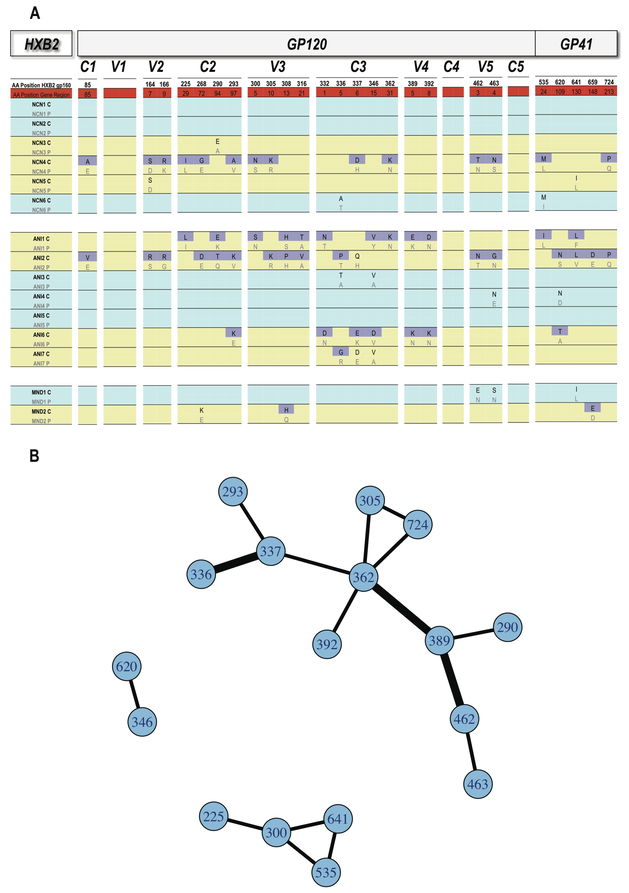


Figure 5. Compartmentalization "hot spots" across full-length HIV-1 *env* and the Network of Correlated Mutations

Composite results of analyses identifying positions (hot spots) at which the dominant amino acid in an individual's CSF and plasma alignments are statistically different (A). All positions listed are significant in > 2 individuals. Subjects with statistically non-compartmentalized (blue) and compartmentalized (yellow) CSF quasispecies are shown. Amino acids present at hot spot positions in the CSF (C) alignments that meet the threshold for statistical significance are shaded in purple. The corresponding plasma (P) amino acids are shown directly below. (B) An adjacency matrix demonstrating compartmentalization hot spots positions with amino acids sharing mutual information (MI) in > 2 subjects with an MI threshold of > 0.10 and Benjamini–Hochberg adjusted p-value for the correlation < 0.05. Positions sharing MI are linked by connecting lines, with the thickness of the line corresponding to the number of studied subjects in which the correlation was identified.

CONCLUSIONS:

Detailed analyses of SGS-derived full length HIV-1 *env* from subjects with the most common HAND diagnoses in the cART era allowed us to confirm and extend current knowledge of the genetic characteristics associated with HIV-1 *env* neuroadaptation.

ACKNOWLEDGEMENTS:

We would like to thank the individuals who participated in the research and gratefully acknowledge the CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER) study researchers and staff for the provision of de-identified, cryopreserved participant samples for study. CHARTER is supported by award N01MH22005 from the National Institutes of Health. THE is the recipient of an NIMH Career Development Award NIH K08MH090900.