学位論文
Doctoral Thesis

Impact of HLA class I-associated genetic variability in HIV-1 accessory gene *vpu*

（HLA クラス I が HIV-1 アクセサリー遺伝子 vpu の多型性に及ぼす影響）

ハッサン　ムハンマド　ザフルール
Hasan Md. Zafrul

Hasan Md. Zafrul
熊本大学大学院医学教育部博士課程医学専攻
エイズ先端研究者育成コース

指導教員

上野　貴将　准教授
熊本大学大学院医学教育部博士課程医学専攻エイズ学Ｖ分野

２０１２年９月

学位論文
Doctoral Thesis

Impact of HLA class I-associated genetic variability in HIV-1 accessory gene *vpu*

（HLA クラス I が HIV-1 アクセサリー遺伝子 vpu の多型性に及ぼす影響）

ハッサン　ムハンマド　ザフルール
Hasan Md. Zafrul

熊本大学大学院医学教育部博士課程医学専攻
エイズ先端研究者育成コース

指導教員　：　熊本大学大学院医学教育部
博士課程医学専攻エイズ学Ｖ分野　　上野　貴将　准教授

審査委員名 ：

審査委員長	松下　修三　教授
審 査 委 員	西村　泰治　教授
審 査 委 員	岡田　誠治　教授
審 査 委 員	前田　洋助　准教授

２０１２年９月

TABLE OF CONTENTS

1. ABSTRACT

[Objective] HLA class I (HLA-I) restricted CTL responses drive HIV evolution through selection of sequence polymorphisms and represent a major selective force toward HIV-1 proteins such as Gag and Nef. Now-a-days it is well known that an accessory protein Vpu acts as crucial enhancer in HIV-1 pathogenesis. Although Vpu represents one of the most variable proteins in the HIV-1 proteome, it is still elusive to what extent HLA-I influence its evolution.

[**Methods]** The *vpu* genes were amplified by nested-PCR using plasma viruses isolated from HLA-I-typed, treatment-naive, chronically-infected individuals in Japan (n=240). The statistical analysis was performed with a phylogenetically-informe d method incorporating the effects of HIV codon co-variation and linkage disequilibrium among HLA-I alleles. Multiple tests were addressed using false discovery rate ($q<0.2$).

[**Results & Discussion**] We successfully obtained *vpu* sequence from 216 out of 240 samples tested. Most codons of Vpu displayed substantial variability, with the average entropy score reaching 0.58.The pattern of amino acid variability was consistent with those observed in HIV-1 subtype B.We only identified 4 different significant HLA-HIV amino acid associations from 3 codons of Vpu, from primary and alternative reading frames (ARFs) of Vpu, suggesting that HLA-I had minor effects on Vpu variability. A mutation arginine (R) to lysine (K) being significantly enriched in subjects having HLA-A*33:03 at position 37 of Vpu, one of the highly immune-dominant epitopic region. Remarkably, we have identified a non-synonymous mutation in Env while the corresponding position is synonymous in Vpu in patients having HLA-B*40:01. However, despite its small size (81 amino acids), Vpu showed 103 codon-codon associations, suggesting that conformation and function may be preserved through many possible combinations of primary and secondary polymorphisms. Noticeably, we also identified a statistically significant association between amino acid residues at position 5 with plasma viral load and therefore it would be interesting to examine further functional effects of amino acid polymorphisms at position 5.

[**Conclusion**] Taken together, we conclude that the influence of HLA-I alleles on Vpu evolution at the population level showed lesser extent compared to other highly variable HIV-1 accessory proteins, providing us with additional insight into differential evolutional pathways among viral accessory proteins.

2. PUBLICATION LIST

I Reference Publications

Zafrul Hasan, Jonathan M. Carlson, Hiroyuki Gatanaga, Anh Q. Le, Chanson J. Brumme, Shinichi Oka, Zabrina L. Brumme, Takamasa Ueno. Minor contribution of HLA class I-associated selective pressure to the variability of HIV-1 accessory protein Vpu. ***Biochem Biophys Res Commun*.** 421:291-295, 2012

II Other Publications

Mwimanzi P, **Hasan Z,** Hassan R, Suzu S, Takiguchi M, Ueno T. Effects of naturally-arising HIV Nef mutations on cytotoxic T lymphocyte recognition and Nef's functionality in primary macrophages. Retrovirology. 22; 8:50, Jun 2011.

Mwimanzi P, **Hasan Z,** Tokunaga M, Gatanaga H, Oka S, Ueno T. Naturally arising HIV-1 Nef variants conferring escape from cytotoxic T lymphocytes influence viral entry co-receptor expression and susceptibility to superinfection. Biochem Biophys Res Commun. 403(3-4):422-7, Nov 2010.

Rahman M, Hassan ZM, **Zafrul H,** Saiada F, Banik S, Faruque AS, Delbeke T, Matthijnssens J, Van Ranst M, Azim T. Sequence analysis and evolution of group B rotaviruses. Virus Res. 125(2):219-25, May 2007.

3. ACKNOWLEGEMENTS

First of all, I express my gratitude in the most humble way to Almighty Allah for enabling me to complete my PhD work and successfully submit my thesis paper in time.

The person, to whom I owe the greatest debt of gratitude and sincere respect, is my respectable teacher and supervisor, Dr. Takamasa Ueno, Associate Professor, Center for AIDS Research, for his enthusiastic guidance, affectionate inspiration, passionate supervision and endearing company all through my PhD tenure herein Kumamoto University, Kumamoto; Japan.

I express my earnest sense of gratification, deep appreciation and thanks to Dr. Masafumi Takiguchi for his help and encouragement throughout the study.

I would like to extend my gratitude to "MEXT" for granting me the scholarship to do my PhD and this thesis work herein Japan.

I am most grateful to Dr. Zabrina Brumme, Simon Fraser University, Burnaby, BC, Canada, for her continuous support and valuable advices during this thesis work and this work would not be achievable without her support and guidance. I feel greatly honored to have such a collaborative work with her.

I would like to thank Dr. Jonathan M. Carlson, Microsoft Research, and Los Angeles, CA, USA for his computational analysis.

I would like to great thank to Dr. Chihiro Motozono and Dr.Philip Mwimanzi for their patience and proper guidance when I started my work in a totally new environment and make me mastered in laboratory techniques. I would like to thank my course adviser Dr. Yousuke Maeda and Dr. Seiji Okada, for their kind comments and suggestions to improve my study. I would like to thank Dr. Hiroyuki Gatanaga and Dr. Shinichi Oka in the International Medical Center of Japan for providing study materials.

My sincere thanks go to my dear colleagues Dr. Tomohiro Akahoshi, Dr. Keiko Sakai, Dr. Masao Hashimoto, Dr Saito Masumichi, Dr. Takuya Naruto, Dr. Yoshinori Sato, Dr. Kazutaka Honda, Dr. Nozomi Kuse, Takayuki Chikata, Meribe, Macdonald and all laboratory members for their nice and generous help during my PhD work in Kumamoto.

I would like to thank Thuy Ngo, Fuyo Koutaki, Hinako Motoyama, Sachiko Sakai, and Keiko Urata for kind cooperation and successfully complete all the documents for PhD work.

Finally, I would like to express my sincere gratitude to all my family members, all friends, and my wife for their moral support, inspiration and encouragement throughout the study.

4. ABBREVIATIONS

AA	**:**	Amino acid
AIDS	**:**	Acquired Immunodeficiency Syndrome
BST-2	**:**	Bone marrow stromal cell antigen 2
CTL	**:**	Cytotoxic T lymphocyte
HAART	**:**	Highly Active Anti-retroviral Therapy
HLA	**:**	Human leukocyte antigen
HIV	**:**	Human Immunodeficiency Virus
HOMER	**:**	HAART Observational Medical Evaluation and Research
IHAC	**:**	International HIV Adaptation Collaborative
MHC	**:**	Major histocompatibility complex
LTNP	**:**	Long Term Non-Progressor
PCR	**:**	Polymerase chain reaction
IHAC	**:**	International HIV Adaptation Collaborative’
PDNs	**:**	Phylogenetic Dependency Networks
Nef	**:**	Negative Regulatory Factor
SIV	**:**	Simian Immunodeficiency Virus
TCR	**:**	T-cell receptor
VPU	**:**	Viral protein U
VIF	**:**	Viral infectivity factor
VPR	**:**	Viral protein R

5. BACKGROUND AND OBJECTIVE

5-1) HIV/AIDS and Accessory Genes

The hurdle of HIV/AIDS has puzzled scientists ever since the illness first came to light in the early 1980s. HIV is a lentivirus, and like all viruses of this type, it attacks the host immune system and cause Acquired Immunodeficiency Syndrome (AIDS). It is well documented that HIV is a descendant of a Simian Immunodeficiency Virus (SIV) because certain strains of SIVs bear a very close resemblance to HIV-1 and HIV-2, the two types of HIV [1]. HIV-1, the most virulent strain of HIV has been the major cause of global pandemic, whereas the less virulent HIV-2 is largely limited to the west of Africa [2-4]. It is estimated that there are 34 million people worldwide living with HIV/AIDS, with 3.4 million new HIV infections per year and 2.6 million annual deaths due to AIDS as of 2010 [5].

Primate lentiviruses encode an array of ‘‘accessory *vif, vpr, vpu* and *nef* genes” in addition to the prototype retroviral *gag, pol,* and *env*, which was coined by the finding that their inactivation resulted in little or no impairment of virus replication in continuous cell lines [6-8]. In contrast, Bour et al, demonstrated that accessory gene products can dramatically change the course and severity of the viral infection *in vivo* or in primary cell types which were susceptible to HIV infection [9]. Now-a-days HIV-1 accessory genes especially *vif, vpr, vpu* and *nef* are well known to function as crucial enhancers of viral pathogenesis, acting as a modifier of local environment within infected cells to ensure viral persistence, replication, dissemination, and transmission [8, 10-14]. In particular, *vpu* gene was specifically acquired in the lentiviral lineage that gave rise to HIV-1 and it is also interesting to note that closely related lentiviruses such as HIV-2 and SIV with less severe pathogenesis and disease outcome, lack Vpu expression in their genome [13, 15].

It is important for a pathogen to create mutation in its genome, enabling it to escape from host immune responses and persist within host system. During zoonotic transmission, pathogens

may evolved some advantageous genes for its survival (e.g., *vpu* gene of HIV-1) and those new genes may also need to adapt in current host and eventually polymorphism occurs [16]. As a result of evolution process, viruses harbor genetic polymorphisms as a footprint from host mediated-immune pressure they had taken in the past. Improvement in DNA sequence technologies and the availability of HIV-infected population allows us to see these immune footprints which could be predictable based on the viral sequences and the immune component of host via contemporary bio-statistical approaches [17-18]. Therefore, it has been suggested that a comprehensive analysis of the polymorphisms in HIV proteins would have worthwhile implications to see its functionality in HIV-1 replication as well as in viral pathogenesis, which may be indispensable towards developing antiviral therapeutics and vaccines [17, 19-22].

5-2) Human Leukocyte Antigen and Diversity

The human leukocyte antigen (HLA) genes are the human versions of major histocompatibility complex (MHC), which is found in most vertebrates. HLA is one of the highly polymorphic and most intensely studied genes in human genome and the diversity of HLAs in the human population is one of the main aspect of immune defense, as this super locus contains a large number of genes related to immune function [23].

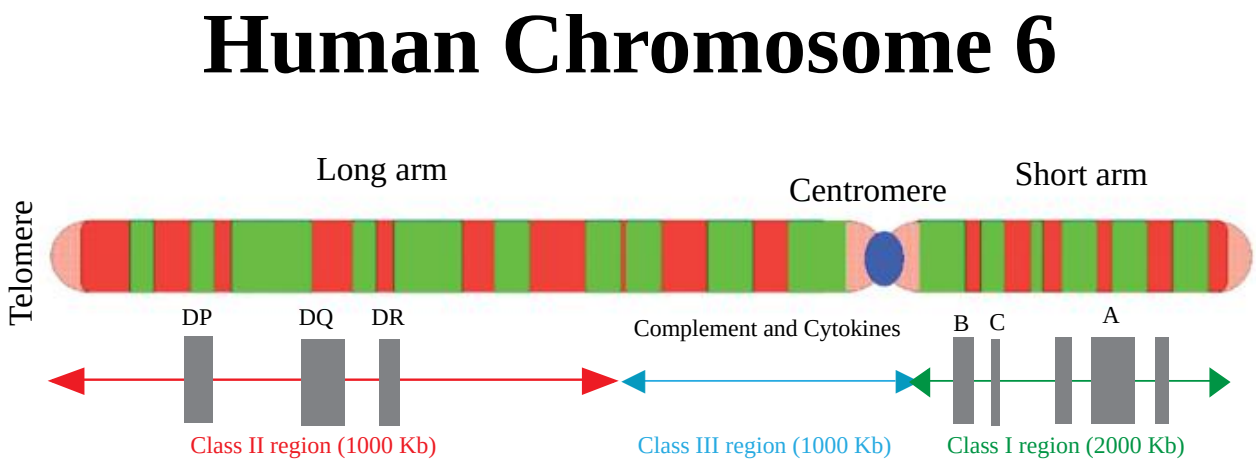

Figure A: **HLA alleles on human chromosome 6.** HLA-A, B, and C represent the three MHC class I α subunit, while HLA-DR, DQ, and DP are MHC class II alleles. Complement and cytokine production from MHC class III molecules

This group of genes resides on chromosome 6, and encodes cell-surface antigen-presenting proteins known as HLAs. In virus-infected cells, antigenic peptides that are processed from viral proteins via the proteasome pathway, bound to MHC-I (HLA class I) molecules, and presented on the cell surface. CTLs recognize antigenic peptide (epitope)-MHC-I complexes on the cell surface by their T cell receptors (TCR) and eliminate the virus-infected cells [24].

.

5-3) Role of CD8+ T-cell and HLA-allele to control HIV-infection

The host immune system detects a wide variety of agents, from viruses to parasitic worms, and may need to distinguish "self" and "non-self" cells or tissues in order to function properly. The immune system can be divided into (i) Innate and (ii) Adaptive immune system; where the innate one provides an immediate but non-specific response. In contrast, adaptive system is able to mount specific responses to pathogen and in such a way eliminate or prevent pathogenic growth in host environment [25-26]. One important arm of adaptive immunity is cytotoxic T-cells, (CTLs or CD8+ T-lymphocytes or killer T-cells) and therefore a key focus of vaccine development efforts towards viruses like HIV-1 [27].

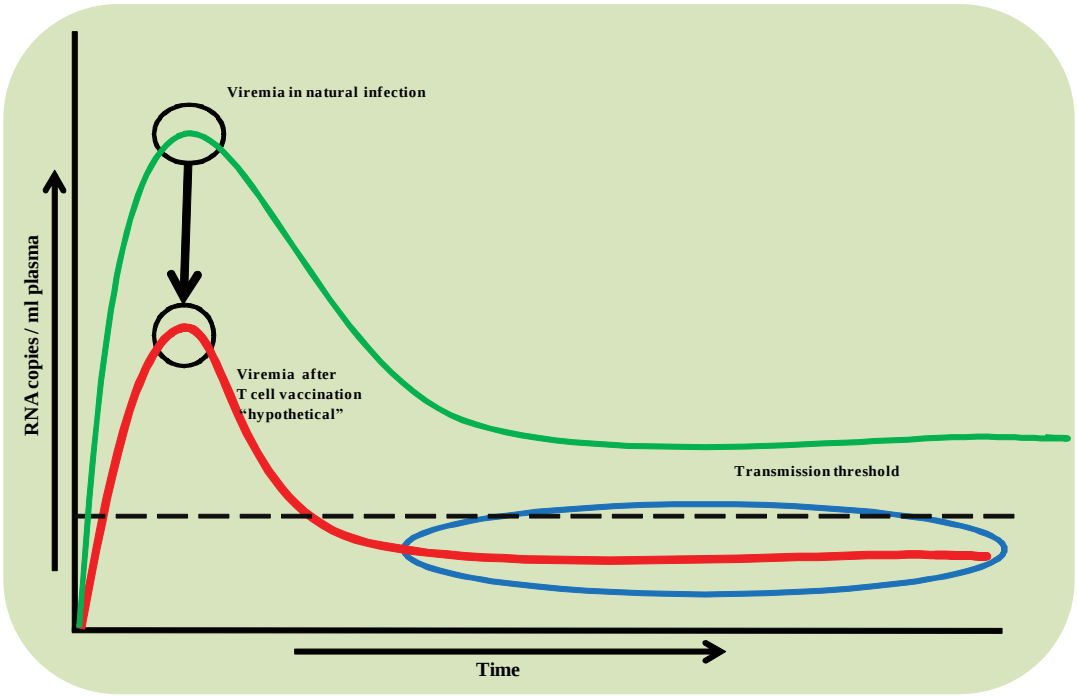

Figure B: A hypothetical proposed "Dogma" regarding HLA mediated CTLs activity toward HIV-vaccination. Figure adapted from "Toward an AIDS Vaccine" by Walker, et al., 2008: Science

Now-a-days, it is well known that host HLA class I-restricted CD8+ cytotoxic T-lymphocyte (CTL) responses play an important role in controlling of HIV replication *in vivo*, evident from the collective studies regarding T-cell immunity and HIV/AIDS [27-29]. Also, the growing numbers of supportive evidence indicated that strong HIV-1 specific CTL activity is detected in long term nonprogressors (LTNP), whereas rapid progressors have low HIV-I-specific CTL activity,

suggesting that HIV-1 infection and the progression to AIDS can be controlled by the maintenance of HIV-1 specific CTL activity [30-32].

In addition, during the natural course of infection, the initial peak of viral replication after primary HIV infection begins to decline simultaneously with the appearance of HIV-specific CD8+ T lymphocytes [33-34]. As a consequence of appeared CTLs, which can eliminate HIV-infected cells directly by MHC-I-restricted cytolysis or indirectly through the production of soluble factors such as cytokines and chemokines [35-36]. Indeed, the biological relevance of HIV-specific CTLs in HIV infection is also supported by *in vivo* studies which demonstrated a dramatic rise of viremia and an accelerated clinical disease progression in SIV-infected macaques, after artificial depletion of CD8+ T-cells [37-38].

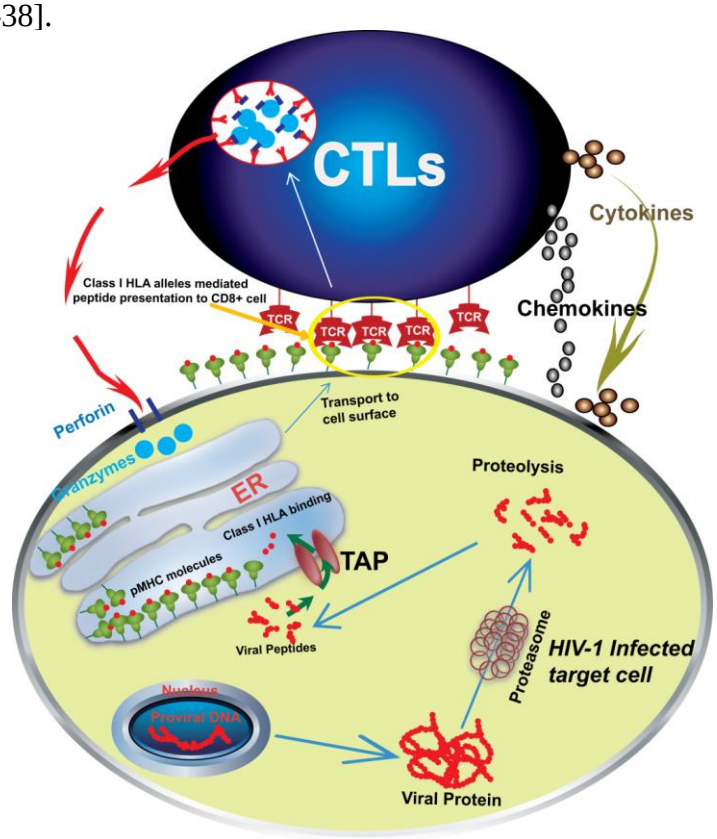

Figure C: MHC class-I antigen presentation pathway. Viral proteins within infected cells are degraded into peptides and transported into the endoplasmic reticulum by the transporter associated with antigen processing (TAP). Some of these peptides can bind to MHC class-I molecules and are transported to the cell surface for immune surveillance by the TCRs of CTLs. CTL recognition can lead to the release of cytotoxic granules, the release of lymphokines and the activation of apoptotic pathways via the FAS/FASL interaction to destroy the infected cell [39].

5-4) Mutational escape from CTLs and sequence polymorphism

Cytotoxic T-lymphocytes (CTL) eliminate HIV-1 infected cells through the recognition of antigenic peptides displayed by HLA class I molecules on the infected cell surface [19]. Multiple mechanisms of immune evasion have been described for HIV-1 and over the natural course of an infection, CTL-mediated selective pressure acts as a major driving force of HIV evolution, resulting in the selection of escape mutations as quasispecies [17, 20, 39-40]. Biefly, to evade CTL-mediated immunesurveillance, HIV possesses the ability to interfere with antigen presentation and recognition by deploying a number of mechanisms, such as mutation in viral genome which ultimately lead to epitope deletion, failure of antigen processing, loss of MHC class-I binding, and impaired recognition by the TCR of CTL [29, 41-45].

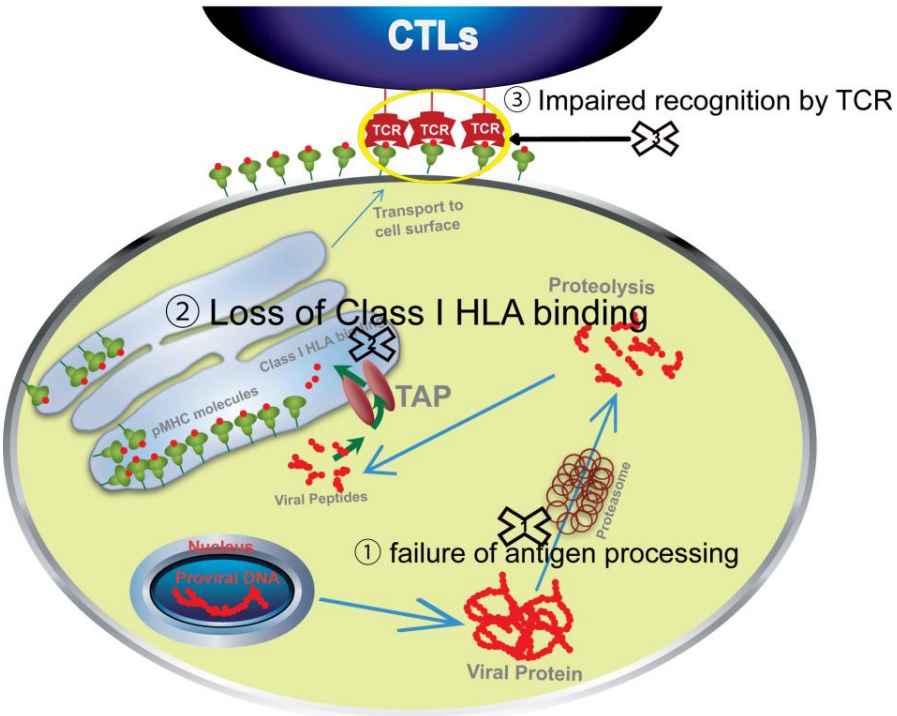

Figure D: Mutational escape mechanism from CTL by HIV-1. 1) Changes in the amino acid residues can lead to altered proteolysis and epitope loss. 2) Mutation in MHC "anchor" residues and reduced binding to MHC class-I molecules which lead to loss of cell surface presentation. 3) Variation in the peptide-MHC class-I surface recognized by the TCR can lead to escape from CTL recognition by a number of mechanisms including TCR antagonism, CTL anergy, distortion of CTL repertoire, and CTL decoy [39].

5-5) HLA class I mediated "footprint/landscape" at population level

Antigenic variation occurs in the context of viral fitness within the host environment. Genetic variation within the highly polymorphic gene of HLA-I alleles allow cytotoxic T lymphocytes (CTLs) to recognize diverse pathogens [46] . This HLA-I imposing immune pressure, on the other hand contribute toward the virus escape and the resultant escape variants could evade HIV specific CTL-mediated immune surveillance, by changing critical amino acids of HLA binding epitopes or flanking regions of the defined epitopes [47-53]. Therefore, immune escape pathway may also represent a significant force shaping viral evolution at the population level. This HLA mediated phenomenon called "imprinting effect" in which escape mutations selected in the context of common HLA class I alleles may become dominat in the circulating viral population if they do not revert upon transmission to new hosts [17, 54].

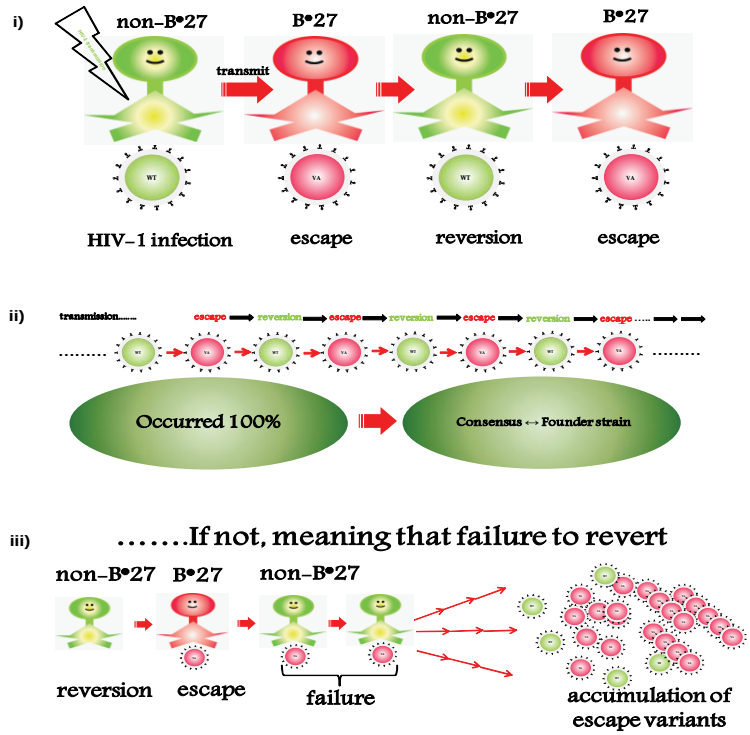

Figure E: HLA mediated polymorphism in the context of escape and reversion at the population level.

However, the reason why we can identify HLA-mediated polymorphism in cross-sectional datasets is due to continued escape and reversion as HIV passes through different human hosts

having different HLA alleles (Figure E [i]). If, the transmission and reversion occurred 100% of the time, HIV polymorphism frequencies would remain stable over time. In addition, the average (consensus) circulating HIV sequence would closely resemble the original founder strain of the epidemic (Figure E [ii]), but, if there is failure to revert, it leads to accumulation of escape variant at the population level (Figure E [iii]).

As such, the frequency and distribution pattern of HLA alleles at the population level may limit or constrain the mutational ability of HIV-1 in response to selective immune pressure and thereby HLA profile leaves "footprints or landscapes" in vivo throughout the entire genome of HIV. Such viral adaptative scenario was first reported at the population level by Moore et al. in 2002 and then followed by the other groups, more explicitly, by correcting few major confounding factors [22, 54-56]. Moreover the identification of HLA-I mediated unique sequence polymorphism have already been reported by well studied multicenter cohorts from Canada, Australia, and USA designated as "International HIV Adaptation Collaborative **[**IHAC]" (e.g., in Gag, Pol, and Nef) and Mexican cohort (e.g., in Pol), suggesting the differential patterns of imposed selective pressure could be existing among the geographical region even within the HIV-1 clade B [21, 57]. It therefore quite prompting, further study to focus on different ethnic population to see the HLA-I mediated polymorphism in HIV-1 genome.

5-6) HIV-1 and VPU

Viral protein U (Vpu) is an accessory protein that is unique to HIV-1 and a subset of related simian immunodeficiency viruses. Vpu is an oligomeric, 81-amino acid type I membrane protein (16 kDa) that is translated from vpu-env bicistronic mRNA by Rev dependent manner [13, 58]. Residues 1-27 constitute the N-terminal hydrophobic membrane anchor, followed by 54 residues that protrude into the cytoplasm. A highly conserved region spanning residues 47–58 contains a pair of serine residues that are constitutively phosphorylated by casein kinase II [59]. The Vpu cytoplasmic domain contains a high proportion of charged residues, which include a membrane-proximal stretch of basic residues followed by a series of acidic residues in the C-terminal part of the protein that confer an overall negative electrostatic charge to the molecule [60].

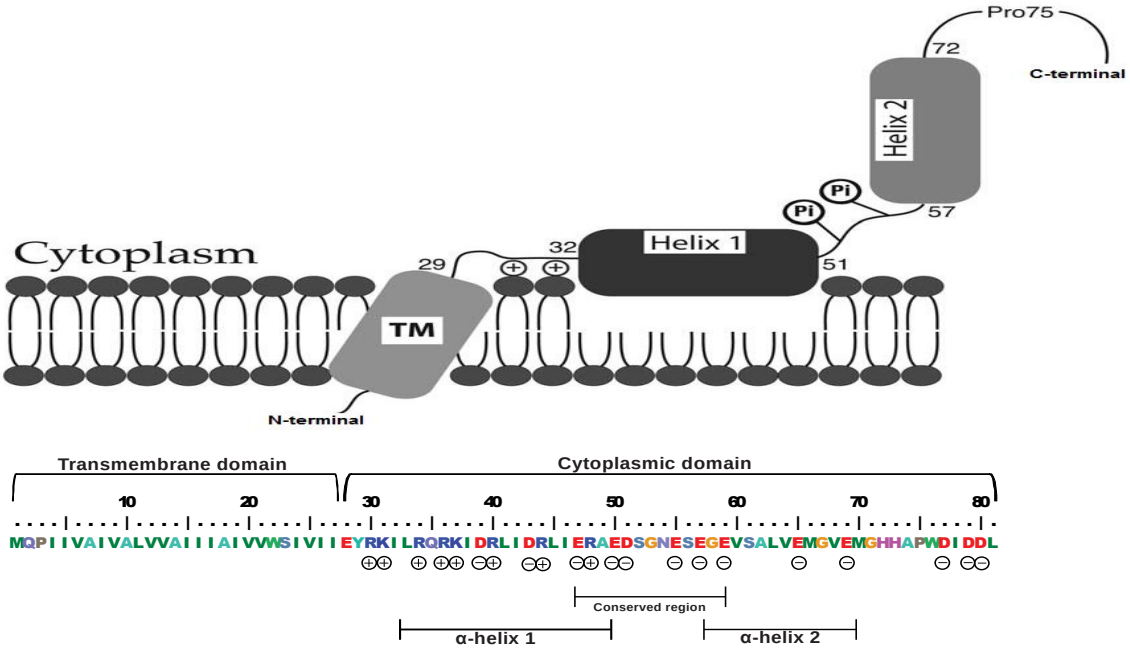

Figure F: Adapted Vpu secondary structure as deduced from available NMR and modeling data (Bour& Strebel et al., 2003)

The influence of Vpu is cell-type specific; some cells require Vpu for virus release, while others do not. It was shown that the Vpu-dependency phenotype is dominant, indicating that such cells express Vpu-sensitive cellular factors that prevent HIV-1 release. The interferon-induced restriction factor that prevents retrovirus release from the plasma membrane which was recently

identified as a cellular protein of previously unknown function called B cell stromal factor 2 (BST-2) or CD317, has also been called ‘‘tetherin’’ to reflect its antiviral activity [12, 14].

The HIV-1 Vpu protein has 2 major functions: degradation of newly synthesized CD4 molecules in the endoplasmic reticulum and enhancement of the release of progeny virions from infected cells, by antagonizing tetherin/CD317, a host restriction factor that directly binds and retains viral particles on the surface of infected cells [13, 15].

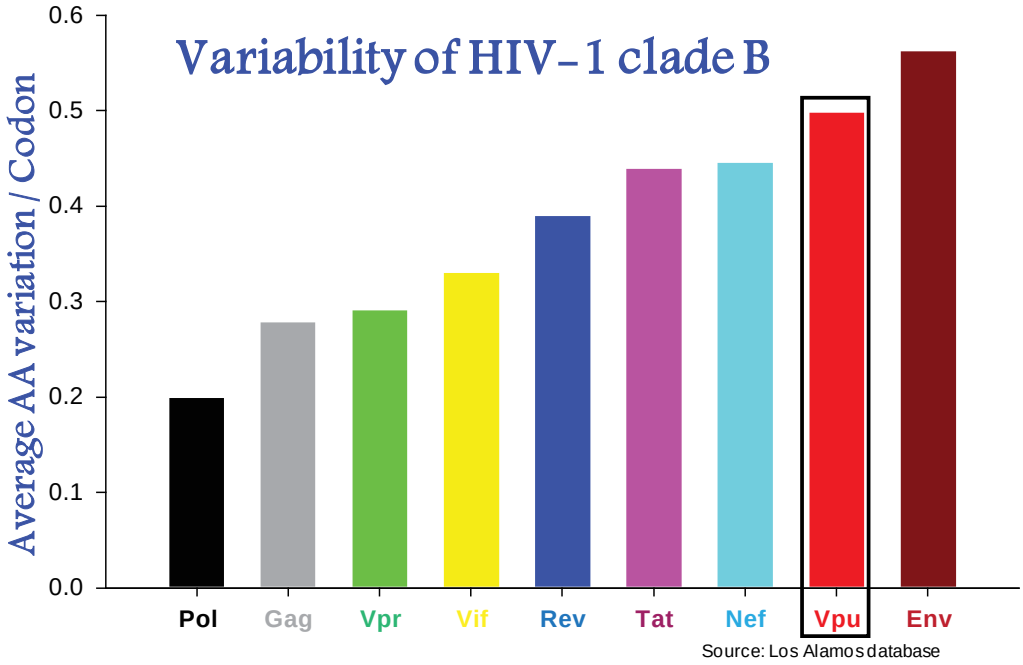

Figure G: Protein variability of HIV-1 clade B. A Shannon entropy score for each of the protein of HIV-1 were calculated to see sequence variability and all of the sequences were retrieved from Los Alamos database.

As such, Vpu is thought to play a role in virus spread and pathogenesis in vivo. Strong lines of evidence point to a role of Vpu in HIV pathogenesis in pig-tailed macaque using SIV/HIV chimeric viruses (SHIV) have shown that mutation of the vpu initiation codon rapidly reverts to an intact ORF of Vpu. This functional Vpu, which correlates with a phase of profound loss of CD4+ cells with a vast increase in the plasma viral RNA levels, after post infection [61-62]. In addition,

comparison of proviral accessory genes between long-term nonprogressors (LTNP) and progressors of HIV-1infection highlighted that only the LTNP harbor viruses with mutated Vpu protein and good control over the virulent viruses which ultimately brings to focus, the consequence of the host mediated immune responses [63].

Interestingly, Vpu is one of the most variable regions among all HIV proteins as evidenced by a cross-sectional comparison of HIV-1 sequences isolated from HIV-infected individuals [64]; raising the possibility that Vpu undergoes adaptation in response to host immune responses.

Figure H: **Host mediated immune components and its interaction with Vpu protein**

Taking into account the possible immune constrains on Vpu by host, it is well appreciated that HLA class I-restricted CD8+ cytotoxic T lymphocyte (CTL) response is thought to play an important role in controlling HIV replication in vivo during an HIV infection [29, 65].

5-7) Aims of the Study

Formidable sequence diversity has been exhibited by HIV-1 Vpu protein, despite its crucial role within HIV-1 infected cells [11, 64, 66]. To date, little is known about the variability of *vpu* gene and host mediated selective pressure at the population level in cross sectional studies. However, how and what extent to which host immune responses drive genetic variability of vpu is incompletely characterized from HIV-1 infected patients. Looking at the highly polymorphic phenotype of Vpu in this present study, we have analyzed the implication of HLA class I associated polymorphisms of Vpu from autologous viral sequences to get comprehensive glimpse about genetic variability and its association by HLA class I molecules.

The specific aims were:

a. To determine how and what extent genetic variability of HIV-1 Vpu is driven by HLA class I alleles?
b. To determine how compensatory mutations of Vpu interplay to act as functional proteins?
c. To determine any correlation between Vpu polymorphism and clinical parameters of HIV-1 infected patients (e.g., plasma viral load and CD4 count).

6. MATERIALS AND METHODS

6-1) Study Population

A total of 240 chronically HIV-1-infected, treatment-naïve subjects (CD4, median 237; IQR, 160-397; viral load, median 33,200; IQR, 222,000-55,400) followed at the AIDS Clinical Center, International Medical Center of Japan were enrolled in this study. All participants provided written informed consent. The total sample collection tenure was between 2005 to 2010 and the subjects were selected based on the availability of plasma and typed HLA class I alleles. The patients HLA types were determined by standard sequence-based genotyping. The study was conducted in accordance with the human experimentation guidelines of the International Medical Center of Japan and Kumamoto University.

6-2) Amplification and sequencing of autologous *vif, vpr and vpu* gene

HIV-1 particles were precipitated by ultracentrifugation (50,000 rpm for 15 min) of Patient's plasma, after which the viral RNA was extracted from them using, ready to use kit according to manufacturer's instructions (QIAamp Viral RNA Mini Kit,QIAGEN). The cDNA was obtained from the extracted viral RNA by Cloned AMV First Strand cDNA Synthesis Kit (Invitrogen Corp., Carlsbad, CA) following the instructions provided by the manufacturer. A total of 20 µl of total HIV cDNA was obtained and stocked at -80 ºC till further use. Gene of interest (vif, vpr, and vpu) were amplified using sets of primers (HXB2 backbone numbering) from the cDNA followed by nested PCR and afterward run in 1-2% agarose gel to check the desirable PCR product size (~1300 bp). The primers were used as follows:

1st round of amplification-

VVVa-F (5`- TTAAAAGAAAAGGGGGGATTGGGGG-3`) and

VVVb-R (5`-ATTCCATGTGTACATT GTACTGT-3`)

2nd round of amplification-

VVVc-F (5`-AGATAATAGTGACATAAAAGTAGTGCCAAGAAG-3`) and

VVVd-R (5`-CCATAATAGACTGTGACCCACAA-3`).

1st and 2nd round of PCR follow the same programme in Bio-Rad thermal cycler

Step	Tem^{0}C	Min	
Initialization	98^{0}C	0:10	
Denaturation	98^{0}C	0:10	} x 30
Annealing	60^{0}C	0:20	
Extension	68^{0}C	1:00	
Final elongation	68^{0}C	7:00	
Hold	15^{0}C	for ever	

The products were stocked at -20^{0}C till further use.

After successful amplification, the PCR amplicons of vvv genes were subjected to gel purification using the QIAEX II Agarose Gel Extraction kit (Qiagen, Valencia, CA), according to the manufacturer's instructions. The purified products were then directly bulk sequenced in 5' and 3'directions using ABI BigDye Terminator (v3.1) cycle sequencing reaction kit (Applied Biosystems) and analysed on a high throughput Applied Biosystems 3500, 3500xL automated genetic analyzer.

6-3) Retrieve gene of Interest and sequences analysis

All raw bulk sequence datasets were uploaded in SeqScape software (V2.7, Applied Biosystems, Foster City, CA) in a project basis to get the gene of interest (e.g., *vpu* gene). In brief, sequences were assembled, aligned to the HXB2 sequence (accession no. K03455), and manually inspected to check the integrity of entire length for each gene and then translated to its protein product. If the second peak of sequence spectra at the same position exceeds 25% of the dominant peak, it was assigned as a mixture of bases. Concatenated sequences of *vif*, *vpr*, and *vpu* reading frames were used for subtyping in the Los Alamos HIV Database (http://www.hiv- .lanl.gov/) followed by HIV genotyping tools at the NCBI (http://www.ncbi.nih.gov/projects-/genotyping/formpage.cgi). More than 90% of the subjects were infected with subtype B in this studied cohort.

6-4) Analysis of amino acid sequence variability

Shannon entropy was used to calculate the amino acid variation for each codon of Vpu protein. This tool is potentially useful to simply assess the diversity in the population level as a cross-sectional study. In this application, each column in a sequence alignment was considered independently and Shannon entropy used as a way to assign a score to each column that reflects the variability in that column. However, this entropy approach does not take into consideration the phylogenetic history of the sequences, nor do for patterns of co-variation [64]

6-5) HLA-I associated sequence polymorphism and Codon-covariation

Phylogenetic Dependency Networks (PDN), a multivariate statistical algorithm were employed to evaluate the associations between host HLA class I alleles and viral amino acid variation in this studied population. In brief, this approach fit to account simultaneously HIV founder effects, linkage disequilibrium among host HLA alleles, and HIV codon co-variation in

order to identify the acting primary sources of selection pressure in each codon of protein. A threshold false discovery rate of 20% was used in this present study ($q<0.2$) for statistical significance among the multiple comparisons. For each target amino acid, the PDN algorithm scanned through all possible amino acid from other positions and HLA profile dataset of studied population to identify the predictor HLA/AA, which would have significant association in the context of enrichment, in the phylogenetic alignment of sequences [18]. Therefore, each pair of "target and predictor" associations was summarized into two categories: "Nonadapted" amino acids which are enriched in the absence of the restricting HLA/AA of interest and vice versa. Usually, "nonadapted" forms represent the consensus amino acid at that position, and they can be thought of as the "wild-type" or "susceptible" form. Conversely, "adapted" amino acids are those enriched in the presence of the HLA/AA and vice versa; these can be thought of as the escape variant. In order to differentiate "direct" (covariation-corrected, meaning polymorphism by particular HLA allele) associations from "indirect" (covariation-uncorrected, meaning that covarying amino acid residue rather than HLA allele) associations we ran the analysis with and without the HIV codon covariation, and included all HLA-associated polymorphisms identified by either method in the immune escape maps [17-18, 21, 67].

6-6) Immune Escape Map

All of the polymorphisms associated by HLA alleles within Vpu protein at $q<0.2$, were organized to specify the location, HLA restriction, specific amino acids substitution, and direction of association (adapted vs non-adapted) in easy to visualize form entitled as "immune escape map" with respect to HIV-1 subtype B consensus as index sequence from Los Alamos database [21]. Afterwards the escape map was adopted with the published optimally described CTL epitopes [68].

6-7) Codon-Covariation and Heat map

The intraprotein amino acid codon covariations and HLA-codon association of Vpu were addressed through easy to visualize tool, PhyloDv (http://www.codeplex.com/MSCompBio) a dependency-network of intra-protein viewer which draws the protein as a circle with N-terminus at "3 o'clock" position and protein extending counter-clockwise around the circle. Thereby, any HLA alleles associated with variation at those sites are indicated at the corresponding positions outside the circle, while covarying amino acids are joined together by arcs within the circle and the strength of the association (q-value) is indicated by the color of the arc [18].

6-8) Clinical parameter and Codon-Codon analysis

A non-parametric statistical hypothesis was run to assessing whether one of two samples of independent observations tends to have larger values than the others in the context of sequence polymorphisms and clinical parameters (e.g., plasma viral load and CD4 count). In brief, subjects having amino acid variation in each position of Vpu were compared with the rest of the subjects not having the certain amino acid of interest, to see any significant association of amino acid for this studied population.

7. RESULTS

7-1) Genetic variability of the HIV-1 *vpu* gene

In order to determine the genetic variability of vpu, we enrolled 240 plasma samples from chronically HIV-1 infected patients in Japan. The amplified DNA of 216 samples (90%) encompassing the *vpu* region nucleotides were sequenced and translated to its respective proteins to get the amino acid variation at each codon by applying the Shannon entropy score in this studied population (see material and method).

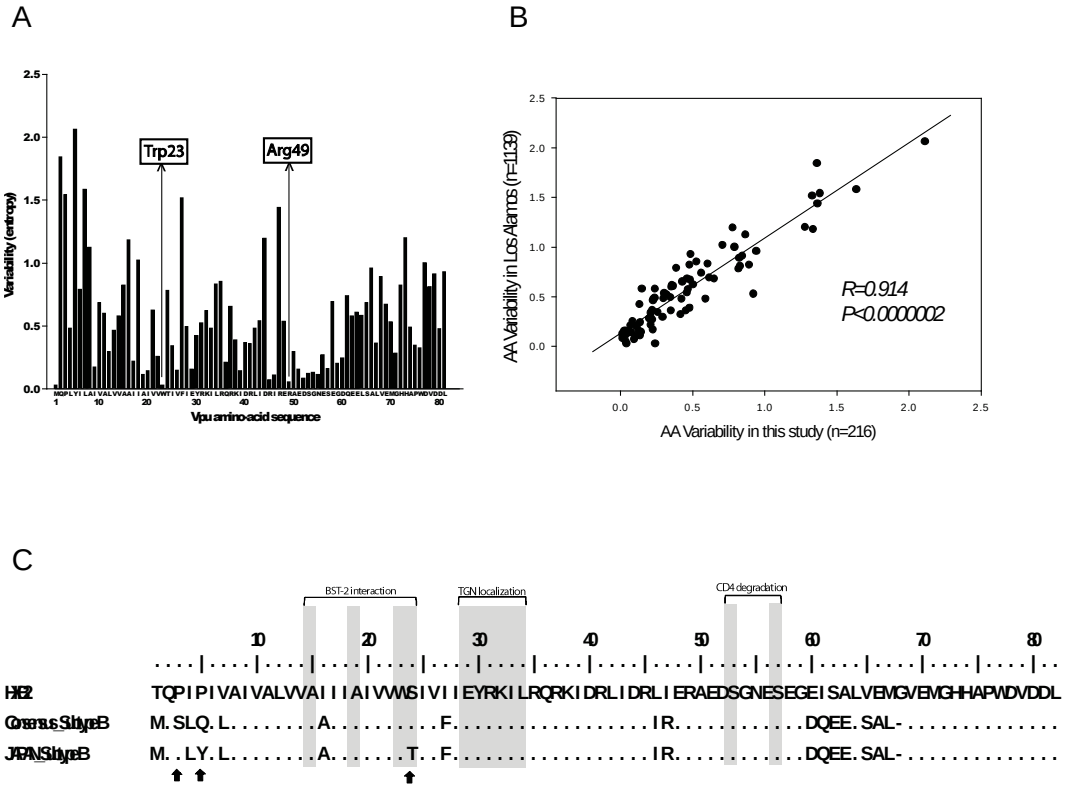

Figure. 1. Variability of the amino acid residues of HIV-1 Vpu. The amino acid sequence of Vpu was analyzed based on the cross-sectional studies on 216 HIV-infected subjects. The amino acid variability at each position of Vpu was analyzed by determining its Shannon entropy score (panel A). The Vpu sequence (subtype B, n = 1139) was retrieved from the Los Alamos HIV sequence database, analyzed for its amino acid variability, and compared with subtype B obtained from this study using Spearman Rank Order Correlation (panel B). The consensus sequences of Vpu obtained from Los Alamos database and this study were aligned with reference strain HXB2 and regions responsible for some key Vpu functions highlighted (panel C).

Out of 81 codon positions only two amino acid residues, Trp23 and Arg49, seem highly conserved (>98%) among individual sequences. Instead, most codons displayed substantial

variability, with the average of the entropy score reaching 0.58 (Fig. 1A), confirming the findings by Yusim et al., which showed that Vpu is a highly variable protein [64] . To see whether or not this variability of Vpu at each codon had a salient difference with respect to global database, we retrieved clade B Vpu sequences from the Los Alamos database. Using Spearman Rank Order Correlation analysis, we observed that each codon in the present study correlated strongly with that of published subtype B sequence dataset, suggesting that our observed pattern of amino acid variation in Vpu was generally representative of the variation observed in HIV-1 subtype B. In fact, the consensus amino acid sequences of subtype B and the present dataset were identical except for 3 amino-acid residues: positions 3, 5, and 24. These amino acid residues were highly variable and not directly associated with known Vpu functions so far reported in literature. This preliminary analysis highlights that, Vpu is being imposed by differential host mediated immune pressures that may be the causes of the observed polymorphisms in HIV-1 at the population level.

7-2) HLA-associated polymorphisms in Vpu

Since, HLA-I-mediated selective pressure contributes to HIV-1 sequence variability over the natural course of infection, especially in the accessory protein Nef [17], we sought to examine whether HLA-I-mediated selective pressure substantially influenced the evolution of Vpu, another accessory protein of HIV-1. We applied a phylogenetic dependency network model [13], which adjusts for the confounding effects of HIV phylogeny, HIV codon co-variation and linkage disequilibrium of HLA-I alleles.

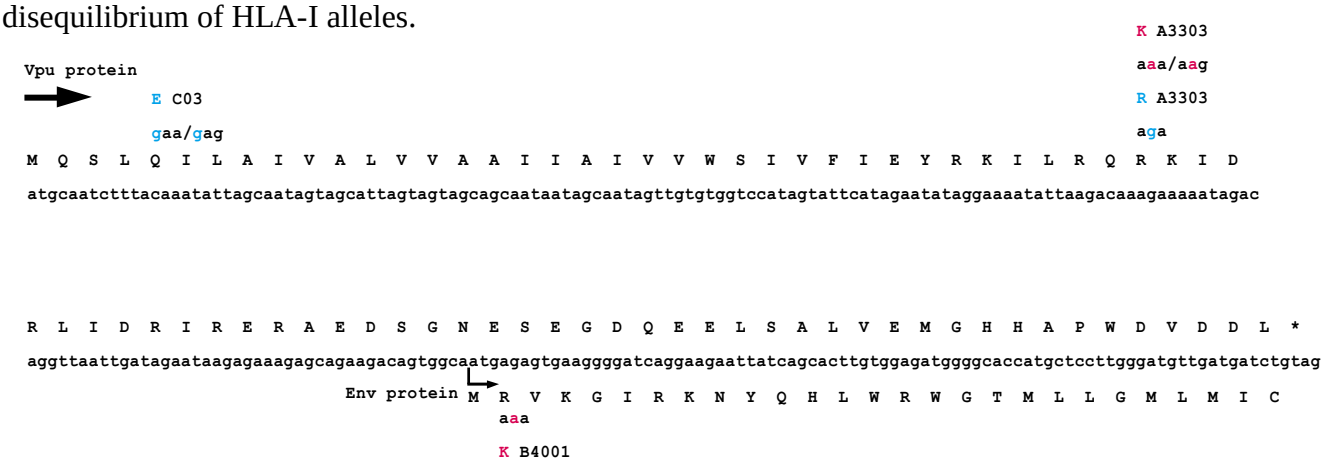

Figure. 2. The Vpu and a part of Env proteins and their associations with host HLA class I alleles. The nucleotide sequence and its deduced amino acid sequence of Vpu and of an overlapping part of Env with reference to the subtype B consensus sequence of Los Alamos database are shown. The amino acid residues associated with the indicated HLA class I alleles ($p < 0.05$, $q < 0.2$) are shown with adapted (red) and nonadapted (blue) residues.

In our dataset of 216 individuals, we identified only three HIV-HLA associations in Vpu: a nonadapted association between C*03 and Glu-5, a nonadapted association between A*33:03 and Arg-37, and an adapted association between A*33:03 and Lys-37 (for adapted vs nonadapted, see material and method). The presence of both nonadapted and adapted A*33:03-associated polymorphisms at Vpu codon 37 is consistent with an Arginine-to-Lysine escape mutation occurring at the C-terminus of the immunodominant HLA-A*33:03-restricted epitope in Vpu, ^{29}EYRKILRQ**R**^{37} [69]. However, there was no HLA-restricted T cell epitopes around Vpu position 5, has been reported. These data suggest that HLA-I-mediated selective pressure toward Vpu does not substantially drive Vpu variability at the population level in this cohort.

7-3) HLA-associated polymorphisms in alternating reading frames (ARFs)

The pathogens which show frequent nucleotide deletions or insertions in their genome may influence the occurrences of ARFs in vivo especially in HIV and other retroviral pathogen [70-71]. CTLs can recognize epitopes encoded by alternate reading frames including the antisense-strand sequences of HIV-1 *gag*, *pol*, and *nef* [72-73]. Therefore, we also investigated HIV-HLA polymorphism associations in peptide sequences encoded by alternative reading frames of the *vpu* gene. We observed no statistically significant HLA-associated polymorphisms in alternate or antisense reading frames, except for a single HLA-B*40:01 associated "adapted" lysine polymorphism at codon 2 of the overlapping Envelope reading frame which is initiated in the middle of the *vpu* gene (ORF +2; Table 1, Fig. 2).

Table1. Summary of HIV-HLA association in the Vpu-encoded region

RF	Protein	Pos HXB2	aa	HLA	Association	p-value	q-value	known epitope	
								sequence	reference
+1	Vpu	5	E	C*03	Nonadapted	2.13×10^{-5}	1.52×10^{-1}	none	
		37	R	A*33:03	Nonadapted	3.40×10^{-6}	5.50×10^{-2}	^{29}**EYRKILRQR**^{37}	Addo et al. 2002
		37	K	A*33:03	Adapted	2.80×10^{-5}	1.52×10^{-1}	^{29}**EYRKILRQR**^{37}	Addo et al. 2002
+2	Env	2	K	B*40:01	Adapted	1.63×10^{-5}	1.67×10^{-1}	none	

RF, reading frame; Pos HXB2, amino acid position when aligned to HXB2 sequence

Although this association was between Lys-2 of Env and *HLA-B*40:01*, no CTL epitopes have been reported in the context of HLA-B*40:01 in this region. Using bioinformatic prediction programs Epipred [74] and BIMAS [75] we attempted to predict B*40:01-restricted CTL epitopes, but found none (data not shown). This failure is most likely due to the presence of several basic amino acids, such as Arg and Lys, in this region of Env, as it has been shown that HLA-B*40:01 preferentially binds peptides with acidic residues at their anchors [76]. This issue needs to be clarified in further studies using immunological assays. Taken together, our results suggest once again that HLA-I-mediated selective pressure do not contribute to a large extent, to population level sequence variation of Vpu in this cohort.

7-4) **Codon covariation path and compensatory dynamics of Vpu**

Given that Vpu is functionally important in viral replication *in vivo*, the highly variable nature of Vpu amino acid sequences could be explained by complex networks of codon-codon covariation and/or secondary/compensatory mutation pathways. We therefore examined the codon-codon covariation of Vpu by using the phylogenetic dependency network model. Although Vpu consists of only 81 amino acids, we identified 103 covarying codon pairs in Vpu, displayed in Fig. 3. The covariation network in Vpu showed an uneven distribution, with a large number of codon-codon covariation networks at the N-terminal membrane-spanning region, a region responsible for BST-2 interaction [77].

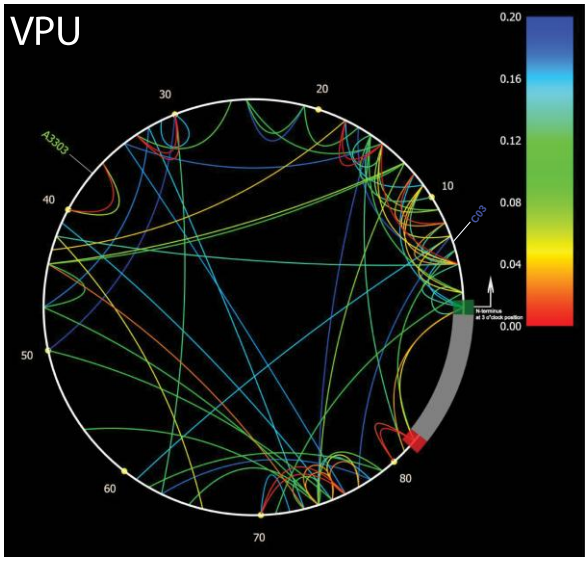

Figure. 3. Amino acid codon–codon covariation in Vpu. The circular map, generated by the PhyloDv software [18], shows Vpu codon–codon covariation associations as inner arcs connecting the association sites, with the HLA associations as tags pointing to their corresponding sites. Q values of individual codon pairs are represented as a heat map shown at the right.

Interestingly, the 3 HIV-HLA associations (Table 1, Fig. 2) were not significantly linked to any other amino acid residues. These data suggest that the conformation and function of Vpu may be preserved through many possible combinations of primary and secondary polymorphisms and that the HLA-I-associated immune-mediated selective pressure may have only a minor influence on such Vpu polymorphisms.

7-5) Association between Vpu polymorphisms and clinical parameters

Finally, we explored associations between Vpu polymorphisms and clinical parameters of HIV-infected patients (*i.e.*, CD4 counts and plasma viral load). To assess the relationship, a non-parametric statistical hypothesis was used. We observed no significant associations between Vpu polymorphisms and CD4 counts. However we identified a statistically significant association between amino acid residues at position 5 and viral load.

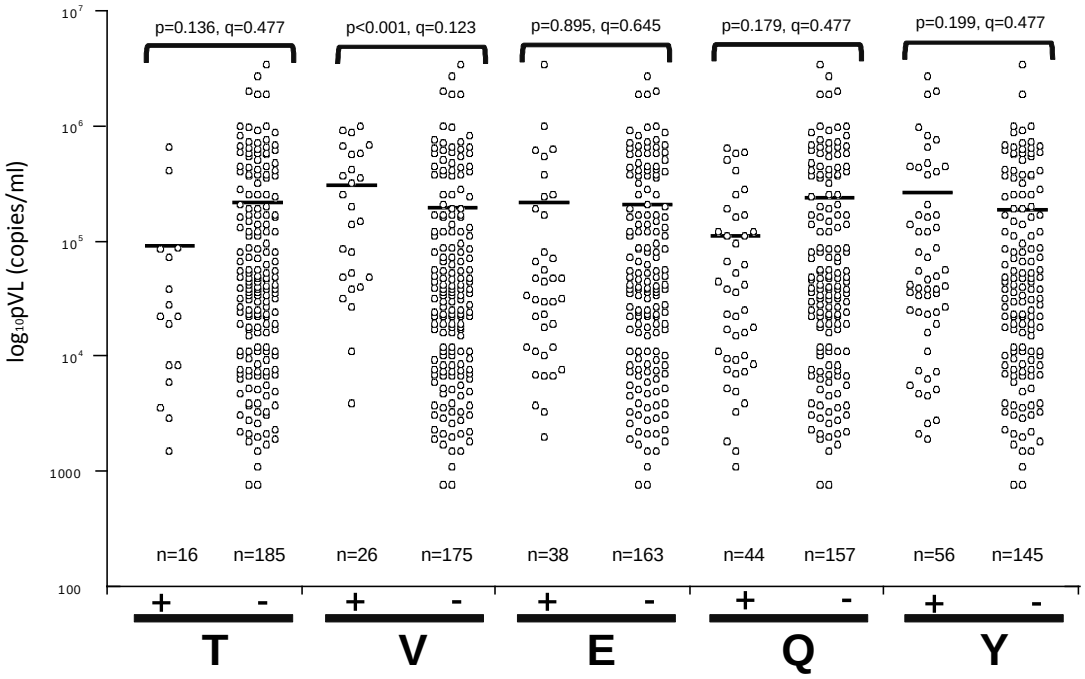

Figure. 4. Association between plasma viral load and amino acid polymorphism at position 5 of Vpu. HIV plasma viral loads, stratified by amino acid expression at Vpu codon 5, are shown. Vpu codon 5 exhibited 11 different amino acids positioning in our dataset; only those observed in >10 patients are shown here. Horizontal bars indicate medians. Statistical analysis was performed using the Mann-Whitney U-test.

The patients harboring Val at Vpu-5 had significantly higher viral loads compared to those with amino acid residues other than Val at this position. Considering that the amino acid residue at this position is located in close proximity to the membrane-spanning region and that this region is functionally important for BST-2 binding, it would be interesting to examine the functional effects of amino acid polymorphisms at position 5, whether they are mediated by host immune responses or otherwise.

8. DISCUSSION

Recent advances in HIV-1 research and compelling evidences highlight that HLA class I-restricted CD8+ cytotoxic T lymphocyte (CTL) response constitute an essential component of protective antiretroviral immunity [22, 27, 65, 67]. Human HLA genotypes largely affect disease progression in HIV infection because presentation of antigenic peptide is restricted by HLA-I alleles to the circulating CTLs. Early control of HIV-1 infection is determined by a balance between the host immune response and the ability of the virus to escape from the imposing response. To be accomplished as a potential intruder in host immune system, HIV has evolved extraordinary mutational capacity which could be evident in the form of HIV-1 resistance to antiviral agents, neutralizing antibodies and CTLs, involving multiple viral gene products [78-80]. It has been suggested that a comprehensive analysis of the polymorphisms in HIV proteins is of value for understanding the virus transmission and pathogenesis as well as for the efforts towards developing anti-viral therapeutics and vaccines [19, 22, 78].

Viral protein U (Vpu) is one of the characteristic genes of primate lentivirus and found exclusively in HIV-1 with the notable exception of SIVcpz and SIVgsn, while the other accessory genes vif, vpr, and nef are expressed in most HIV-1, HIV-2 and SIV strains [60]. This smallest protein (81aa), is well known to play an important role in viral replication through de-novo CD4 degradation and enhancement of virion release by antagonizing tetherin/CD317 from HIV-infected cells [12]. Remarkably, Vpu represents one of the highly variable proteins in the HIV-1 proteome [64], but it is still not clear to what extent HLA-I allele contributes to its evolution. Although, it has been shown that Vpu is a minor target for CTLs as revealed by IFN-γ Elispot assays with overlapping peptides based on the subtype B consensus sequence. Considering the highly variable nature of Vpu, it is possible to miss responses if the autologous virus sequence is markedly different from the peptide sequence when using this Elispot assay system [81].

In the present study, we sought to identify HLA-associated polymorphisms in Vpu primary and alternate reading frames (ARFs) and examine to what extent they are involved in amino acid variability at the population level. We utilized a published phylogenetic dependency network model, a comprehensive evolutionary model that considers all important confounding effects such as HIV phylogeny, HIV codon covariation, and linkage disequilibrium of HLA alleles [18]. We enrolled 240 HLA-I-typed, treatment naïve, chronically HIV-infected subjects in Japan, and analyzed plasma HIV RNA nucleotide sequences of the Vpu region to identify viral sites under contemporaneous immune pressure by HLA class-I alleles.

In the preliminary genetic analysis of Vpu, only two codon positions Trp23 and Arg49 have shown high conservative nature (~98%) of all 216 individual patients sequence alignment, in good agreement with Cordes et al., in their computer simulation study that Trp23 need to be conserved to act as a gatekeeper of controlling ion channel activity of Vpu for viral particle release [82]. In addition, Trp23 also takes part in counteracting BST-2, in conjunction with other amino acids of Vpu (Fig. 1C) [77]. On the other hand Arg49 is located between the two α-helix domains which may assist in proper orientation of Vpu for the two serine residues at position 53 and 57, known to play a pivotal role in CD4 degradation [60]. However, the observed variability pattern of this studied cohort was pointing to the polymorphic nature of Vpu which is a representative model of global subtype B, although noticeable differences were also found at position 3, 5 and 24 in this cohort (Fig. 1C, upward arrow). It therefore warrants for further investigation, with larges number of sequence dataset not only in clade B but also in other clades as well.

To see the sequence polymorphism and HLA enrichment in Vpu using the contemporary bio-statistical approach [18], we identified only 3 HLA-HIV associations from the primary reading frame of Vpu, suggesting less association by HLA molecules. Indeed, high polymorphic nature and low HLA class I allele association from primary open reading frame guided us to look at the alternative reading frames of Vpu. Recent studies have identified the existence of HLA associations

in ARFs of Gag, Pol and Nef, highlighting the presence of HLA class I-restricted CTL responses, targeting epitopes in HIV frame shift sequences [71] and such ARFs encoded-antigenic epitopes, have been reported in a variety of human diseases, including influenza infection, malignancies, and autoimmunity [83-84]. Once again, only 1 HLA-B*40:01 association was identified in RF2 of *vpu* gene, which is actually the overlapping (27aa) region with Env and the adapted amino acid was lysine (K) (Fig.2). However, the B*40:01 associated lysine could be the "landscape/footprint" of this particular overlapping (*vpu-env*) region of HIV-1 proteome as the B*40 allele HIV-infected subjects are relatively high (~17%) in this studied population. It is still enigmatic, from our preliminary analysis how HIV modulate immune pressure from the conserved overlapping genes between *vpu* and *env* by B*40:01 allele, while acting as a functional protein in vivo. This issue needs to be clarified by further studies using immunological assays not only by B*40:01 but also other allelic variants (e.g., B*40:02, B*40:06) as well.

Moreover, we also found that lysine (K) at position 37 is positively enriched (Fig.2 and Table.1) in patients having A*33:03 alleles from the ^{29}EYRKILRQ**R**^{37} epitopic region of Vpu which partially cover two domains (TGN & α-helix-1). Also, this is one of the optimally defined immunogenic part (A*33:03 restriction), identified from long-term non-progressors (LTNP) [81]. This result highlighted that, region 29-37 of Vpu have been undergoing immune constrain by A*33:03 allele which could be seen as a mutation from arginine (R)- to-lysine (K) as adaptive, despite subtle effect of HLA class-I alleles on Vpu.

However, we can't rule out the possible causes behind the low level of HLA association, exemplified by number of patients sample, homogenous population, HLA class-I distribution pattern and its frequency, which remains to be determined by future studies. This observation is consistent with the previously well studied British Columbia, HOMER cohort and combined IHAC cohort from USA, Australia and Canada; which narrated that the HLA associated polymorphisms

were highly associated in Gag, Pol, and Nef rather than other accessory proteins (e.g., Vpr) [17, 21].

In contrast, HIV codon co-evolution domain was also used to see the complex interaction among amino acid within Vpu proteins as substantial sequence diversity was observed in Vpu proteins (Fig.1, entropy score 0.58). To be adapted in host immune system generally, viruses introduce a mutation in one position which in turn may be compensated by creating another secondary mutation for viral fitness cost of primary mutations [18, 21]. Strikingly, despite its relatively small size (81 aa), Vpu harbored 103 codon-codon association (Fig. 4) with uneven distribution. More than 50% of its association was found in the trans-membrane domain (TM), which is known to function as antagonizing region of Vpu toward BST2, which is one of the host restriction factors in viral replication [14] .

Taken together, these data suggest that the conformation and function of Vpu may be preserved through many possible combinations of primary and secondary polymorphisms and the HLA-I-associated immune pressure may have a minor effect on such Vpu polymorphisms in this cohort. Indeed, to get a deep insight of Vpu variability, it is inevitable to look at other differential influencing factors, for example; NK-cell mediated (KIR-associated) immune pressure, evolutionary pressure from SIVcpz to HIV-1, compensatory effect of certain selective pressure as Env translated from vpu-env bicistronic mRNA [58], or variability acts as a key factor to manipulate the viral pathogenesis [15].

A previous study described that sequence polymorphism of Gag protein were associated positively and negatively with higher viral load and low CD4 cells respectively from chronically untreated HIV-infected patients [67]. Such findings directed us to investigate any possible correlation of Vpu polymorphism and the clinical parameters (e.g., plasma viral load and CD4 count) of this HIV-infected population. There was virtually no significant association between Vpu

polymorphisms and CD4 counts. In contrast, we identified a statistically significant association between amino acid residues at position 5 and viral load (Fig. 4). The viral load values of the patients having Val at position 5 showed significantly higher viral load than those having amino acid residues other than Val. In order to further confirm this result, we recruited additional 67 patients and analyzed the Vpu sequence and their viral load. Again, this association between viral load and Val at Vpu-5 still remained statistically significant (data not shown). Indeed, the amino acid residues at position 5 of Vpu showed several interesting features; the highest variability all over Vpu amino acids (Fig. 1A), nonadapted association of Glu-5 with *HLA-Cw*03* (Fig. 2), and association of Val-5 with the increased viral load ((Fig. 4). Considering that the amino acid residue at this position is located in close proximity to the membrane-spanning and functionally important region for BST-2 binding, it would be interesting to see functional effects of amino acid polymorphisms at position 5, whether or not it is immune mediated.

9. CONCLUSION

Since its discovery three decades ago, control of the HIV-1 virus has remained elusive. Various strategies to control the virus, through Highly Active Anti-retroviral Therapy (HAART) and antibody based vaccines have been proven to be futile in the total control of HIV-1. This has prompted further research and studies into other strategies and also refining the current available strategies to control the virus. CTL based vaccines have been said to be a potent strategy to control HIV-1 infection, leading current researches into the development of CTL based HIV-1 vaccines design.

The global AIDS pandemic, almost caused by the HIV-1 M viruses and it evolved a fully functional Vpu, unlike the O or N group of HIV viruses. So far, the accessory genes except Nef, have been underestimated compared to other structural, enzymatic and regulatory part of HIV-1 proteome in vaccine design candidates [16, 66, 81, 85]. Ultimately, Vpu polymorphisms and the integration of population-level analyses will help for comprehensive understanding of the sites, immune dynamics and escape at the population level. As such, Vpu would be a suitable candidate to foster in the current HIV-vaccine therapeutic arena.

In summary, we are suggesting from this study that despite the minor contributions of HLA-I alleles, still, few positions (e.g., 5, 37 and overlapping region with Env) may be under immune constrain by host HLA-I alleles. In the development of CTL based HIV vaccines, the candidate regions should be highly immunogenic, functionally active and conserved, as such regions impose high fitness cost to the virus. This knowledge, in turn, will help us overcome a bit, the complex challenges of HIV-1 diversity and evolution in vaccine design.

10. REFERENCES

1. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, *et al.* Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 1999,**397**:436-441.
2. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983,**220**:868-871.
3. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, *et al.* Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 1983,**220**:865-867.
4. van der Loeff MF, Awasana AA, Sarge-Njie R, van der Sande M, Jaye A, Sabally S, *et al.* Sixteen years of HIV surveillance in a West African research clinic reveals divergent epidemic trends of HIV-1 and HIV-2. *Int J Epidemiol* 2006,**35**:1322-1328.
5. WHO/UNAIDS. Global summary of the AIDS epidemic. *http://www.who.int/hiv/data/en/* 2010.
6. Strebel K, Klimkait T, Martin MA. A novel gene of HIV-1, vpu, and its 16-kilodalton product. *Science* 1988,**241**:1221-1223.
7. Fan L, Peden K. Cell-free transmission of Vif mutants of HIV-1. *Virology* 1992,**190**:19-29.
8. Miller RH, Sarver N. HIV accessory proteins: emerging therapeutic targets. *Mol Med* 1995,**1**:479-485.
9. Bour S, Strebel K. HIV accessory proteins: multifunctional components of a complex system. *Adv Pharmacol* 2000,**48**:75-120.
10. Barraud P, Paillart JC, Marquet R, Tisne C. Advances in the structural understanding of Vif proteins. *Curr HIV Res* 2008,**6**:91-99.

11. Bell CM, Connell BJ, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA. Molecular characterization of the HIV type 1 subtype C accessory genes vif, vpr, and vpu. *AIDS Res Hum Retroviruses* 2007,**23**:322-330.

12. Malim MH, Emerman M. HIV-1 accessory proteins ensuring viral survival in a hostile environment. *Cell Host Microbe* 2008,**3**:388-398.

13. Nomaguchi M, Fujita M, Adachi A. Role of HIV-1 Vpu protein for virus spread and pathogenesis. *Microbes Infect* 2008,**10**:960-967.

14. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008,**451**:425-430.

15. Dube M, Bego MG, Paquay C, Cohen EA. Modulation of HIV-1-host interaction: role of the Vpu accessory protein. *Retrovirology* 2010,**7**:114.

16. Kirchhoff F. Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? *Nat Rev Microbiol* 2009,**7**:467-476.

17. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, *et al.* Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS Pathog* 2007,**3**:e94.

18. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, *et al.* Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS Comput Biol* 2008,**4**:e1000225.

19. Brumme ZL, Art FYP, Jonathan MC, Walkerd BD. Identifying HLA-Associated Polymorphisms in HIV-1. *HIV Molecular Immunology 2010* 2010:3-16.

20. Brumme ZL, Brumme CJ, Chui C, Mo T, Wynhoven B, Woods CK, *et al.* Effects of human leukocyte antigen class I genetic parameters on clinical outcomes and survival after initiation of highly active antiretroviral therapy. *J Infect Dis* 2007,**195**:1694-1704.

21. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, *et al.* HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS One* 2009,**4**:e6687.

22. Brumme ZL, Walker BD. Tracking the culprit: HIV-1 evolution and immune selection revealed by single-genome amplification. *J Exp Med* 2009,**206**:1215-1218.

23. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 1999,**401**:921-923.

24. Kimball's Biology. *Histocompatibility Molecules* 2011.

25. Mayer. Microbiology and Immunology On-Line Textbook. USC School of Medicine. *Gene* 2006 **"Immunology-Chapter One: Innate (non-specific) Immunity"**

26. Beck G, Habicht GS. Immunity and the invertebrates. *Sci Am* 1996,**275**:60-63, 66.

27. Walker BD, Burton DR. Toward an AIDS vaccine. *Science* 2008,**320**:760-764.

28. Ogg GS, Jin X, Bonhoeffer S, Dunbar PR, Nowak MA, Monard S, *et al.* Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* 1998,**279**:2103-2106.

29. Ueno T, Motozono C, Dohki S, Mwimanzi P, Rauch S, Fackler OT, *et al.* CTL-mediated selective pressure influences dynamic evolution and pathogenic functions of HIV-1 Nef. *J Immunol* 2008,**180**:1107-1116.

30. Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, Rathod A, *et al.* HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphotye recognition. *J Virol* 2009,**83**:2743-2755.

31. Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* 2007,**27**:406-416.

32. Klein MR, van Baalen CA, Holwerda AM, Kerkhof Garde SR, Bende RJ, Keet IP, *et al.* Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. *J Exp Med* 1995,**181**:1365-1372.

33. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol* 1994,**68**:6103-6110.

34. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, *et al.* Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* 1994,**68**:4650-4655.

35. Tomiyama H, Akari H, Adachi A, Takiguchi M. Different effects of Nef-mediated HLA class I down-regulation on human immunodeficiency virus type 1-specific CD8(+) T-cell cytolytic activity and cytokine production. *J Virol* 2002,**76**:7535-7543.

36. Yang OO, Kalams SA, Trocha A, Cao H, Luster A, Johnson RP, *et al.* Suppression of human immunodeficiency virus type 1 replication by CD8+ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms. *J Virol* 1997,**71**:3120-3128.

37. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, *et al.* Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *J Exp Med* 1999,**189**:991-998.

38. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, *et al.* Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. *Science* 1999,**283**:857-860.

39. Sewell AK, Price DA, Oxenius A, Kelleher AD, Phillips RE. Cytotoxic T lymphocyte responses to human immunodeficiency virus: control and escape. *Stem Cells* 2000,**18**:230-244.

40. Messaoudi I, Guevara Patino JA, Dyall R, LeMaoult J, Nikolich-Zugich J. Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* 2002,**298**:1797-1800.

41. Chen W, Norbury CC, Cho Y, Yewdell JW, Bennink JR. Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. *J Exp Med* 2001,**193**:1319-1326.

42. Brander C, Riviere Y. Early and late cytotoxic T lymphocyte responses in HIV infection. *AIDS* 2002,**16 Suppl 4**:S97-103.

43. Altfeld M, Allen TM, Kalife ET, Frahm N, Addo MM, Mothe BR, *et al.* The majority of currently circulating human immunodeficiency virus type 1 clade B viruses fail to prime cytotoxic T-lymphocyte responses against an otherwise immunodominant HLA-A2-restricted epitope: implications for vaccine design. *J Virol* 2005,**79**:5000-5005.

44. Probst HC, Tschannen K, Gallimore A, Martinic M, Basler M, Dumrese T, *et al.* Immunodominance of an antiviral cytotoxic T cell response is shaped by the kinetics of viral protein expression. *J Immunol* 2003,**171**:5415-5422.

45. Ali A, Lubong R, Ng H, Brooks DG, Zack JA, Yang OO. Impacts of epitope expression kinetics and class I downregulation on the antiviral activity of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes. *J Virol* 2004,**78**:561-567.

46. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med* 2000,**343**:702-709.

47. Borrow P LH, Wei X, Horwitz MS, Peffer N, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* 1997,**3**:205–211.

48. Allen TM AM, Yu XG, O'Sullivan KM, Lichterfeld M, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic Tlymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J Virol* 2004,**78**:7069–7078.

49. RA K. Virus escape from CTL recognition. . *J Exp Med* 1994,**180**:779–782.

50. Price DA GP, Klenerman P, Sewell AK, Easterbrook PJ, et al. . Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. . *Proc Natl Acad Sci U S A* 1997,**94**.

51. Phillips RE R-JS, Nixon DF, Gotch FM, Edwards JP, et al. . Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* 1991, **354**:453–459.

52. Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, Prendergast A, *et al.* HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. *J Virol* 2008,**82**:6434-6446.

53. Meier UC, Klenerman P, Griffin P, James W, Koppe B, Larder B, *et al.* Cytotoxic T lymphocyte lysis inhibited by viable HIV mutants. *Science* 1995,**270**:1360-1362.

54. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 2002,**296**:1439-1443.

55. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, Addo M, *et al.* Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 2009,**458**:641-645.

56. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan K M, *et al.* Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J Virol* 2005,**79**:13239-13249.

57. Avila-Rios S, Ormsby CE, Carlson JM, Valenzuela-Ponce H, Blanco-Heredia J, Garrido-Rodriguez D, *et al.* Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Retrovirology* 2009,**6**:72.

58. Schwartz S, Felber BK, Fenyo EM, Pavlakis GN. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J Virol* 1990,**64**:5448-5456.

59. Schubert U, Henklein P, Boldyreff B, Wingender E, Strebel K, Porstmann T. The human immunodeficiency virus type 1 encoded Vpu protein is phosphorylated by casein kinase-2 (CK-2) at positions Ser52 and Ser56 within a predicted alpha-helix-turn-alpha-helix-motif. *J Mol Biol* 1994,**236**:16-25.

60. Bour S, Strebel K. The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. *Microbes Infect* 2003,**5**:1029-1039.

61. Stephens EB, Mukherjee S, Sahni M, Zhuge W, Raghavan R, Singh DK, *et al.* A cell-free stock of simian-human immunodeficiency virus that causes AIDS in pig-tailed macaques has a limited number of amino acid substitutions in both SIVmac and HIV-1 regions of the genome and has offered cytotropism. *Virology* 1997,**231**:313-321.

62. McCormick-Davis C, Zhao LJ, Mukherjee S, Leung K, Sheffer D, Joag SV, *et al.* Chronology of genetic changes in the vpu, env, and Nef genes of chimeric simian-human immunodeficiency virus (strain HXB2) during acquisition of virulence for pig-tailed macaques. *Virology* 1998,**248**:275-283.

63. Yamada T, Iwamoto A. Comparison of proviral accessory genes between long-term nonprogressors and progressors of human immunodeficiency virus type 1 infection. *Arch Virol* 2000,**145**:1021-1027.

64. Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, *et al.* Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J Virol* 2002,**76**:8757-8768.

65. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol* 2004,**4**:630-640.

66. Malim MH, Emerman M. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* 2008,**3**:388-398.

67. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, *et al.* Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS* 2008,**22**:1277-1286.

68. Frahm N LC, Brander C. Identification of HIV-derived, HLA class I restricted CTL epitopes: Insights into TCR repertoire, CTL escape and viral fitness. In: Korber BB, C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI, eds (2006) HIV Molecular Immunology 2006/2007: Los Alamos National Laboratory, Theoretical Biology and Biophysics. 2006:3-28.

69. Addo MM, Altfeld M, Rathod A, Yu M, Yu XG, Goulder PJ, *et al.* HIV-1 Vpu represents a minor target for cytotoxic T lymphocytes in HIV-1-infection. *AIDS* 2002,**16**:1071-1073.

70. Cervantes-Acosta G, Welman M, Freund F, Cohen EA, Lemay G. CD4/CXCR4 co-expression allows productive HIV-1 infection in canine kidney MDCK cells. *Virus Res* 2006,**120**:138-145.

71. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, *et al.* Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *J Exp Med* 2010,**207**:61-75.

72. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, *et al.* CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J Exp Med* 2010,**207**:51-59.

73. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, *et al.* Viral adaptation to immune selection pressure by HLA class I–restricted CTL responses targeting epitopes in HIV frameshift sequences. *J Exp Med* 2010,**207**:61-75.

74. Heckerman D, Kadie C, Listgarten J. Leveraging information across HLA alleles/supertypes improves epitope prediction. *J Comput Biol* 2007,**14**:736-746.

75. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* 1994,**152**:163-175.

76. Falk K, Rotzschke O, Takiguchi M, Gnau V, Stevanovic S, Jung G, *et al.* Peptide motifs of HLA-B58, B60, B61, and B62 molecules. *Immunogenetics* 1995,**41**:165-168.

77. Vigan R, Neil SJ. Determinants of tetherin antagonism in the transmembrane domain of the human immunodeficiency virus type 1 Vpu protein. *J Virol* 2010,**84**:12958-12970.

78. Srinivasan A, Ayyavoo V, Mahalingam S, Kannan A, Boyd A, Datta D, *et al.* A comprehensive analysis of the naturally occurring polymorphisms in HIV-1 Vpr: potential impact on CTL epitopes. *Virol J* 2008,**5**:99.

79. Ode H. [Bioinformatics studies on drug resistance against anti-HIV-1 drugs]. *Uirusu* 2011,**61**:35-47.

80. Wijesundara DK, Jackson RJ, Ramshaw IA, Ranasinghe C. Human immunodeficiency virus-1 vaccine design: where do we go now? *Immunol Cell Biol* 2011,**89**:367-374.

81. Addo MM, Altfeld M, Rathod A, Yu M, Yu XG, Goulder PJ, *et al.* HIV-1 Vpu represents a minor target for cytotoxic T lymphocytes in HIV-1-infection. *AIDS* 2002,**16**:1071-1073.

82. Cordes FS, Kukol A, Forrest LR, Arkin IT, Sansom MS, Fischer WB. The structure of the HIV-1 Vpu ion channel: modelling and simulation studies. *Biochim Biophys Acta* 2001,**1512**:291-298.

83. Bullock TN, Eisenlohr LC. Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *J Exp Med* 1996,**184**:1319-1329.

84. Rimoldi D, Rubio-Godoy V, Dutoit V, Lienard D, Salvi S, Guillaume P, *et al.* Efficient simultaneous presentation of NY-ESO-1/LAGE-1 primary and nonprimary open reading frame-derived CTL epitopes in melanoma. *J Immunol* 2000,**165**:7253-7261.

85. Altfeld M, Addo MM, Eldridge RL, Yu XG, Thomas S, Khatri A, *et al.* Vpr is preferentially targeted by CTL during HIV-1 infection. *J Immunol* 2001,**167**:2743-2752.