Winter 12-15-2015

Human Immunodeficiency Virus Type 1 Envelope Protein Interaction with Host Cell Coreceptor C-X-C Chemokine Receptor Type 4

Jie Zhang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Immunology

Dissertation Examination Committee:

Lee Ratner, Chair
Thomas Baranski
Marco Colonna
Robyn Klein
Makedonka Mitreva
Christina Stallings

Human Immunodeficiency Virus Type 1 Envelope Protein Interaction with Host Cell Coreceptor C-X-C Chemokine Receptor Type 4

by

Jie Zhang

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2015
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my sincere gratitude towards my thesis advisor and mentor, Professor Lee Ratner, for his support and guidance throughout the years. He is not only a great scientist, an excellent mentor, but also a role model for my career and life.

I would like to thank the members of my thesis committee who provided thoughtful suggestions to my research and graduate career. In particular, I would like to thank Robyn Klein for serving on my committee chair, and Makedonka Mitreva for facilitating and helping with data analysis for the next generation sequencing section. I would also like to thank Thomas Baranski, Daved Fremont, Marco Colonna, and Christina Stallings for always being positive mentors and role models.

I would like to thank the members of the Ratner lab, both present: Daniel Rauch, John Harding, George Kyei, Xiaogang Cheng, Mathew Cherian, Hicham Baydoun, Rashmi Ramani, Austin Niu, and past: Nancy Campbell, Anna McCulley, Tenira Townsend, Esa Seegulam, Sirosh Bokhari and Sidney Leibel, for their friendship, patience, intellectual discussion, and guidance over the years. They really make the lab a wonderful and pleasant working environment.

I would like to thank my collaborators, in particular, John Martin and Bruce Rosa in the Mitreva lab, Zheng Chen who now has moved on to Facebook Inc., Junchen Gu in Ting Wang lab, Kristian Wylie, Guoyan Zhao in David Wang lab, Henry Huang, and Justin Fay, for their generous help on next generation sequencing data analysis. I would like to thank our collaborators Daniel Kuritzkes and Timothy Henrich at Harvard for providing clinical samples and valuable comments on the manuscript.

I would like to thank everyone who have helped and supported me over my graduate years in the Immunology Program and in the Molecular Oncology Department, including but not limited to, Kenneth Murphy, Paul Allen, Emil Unanue, Melanie Relich, Cheryl Lyons, Sharon Becker, and Helen Black. I am grateful for their kindness and support to me by going beyond their everyday work.

I have been extremely fortunate to a group of friends at Washington University, who have supported me in one way or another, and given my wonderful memories: Xinxin Wang, Xiaolu Xu, Cheng Li, Wan Shi, Fei Wang, Pei Xu, Yinzi Liu, Lu Yu, Shuo Luo, Cong Li, Fengzhi Tang, Haowei Yuan, Peng Lin, Yanfei Jiang, Yu Zhu, Yiqun Zheng, Gregory Watson, Deepak Nayak, Jamie Rimer, and many more, for stimulating discussions, career suggestions, emotional counselling, feeding me, driving me around, looking out and taking care of me. Without them, my graduate school experience would not be the same.

Most importantly, I thank my parents, Qiong Liu and Shaohua Zhang, my grandparents, Pingping Zhou and Yanling Liu, for their unconditional love, support, and encouragement at every step of the way. I know I can always share with them my happy moments, and turn to them during difficult times.

Words fail me when I try to express my appreciation and gratitude to my husband, Xiang Gao, who has been supporting me both in spirit and physically, over the course of my graduate career. Last but not least, I would like to thank my two feline family members, Happy and Dong Dong, for their powerful healing magic through cuddling and purring.

Abstract of the dissertation

Human Immunodeficiency Virus Type 1 Envelope Protein Interaction with

Host Cell Coreceptor C-X-C Chemokine Receptor Type 4

by

Jie Zhang

Doctor of Philosophy in Biology and Biomedical Sciences

Immunology

Washington University in St. Louis, 2015

Professor Lee Ratner, Chair

Human Immunodeficiency Virus Type 1 (HIV-1) envelope protein is the sole determinant for viral entry and tropism. The ability of HIV-1 to infect susceptible host cells depends on the ability of its envelope protein to engage host cell receptor CD4 and coreceptor C-C Chemokine Receptor Type 5 (CCR5) and/or C-X-C Chemokine Receptor Type 4 (CXCR4). Most naturally occurring infections start with a single CCR5-tropic virion. In approximately 50% of HIV-1 Clade B infected patients, the viral population spontaneous develops the ability to utilize CXCR4 at a late disease stage, and this coreceptor shift corresponds to a poor prognosis for the patients. Clinical application of a CCR5 antagonist drug accelerates this coreceptor shift process. Despite the important implications of HIV-1 tropism on disease pathogenesis, prognosis, and treatment, molecular mechanisms for coreceptor shift and the contributing envelope determinants have not yet been clearly defined due to the dynamic, multimeric, multi-step nature of envelope-coreceptor interaction. In order to better understand the regions on envelope that are important for coreceptor specificity, we studied a panel of HIV-1 envelope

amplicon samples from patients who experienced coreceptor shift and treatment failure in a CCR5 antagonist clinical trial. In order to further understand the structure-function relationship of HIV-1 envelope, we utilized both in-depth single clone analysis to fully characterize the functional attributes of individual envelope isolates, and high-throughput deep sequencing analysis to investigate the global envelope variant landscape before and after treatment.

In the single clone analysis, we applied both tissue-culture based phenotypic tropism characterization, as well as sequence-based genotypic analysis towards 97 single envelope clones isolated from four samples of two study subjects. Unique quasispecies composition were found post coreceptor shift in two individuals who had different courses of disease. In addition, one subject showed drastic sequence variation between isolates from before and after treatment, and a highly homogeneous viral population post treatment. This suggested a rapid shift to CXCR4-using variants that accounted for failure to respond to CCR5 antagonist treatment.

In the deep sequencing analysis, we developed a novel comprehensive approach to sequence a large number of phenotypic validated variants on two next-generation sequencing platforms. This approach offered an unprecedented view of the viral quasispecies landscape *in vivo*, as well as on the dynamic population change in response to drug selection. Regions of interest that strongly associated with the usage of CXCR4 coreceptor were identified, including previously reported coreceptor specificity sites within and outside V3, the CD4 binding site, and gp41-gp120 interaction site within the same protomer and between neighboring protomers in the envelope trimer spike. These findings have potentially implications in rational design of better coreceptor inhibitors, and development of more accurate predictive algorithms for HIV-1 envelope tropism.

Chapter 1

Introduction

Human Immunodeficiency Virus (HIV) is an enveloped virus that enters into a susceptible host cell via a receptor-mediated fusion event between the host cell membrane and viral membrane (1-3). Like most enveloped viruses, this receptor binding process is mediated by a viral glycoprotein located on the lipid envelope of the virus, such as the envelope protein on HIV and the hemagglutinin protein on influenza virus (4, 5). The specificity of viral protein and host cell receptor interaction determines the range of host cell types that a virus could potentially infect, namely the viral tropism. Viral tropism is critical in disease pathogenesis and progression, prognosis, and treatment options. In this chapter, I will introduce some background knowledge on HIV virology and pathogenesis, envelope structure and cell entry process, tropism shift and its potential therapeutic implications, as well as the current and future tropism testing approaches.

1.1 HIV-1 and AIDS

HIV is a complex retrovirus and the etiologic agent of acquired immunodeficiency syndrome (AIDS) first discovered in 1983. A virion comprises of a nucleocapsid core, where its genetic information is stored in two copies of its positive sense single stranded RNA genome, and encapsulated by viral matrix proteins and nucleocapsid protein. On the exterior of the nucleocapsid core there is a roughly spherical-shaped lipid bilayer membrane called the envelope (6). On the lipid bilayer membrane, there are viral envelope proteins forming trimeric spikes that can interact with the host cell surface receptor and coreceptors to initiate an infection (7).

HIV/AIDS remains a major global public health problem today. About 37 million people world-wide are living with HIV infection as of the end of 2014. Over 34 million deaths have occurred to-date as a result of HIV infection. Each year there are 2 million new infection cases and approximately 1.2 million HIV-associated deaths. The pandemic is the most severe in regions of sub-Saharan Africa where there is heavy disease burden and poor access to testing and treatment (Figure 1.1) (UNAIDS) (8).

The term HIV refers to two distantly related viruses, HIV-1 and HIV-2, which were independently introduced from their zoonotic origins to humans (9, 10). HIV-1 is more pathogenic and responsible the global AIDS epidemic, whereas HIV-2 is less infectious, with cases mostly restricted to West African regions, and has a slower progression to full-blown AIDS. HIV-1 can be further categorized into Groups including M, N, O, and P, of which Group M (main) causes the majority of HIV-1 infection world-wide. Based on sequence diversity, Group M is further classified in subtypes A, B, C, D, F, G, H, J, K, and circulating recombinant forms (CRFs) (11, 12). The dominant subtype in the Americas and Europe is Subtype B, while over half of all HIV cases are caused by Subtype C, which is also the most dominant subtype in

Africa. The different subtypes represent rapid evolution of HIV-1 in various geographic areas as a direct consequence of its high mutation and recombination rate as well as its high turnover rate in the host.

The current standard of care for HIV patients is typically a combination of antiretroviral drugs in a treatment called highly active antiretroviral therapy (HAART) (13). HAART drugs target primarily three viral proteins, protease, reverse transcriptase, and integrase. Current WHO recommendation on first time baseline combination includes two nucleoside reverse transcriptase inhibitors (NRTI) and one non-nucleoside reverse transcriptase inhibitor (NNRTI), one protease inhibitor (PI) or one integrase inhibitor (WHO) (14, 15). HAART can effectively suppress circulating viral load, therefore, significantly reducing mobility and mortality of AIDS, but it does not provide a cure (16, 17). A drug holiday for one or two weeks can result in a blood viral load rebound back to the pre-treatment level (18). Therefore, the treatment has to be taken continuously for life. Apart from the public health issues that accessibility to HAART drugs are far from universal, taking antiretroviral drugs long term could lead to other complications including toxicity and development of drug-resistant strains of HIV (19-21). Taken together, the present treatment options for HIV/AIDS are insufficient to address the global AIDS pandemic, and more efforts are needed to understand the biology of HIV infection in the hope to provide guidelines in future drug and vaccine development processes.

1.2 HIV life cycle and pathogenesis

A productive HIV-1 infection at the cellular level starts with viral entry (Figure 1.2) (1, 6). Following attachment to the extracellular matrix or cell surface receptors of a susceptible cell, the viral envelop protein engages with a cell surface receptor CD4, undergoes a large conformational change, and then engages with its coreceptor, CCR5 or CXCR4 (22, 23). Upon coreceptor binding, further conformational changes occur, inserting the fusion peptide into the cellular membrane, pulling together the viral membrane and the cellular membrane to a critical distance, resulting in hemi-fusion and fusion of the two membranes. After the formation of fusion pore, the virion nucleocapsid is transported into the cytoplasm of the infected cells.

Once the HIV nucleocapsid core enters the cytoplasm, uncoating occurs and converts the nucleoprotein core into the reverse transcription complex (RTC) (24) where reverse transcription of the viral RNA genome into a double-stranded DNA copy takes place (25, 26). The RTC is then transformed into the pre-integration complex (PIC), chaperoned to the nuclear pore where the PIC enters the nucleus for integration. Upon integration, the viral DNA genome becomes part of the chromosomal DNA in the infected cell and is called a provirus. Expression of viral RNA and proteins are realized through cellular transcription and translation machineries. Newly manufactured viral constituents are transported to the plasma membrane of the infected cell, where viral assembly and budding occur. The gag-pro-pol polyprotein mediates the assembly process by packaging two copies of positive sense single-strand viral genomic RNA into the budding virion, recruiting accessory proteins and envelope protein, and generating a spherical viral particle. The host endosomal sorting complexes required for transport (ESCRT) system releases the particles from the plasma membrane of the infected cells (27). Upon budding, the

viral protease cleaves the Gag-Pol polyprotein, resulting in fully processed structural components that rearrange to form a mature nucleocapsid core capable of infecting a new host cell (6, 28, 29).

At an organismal level, HIV can be transmitted through sexual, percutaneous, and perinatal routes. A new infection via mucosal transmission is typically established by a single CCR5-tropic founder virus that infects a partially activated memory $CD4^+$ T lymphocyte near the site of exposure (30). The infected T cell then migrates to a draining lymph node to start a peripheral infection, which then quickly proceeds to a systemic infection within weeks. During the initial stages of infection, $CD4^+$ T cells in gut-associated lymphoid tissues (GALT) are massively depleted due to active infection as well as abortive infection induced through by-stander T cell pyroptosis (31). Cellular, and later, humoral immune responses counteract active viral replication and infection but are unable to clear the virus, resulting in a viral load set-point that usually is stable throughout the seven to ten years of the clinical latency period. In the later phase of infection, especially by the time immunodeficiency starts, the virus has largely depleted CCR5-expressing memory T cell compartment, and sometimes evolved to infect CXCR4-expressing naïve T cells. HIV/AIDS patients, if untreated, usually become so immunocompromised that they succumb to multiple severe opportunistic infections that are rare and non-life-threatening in immunocompetent individuals (32, 33).

A thorough understanding of the HIV viral life cycle and its disease pathogenesis is central to the development of intervening methods that block the replication cycle or stop disease progression. For example, entry inhibitors prevent the virus from infecting a target cell in the first place, thus, such agents could potentially be used in pre-exposure microbiocide creams or as post-exposure prophylactics (34). Entry inhibitors also prevent the newly released virus from undergo another round of infection, thus, they could be used as one component of the HAART

regimen cocktail in combination with other drugs to suppress viral replication in a chronically infected patient.

1.3 HIV-1 cell entry and entry inhibitors

Viral entry into a susceptible target cell is the first step in the virus replication cycle. The entry event might occur at the plasma membrane or within an endocytic compartment of a targeted cell in a pH-independent manner (1). Viral entry, mediated by the envelope protein, is a dynamic, multimeric, multi-step process that involves several protein-protein interactions, consecutive conformational changes, membrane fusion, and the eventual translocation of the viral nucleocapsid core (Figure 1.3) (28). Several key steps in the entry process are being targeted for the development of entry inhibitors (35-40).

During viral attachment, nonspecific interactions with the extracellular heparan sulfate proteoglycans, and specific interaction with cell surface proteins DC-SIGN and α4β7 integrin, bring the virion to a close proximity with the plasma membrane of a target cell. The viral envelope trimeric spike encounters and engages CD4, the primary host receptor required for HIV-1 infection (41-43). CD4 is an immune cell surface marker expressed on a subset of T lymphocytes ($CD4^+$ T cells), dendritic cells ($CD4^+$ DCs), and macrophages, all of which are cell types susceptible to HIV-1 infection. CD4 binding to gp120 leads to a conformational change that forms a bridging sheet, and allows an otherwise buried region including variable loop (V3) to be repositioned and exposed, forming a critical contact platform for the subsequent coreceptor interaction step (44).

Human chemokine receptors, CCR5 and CXCR4, are the two major coreceptors playing significant roles in viral entry *in vivo*. Both CCR5 and CXCR4 are G-protein coupled seven-transmembrane receptors that are expressed on a variety of cell types including certain immune cell subsets. In a normal immunological context, interaction between chemoattractant cytokines and their receptors interact and modulate immune cell migration and sometimes influx into sites

of injury and inflammation. CXCR4 is widely and constitutively expressed on many cell types, including thymocytes and naïve T cells that require CXCR4-CXCL12 interaction for development and homeostasis (45). In a pathological context, CXCR4 also plays an important role in cancer cell survival and metastasis. On the other hand, CCR5 has a narrower expression profile including activated and memory T cells. CCR5 interacts with its ligands MIP1α, MIP1β and RANTES for T cell costimulation and inflammatory immune responses, though it appears to be dispensable, with no severe immunodeficiency in people carrying homozygous null alleles (46).

The process of envelope-coreceptor interaction is not clearly defined but is thought to occur according to a two-site model. Site one is the amino-terminus of a coreceptor that interacts with the base of V3 loop, and site two is the extracellular loops of a coreceptor that interacts with the now exposed V3 region on gp120 upon CD4 binding (47, 48). As the coreceptors fully engage gp120, further conformational changes allow the hydrophobic gp41 fusion peptide to inserts into the host membrane, while the transmembrane domain of gp41 still anchors its carboxyl terminus in the viral membrane. As a Class I fusion machinery, two heptad repeat (HR) regions in gp41, HR1 and HR2, fall back on each other in antiparallel orientation, forming a coiled-coil six-helical bundle that pulls together the viral and host membrane to a proximity close enough for mixing of the outer leaflet of the two membranes, and eventual formation of a fusion pore (49-52). The viral nucleocapsid is then translocated through the pore into the host cell cytoplasm.

A number of therapeutic developments focus on blocking different steps in the entry process, such as CD4 binding, coreceptor binding, and fusion (34, 53, 54). For example, vaccine engineering research focuses on broadly neutralizing antibodies that target native envelope

trimers in an attempt to block viral attachment or entry (55-60). Small molecules and monoclonal antibodies are tested for their ability to target CD4 or the CD4 binding site on gp120 (61-63). Fusion inhibitors include Enfuvertide (Fuzeon, Roche), which is a 36-amino-acid-long HR2-based peptide that competes with HR2 and binds to HR1 in gp41 region , therefore preventing the formation of the six-helical bundle and stopping fusion (64-70). With the discovery of a rare, naturally occurring, protective loss-of-function CCR5Δ32 mutation, blocking or removing CCR5 from the host cell became an attractive strategy (71, 72). Small molecule antagonists and gene therapy agents such as Zinc finger nucleases (ZFN) have been studied in the laboratories and in clinical trials (73-78). Maraviroc (MRV), a small molecular CCR5 antagonist, is the only coreceptor inhibitor approved by Food and Drug Administration (FDA) as a second line treatment for patients who have exhausted first line options (48, 79-83). Inhibitors to CXCR4 were developed but not approved for long-term HIV treatment due to its adverse effects. Single dose administration of AMD3100 (Mozobil or Plerixafor), a potent CXCR4 inhibitor, is approved to be used in bone marrow donors to mobilize bone-marrow residing hematopoietic stem cells before transplant but is not used for HIV-1 treatment due to side effects (84, 85).

Overall, HIV-1 viral entry is a multi-step process that remains an attractive target for active development of entry inhibitors (86-88). Although a number of small molecule drugs and biologics are being studied in clinical trials, only two were FDA approved to be used as second line HIV drugs. More effort to understand the molecular details between HIV-1 envelope and coreceptor interaction during cell entry can provide valuable guideline for the future development of efficacious, precisely-targeting therapeutic agents.

1.4 Envelope structure and tropism determinants

The native HIV-1 envelope protein exists as a trimer of gp120-gp41 heterodimers on the viral membrane. The full-length envelope precursor protein gp160 is cleaved into a transmembrane subunit called gp41 and a surface subunit called gp120 (89). The two subunits are non-covalently attached, and form a trimer of heterodimers on the viral lipid membrane, called an envelope trimer spike (90). Based on structural studies and models, one HIV-1 virion has approximately 14 envelope spikes, on average, on its lipid membrane, which is a very low number compared to other enveloped viruses, such as influenza and dengue viruses (7, 91). The sparse location of envelope spikes on a virion is thought to minimize crosslinking of B cell receptor and elicitation of a humoral immune response (92). In addition, gp120 is heavily glycosylated with over 20 glycosylation sites (93). The glycan shield on the exterior of the envelope spike also helps to evade the neutralizing antibody response (94, 95).

HIV-1 gp120 can be further divided into five relatively conserved regions (C1 through C5) that form the gp120 core, interspersed by five variable regions (V1 through V5) that are mostly located at or near the outer surface of gp120 to shield the more conserved inner core from the immune system (Figure 1.4) (96-99). Disulfide bridges link the two cysteine residues at the base of each variable loop except for V5. The variable loops are named due to their sequence hyper-variability among different envelope isolates. Among the five variable loops, the V3 loop is much less variable than V1/2 and V4, especially in terms of insertions and deletions (44).

The V3 loop has been extensively studied for its critical role in coreceptor binding (Figure 1.5) (96). Sub-regions in V3 have been shown to be conserved, especially in CCR5-using viruses (R5 viruses), highlighting the possibility that V3 interacts directly with CCR5 with some degree of structural stringency. On the other hand, the V3 sequences in CXCR4-using viruses

(X4 viruses) are more diverse. Currently, a plethora of mutagenesis and structural studies support the theory of a CCR5 amino-terminal interaction with the base of V3, in particular a sulphated tyrosine at Position 10 of CCR5 forming a salt bridge interaction with a highly conserved arginine at Position 327 of gp120. Computational docking simulations also demonstrated a possible interaction between the CCR5 amino-terminus with the CD4-bound gp120 structure. However, there is no direct crystallographic evidence or molecular details of the direct interaction between the V3 loop and CCR5 or CXCR4 extracellular domains.

Besides being a critical piece in coreceptor binding, the V3 loop is also the principal determinant in coreceptor specificity (100, 101). The functional capacity of gp120 to interact with CCR5 and/or CXCR4 bears clinical significance in treatment options and disease prognosis. Studies show that in an artificial pseudotyping system, exchanging the V3 loops between CCR5-tropic and CXCR4-tropic envelope strains can result in a switch in tropism, demonstrating the importance of V3 in coreceptor choice (102). However, *in vivo* envelope mutation and evolution in a physiological context is more complex. For example, the 11/25 Rule states that positively charged residues on the V3 loop in particular at Position 11 and 25, as well as the net positive charge of V3, are strongly associated with CXCR4 usage.

A large body of literature has focused on the V3 loop, as it is the principal determinant of HIV-1 tropism (103-107). However, HIV-1 envelop also modulates its interaction with the coreceptor via regions outside the V3 loop (108-112). Covariance of a negatively charged amino acid at Position 440 in C4 region and a positively charged residue at Position 322 in V3 were shown to associate with strong with R5-using Subtype B virus; incorporating the status of Position 440 could improve the predictive power of existing tropism predictive algorithms. Mutations in gp41 correlate highly with X4-usage and are validated in patients in MRV clinical

trials (113). Biochemical and biophysical analysis also showed that residues that influence tropism are identified in V1, V2, C4, and gp41 regions (114). In particular, the tropism-determining residues of gp41 likely influence gp120-coreceptor interaction via modulating binding affinity and complex stability (115). Since the envelope-coreceptor interaction is a multimeric, highly dynamic process, residues that are not located at the contact surface might still influence the tropism indirectly by stabilizing or destabilizing the interaction with a coreceptor (116). A more thorough and nuanced understanding of gp120-gp41-CD4-coreceptor interaction, including the role of V3 and determinants outside V3, would potentially improve tropism predictive algorithms and reveal new drug targets (117).

1.5 HIV-1 quasispecies and coreceptor shift

HIV-1 exists in patients not as a homogenous clone of virus with identical genomes, but as a quasispecies, or a group of related but non-identical variants. This population of variants is a result of high levels of viral replication soon after the initial infection (118). As previously mentioned, a new infection usually starts from a single R5-tropic transmitter / founder virus. Upon establishment of a productive infection, rapid replication occurs, with mutations and recombination accumulating, and selected along the way for improved fitness in the host. The error-prone viral reverse transcriptase, in combination with rapid viral reproduction, results in an estimated mutation rate that allows changes at every single nucleotide position in the viral genome within a day (119-122). Therefore, a viral quasispecies population in the host is adaptable to various selection pressures coming from the host immune system or from antiretroviral therapies (123). The HAART regimen was developed based on this understanding, that the virus can quickly become resistant to any mono-therapy, but it is less likely to gain resistance to multiple drugs administered simultaneously.

In the context of tropism, at the early stage of HIV-1 infection, CCR5-tropic virus predominates largely over the asymptomatic period, probably reflecting a bottleneck event in the transmission process. As the CCR5-expressing T cells are depleted, selection pressure drives the quasispecies to utilize the alternative coreceptor CXCR4 and in the presence of less CD4 expression. Current nomenclature broadly categorizes HIV-1 isolates into CCR5-tropic (R5), CXCR4-tropic (X4), and CCR5/CXCR4-tropic (dual or R5X4) clones. In reality, however, strictly R5-tropism and strictly X4-tropism are the opposite ends of a spectrum, where most physiological variants fall on the spectrum in between the two ends. Strictly X4-tropic variants are rare, as most infections start with a R5-tropic virus and evolve towards different levels of

capacity in using CXCR4. Approximately 50% of HIV-1 Clade B patients spontaneously develop X4-using viruses in the later stage of infection, and the presence of X4-using virus (including dual-tropic and X4-tropic) correlates with a worse clinical prognosis (124). It is unclear whether the emergence of X4-using virus is the cause or result of a more severe disease.

Some patients receiving CCR5 antagonist therapy experience an expedited course of X4-using virus emergence and become unresponsive to the coreceptor inhibitor treatment (Figure 1.6). This could lead to treatment failure and viral load rebound. The conversion from R5-tropism to dual- or X4-tropism is called coreceptor shift. Again it is unclear if this presence of X4-using virus will cause worse disease outcome, or whether it is the result of disease progression. In some of the individuals who undergo coreceptor shift, the viral tropism reverts back to R5 upon withdrawal of the CCR5 antagonist therapy, reflecting the adaptability of the quasispecies population in response to external selection pressures, as well as a potential fitness cost in X4-usage (125).

1.6 Phenotypic tropism testing

The current gold standard for viral tropism determination is tissue-culture based phenotypic tests, such as the Enhanced Sensitivity Trofile Assay (ESTA) (126-128). The phenotypic testing is usually performed by pseudotyping patient-derived gp160 coding sequence on infectious particles, and infecting cell lines expressing CD4 and CCR5 or CXCR4 using these pseudotyped viruses. The infection outcome is detected with a reporter system such as fluorescence or luminescence. The advantage of a phenotypically based assay is that it is highly accurate, sensitive, and includes all tropism determinants on the full-length gp160 envelope region. The ESTA reports 100% sensitivity at detecting 0.3% X4-using minor variant at a starting viral load larger than or equal to 1000 copies per milliliter of blood. The disadvantages of a phenotypic assay include high cost, long turn-around time, sometimes inconclusive results, and being labor-intensive. Since clinical administration of a CCR5 antagonist requires prior tropism determination that the patient does not harbor X4-using variants, such drawbacks of the phenotypic tropism testing directly limit the application of the CCR5 antagonist drug.

1.7 Genotypic tropism testing

With the impressive advancement of the next-generation-sequencing (NGS) technology over the past decade, genotypic analysis of HIV-1 quasispecies becomes feasible (129-136). NGS technology was developed to meet the demand for higher-throughput, cheaper sequencing methods in the Human Genome Project, and has become increasingly useful in the study of microbial communities called the microbiome. The ability to produce sequencing libraries in a cell-free system without cloning, the massive parallelization of sequencing reactions from the advancement of microfluidics technology, and the real-time base-interrogation technology without the need for electrophoresis, allow millions and billions of sequencing reads to be produced in one run (137). The unprecedented scale and speed of NGS opened doors for genomic and genotypic approaches in disease diagnosis and testing, including HIV-1 tropism testing.

Illumina sequencing technology is the most widely used mainstream NGS platform today. It has different sequencing systems developed for different applications based on genome size and output volume. The MiSeq system is designed to sequence small genomes, amplicons or targeted gene panels with a fast turnaround time (within 6 hours), ideal for genotypic determination in clinical laboratories (138). The basic workflow of Illumina NGS system includes fragmentation of input DNA into several hundred base-pairs, ligation of an adaptor on both ends of the fragments, immobilization of the adaptor-DNA fragment on a flow cell, solid-phage bridge amplification to produce millions of densely packed clusters, and sequencing by synthesis using labelled nucleotide terminators (Figure 1.7). Each cluster comes from one DNA fragment in the input sample, and will produce one read after sequencing. A CCD camera is used to capture fluorescence as each terminator base is being added to the growing DNA polymer, and

the output reads are typically 50 – 250 nucleotides in length, depending on the sequencer and customized settings. The advantage of Illumina sequencing lies in its massive amount of output, ranging from millions to billions of reads per run. This allows hundreds to tens of thousands folds of coverage for a small genome or an amplicon gene. It is ideal to uncover rare variants that comprises of less than 1% of the input population, for example in a patient-derived HIV-1 envelope amplicon sample where the majority of the variants are R5-tropic but minor X4-tropic variants exist at a low percentage (139-142). The challenge for genotypic testing by Illumina sequencing comes from the need to reconstruct short reads into a longer haplotype. It is difficult to analyze mutations that are longer than the length of the reads, as the linkage information would be lost over the fragmentation and sequencing process.

In order to address the issues brought about by the short read length in Illumina technology, several new cutting-edge "third-generation sequencing" technologies that focuses on longer sequence of single DNA molecules emerged. I will still call them next-generation sequencing for the sake of simplicity. Pacific Biosciences (PB) single-molecule, real-time (SMRT) sequencing is one of them (Figure 1.8) (143, 144). Some breakthroughs for the PB sequencing technology include (1) utilization of zero-mode waveguides that are small wells on a SMRT cell to house single DNA molecule, therefore blocking light from other sequencing reactions and reduce background noise; a novel DNA polymerase and sequencing chemistry that can generate long reads up to tens of kilobases in length; and (3) sequencing in a circular fashion to allow multiple passes on the input DNA fragment on both strands, therefore resulting in a circular consensus sequence (CCS) for each single molecule sequenced and reducing error rate. Comparing to Illumina platforms, PB sequencing generates fewer reads in the range of thousands to tens of thousands, therefore the coverage is not as high but still more than sufficient in

analysis of a small genome or amplicon. In the context of genotypic testing for HIV-1 tropism, the added advantage of PB sequencing is that mutation located at different domains of the envelope gene can be correlated thanks to the long CCS reads. It affords analysis of covariance between different loci and allows better understanding of structure-function relationship of the envelope.

Since the NGS-based genotypic assays routinely multiplex samples to maximize the utilization of its massive sequencing capacity, development of high-throughput low-cost genotypic testing methods is a possibility. With a deep coverage and robust analysis pipeline, it is possible to detect minor variants at a comparable or higher sensitivity than the current ESTA approach (146-149). The limit of detection of NGS-based genotypic assay is based on the level of noise, namely the intrinsic combined error rate of the sample preparation procedures and the sequencing chemistry, which hopefully decreases as the NGS technology advances. Due to its incredible depth of coverage, NGS-based genotypic testing offers a more detailed, high resolution picture of quasispecies landscape in patients, allowing doctors to make more informed decisions (150). In addition, a NGS-based genotypic assay could potential sequence the entire genome of HIV-1, and detect mutations that confer resistance in multiple viral targets across different classes of antiretroviral drugs in an all-inclusive multiplex style (151), providing a one-stop complete diagnostic and monitoring assay for HIV patient management.

Major disadvantages of the NGS-based genotypic testing include bioinformatic challenges in data analysis, and prior knowledge of structure-function relationship in result interpretation (152). In HIV-1 genotypic tropism determination, the latter is especially challenging, given that the envelope sequence is so variable, and we do not understand all tropism determinants in gp160 sequence. Currently available computational algorithms,

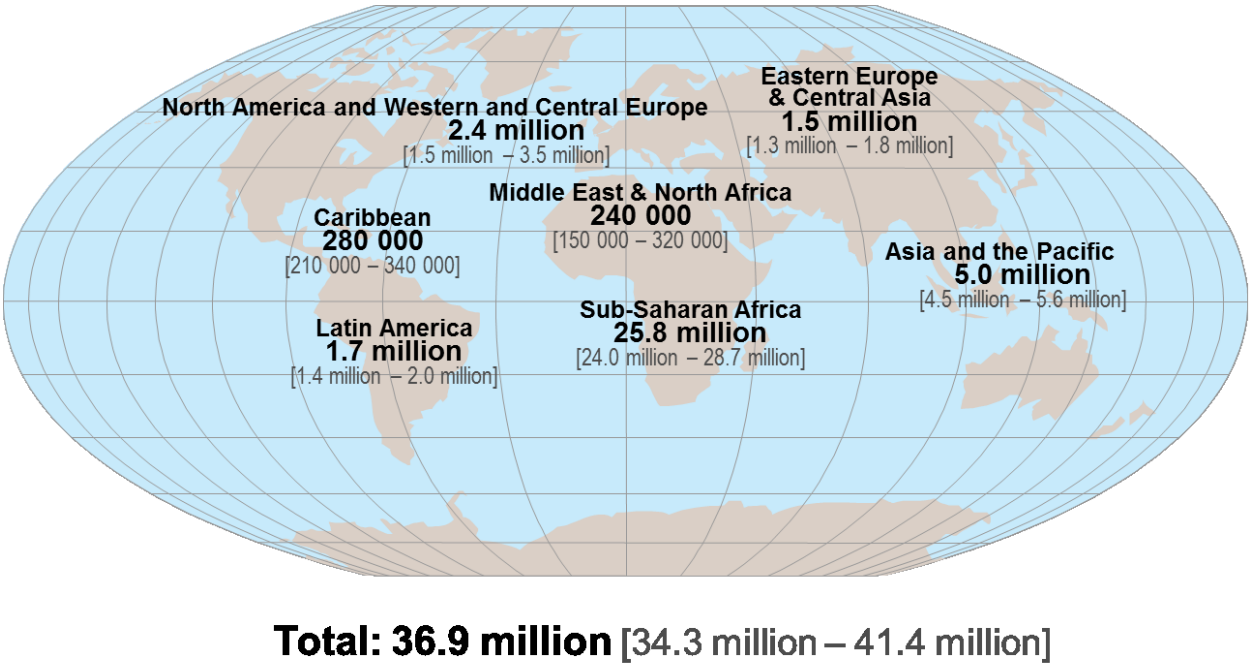
Geno2Pheno[coreceptor] and Web PSSM, are designed to make tropism predictions based on the V3 loop sequence, the major but not only determinant for tropism (153-159). These algorithms are trained on a limited set of Subtype B sequences that only accounts for 10% of the global circulating strains, and are not as accurate in other subtypes that account for the majority of HIV-1 infections worldwide (160, 161). In addition, even for Subtype B V3 sequences, the algorithms achieve high specificity (few false positives) but low to moderate sensitivity (many false negatives) at predicting X4-using variants (162). The current US and European guidelines for MRV administration based on genotypic testing results predicted by Geno2Pheno[coreceptor] recommend a false positive rate cut off that ranges from 2 – 5.75% based on MOTIVATE trial data, up to 10%, highlighting the inadequacy of the current predictive algorithms (163-165). One potential way to improve the predictive power is to incorporate other non-V3 determinants that differentiate X4-using variants into the algorithms (166). This requires more research to understand the structure-function relationship of HIV-1 envelope in the context of coreceptor binding and specificity. A streamlined, optimized genotypic assay would ideally be high-throughput, fast-turnaround, low cost, and highly specific and sensitive, so more patients eligible for MRV and other future coreceptor inhibitors could benefit from the respective therapy.

1.8 Conclusion

Over the past three decades, we have come a long way in understanding HIV biology, designing optimized treatment methods, and alleviating the morbidity and mortality associated with the disease. However, HIV/AIDS remain a public health priority today, with 2 million new cases each year, increasing prevalence of drug-resistant strains, and lack of access to testing and treatment. The current HAART drugs can effectively suppress viral load and pathology but do not provide a cure. As a result, HIV patients need to take HAART therapy for the rest of their lives, raising concerns with long term toxicity and resistance development. Therefore, there is a continuous need for development of new treatment options and vaccines. Coreceptor inhibitors are a new class of drugs that could potentially benefit patients if administered properly. Further understanding of HIV-1 envelope protein interaction with coreceptors could shed light on future inhibitor design and tropism predictive algorithm improvement.

1.9 Figures

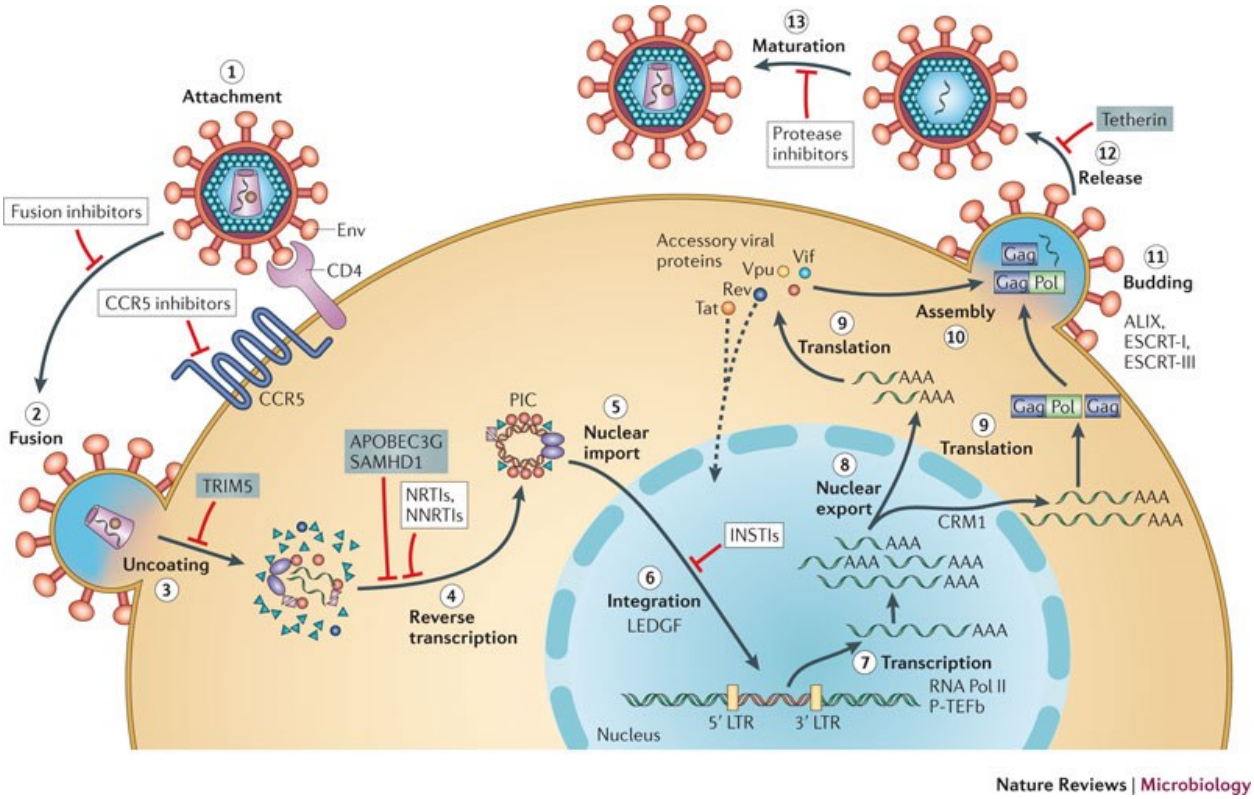
Figure 1.9.1 Adults and children estimated to be living with HIV by the end of 2014.



An estimated 37 million people are living with HIV worldwide. Approximately 70% of all people infected with HIV reside in Sub-Saharan Africa.

This figure is adapted from UNAIDS Core Epidemiology Slides (July 2015).

Figure 1.9.2 HIV-1 replication cycle.

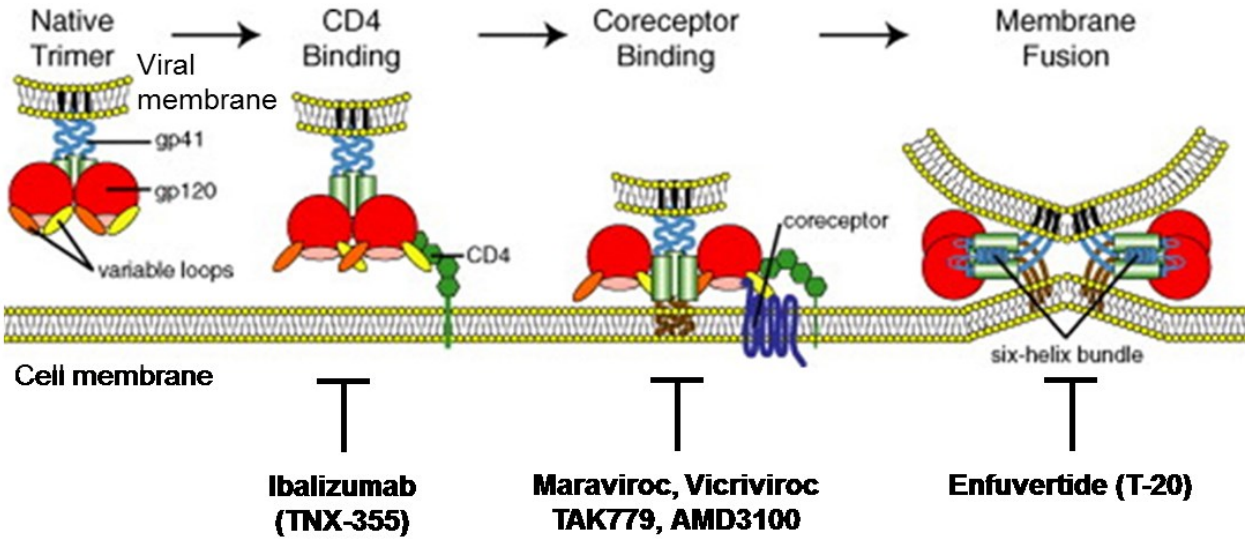


A range of host cell proteins play a role as restriction factors or dependency factors in the process of HIV life cycle. Different steps in the life cycle are numbered as the following: (1) attachment and binding to cell surface receptors CD4 and CCR5 or CXCR4; (2) viral fusion and translocation of viral nucleocapsid into host cell cytoplasm; (3) uncoating of viral nucleocapsid; (4) formation of reverse transcription complex and reverse transcription; (5) formation of pre-integration complex (PIC) and nuclear import of PIC through nuclear pore; (6) integration of pro-viral genome into host cell genome; (7) transcription from the provirus to make viral genomic RNA and mRNA; (8) nuclear export of viral genomic RNA and mRNA; (9) translation of viral mRNA into accessory proteins and structural proteins (gag-pol); (10) assembly of viral RNA genome and gag-pol on the cell plasma membrane; (167) budding of new virions on the cell plasma membrane; (12) release of new virions from infected cell membrane via ESCRT machinery; and (13) maturation of newly release virion into infectious particles via proteolytic

cleave and rearrangement to form a viral nucleocapsid core. Each of the above steps can be targeted for intervention. White and blue boxes indicate therapeutic inhibition and cellular restriction. INSTI, integrase strand transfer inhibitor; LTR, long terminal repeat; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor.

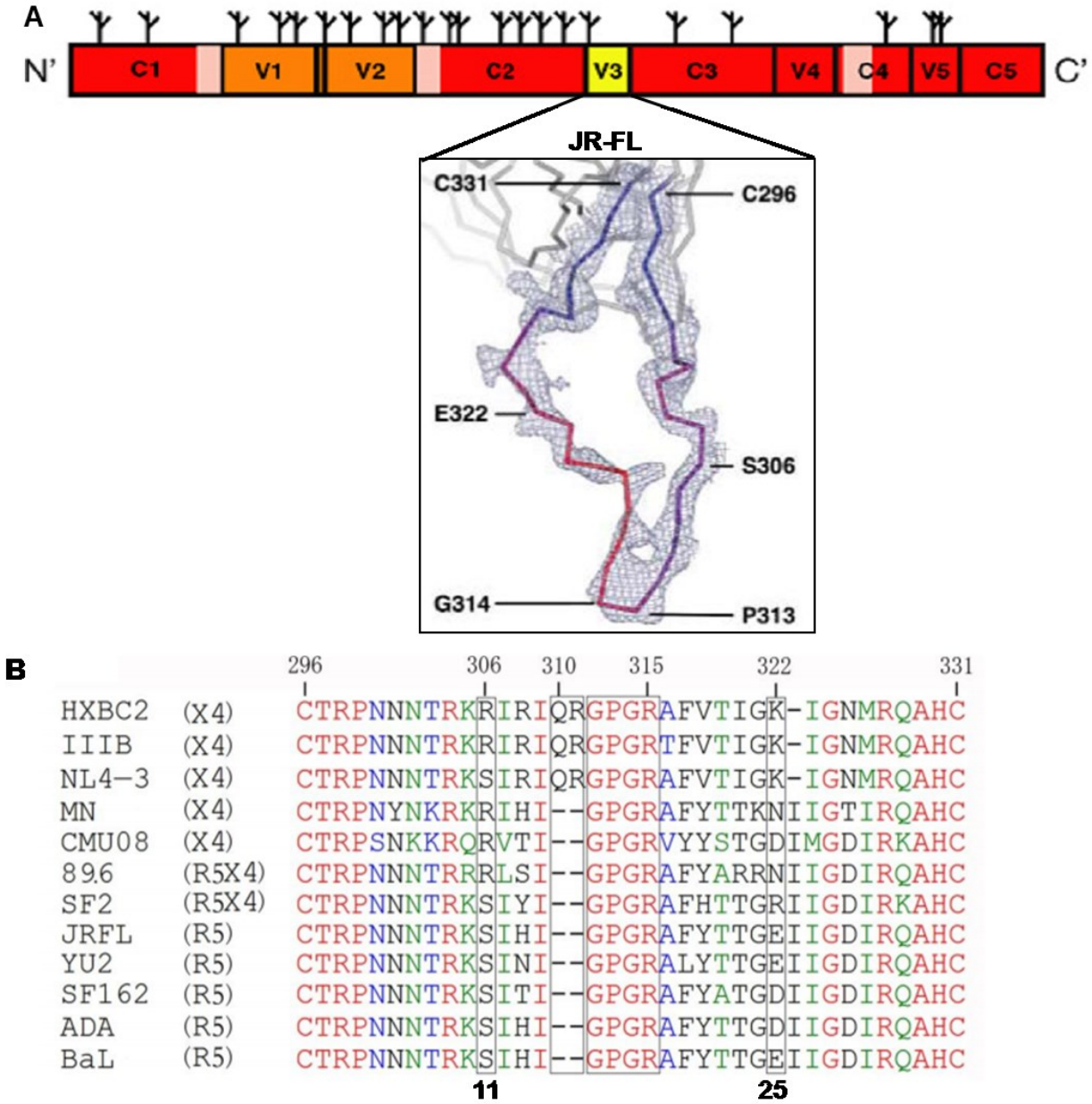
This figure is adapted from Alan Engelman and Peter Cherepanov, Nature Reviews Microbiology 10, 279-290 (April 2012)

Figure 1.9.3 The multi-step HIV entry process mediated by envelope protein.



The HIV-1 envelope protein in native trimer form (left) is shown with gp41 transmembrane subunit (blue) and gp120 surface subunit (red). Variable loop regions on gp120 are shown in yellow and orange. Upon gp120 engagement with CD4 (green) on cell membrane, the variable loops change conformation, exposing the contact site for the subsequent interaction with coreceptor CCR5 or CXCR4 (navy). Coreceptor binding results in the insertion of gp41 fusion peptide (brown) into the cell membrane, which eventually leads to six-helix bundle formation and membrane fusion (right). Therapeutic and research agents that inhibit the above-mentioned processes are shown below each step. Ibalizumab (TNX-355) is a monoclonal CD4 antibody that binds to CD4 and blocks CD4 binding. Maraviroc, Vicriviroc, and TAK779 are CCR5 antagonists that bind to CCR5 and blocks coreceptor interaction. AMD3100 is a CXCR4 inhibitor that blocks CXCR4-tropic viral entry. Enfuvertide (T-20) is a peptide drug that competitively blocks the formation of six-helix bundle and prevents membrane fusion.
This figure is adapted from Bridgette J. Connell and Hugues Lortat-Jacob, Front. Immunol.(20 November 2013)

Figure 1.9.4 Envelope gp120 structure and the V3 loop



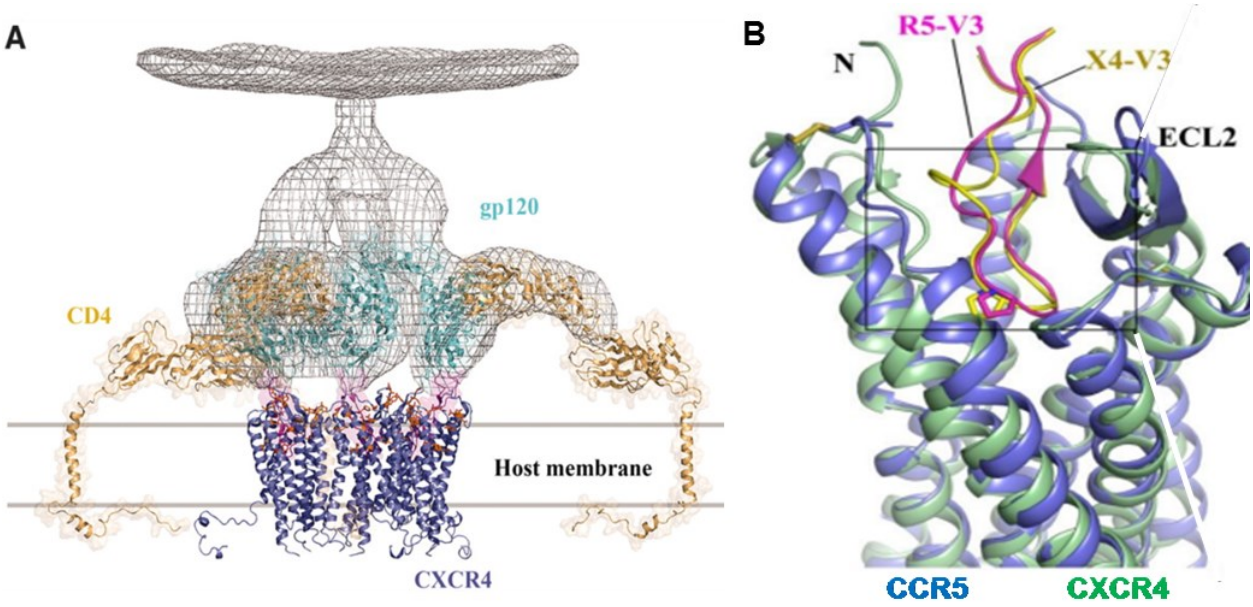
(A) Schematic representation of HIV-1 envelope gp120, with V3 crystal structure underneath. The envelope surface subunit gp120 is shown in its primary structure, and annotated with its conserved regions C1-C5, and variable regions V1-V5. Glycosylation sites are represented as branches on the primary structure. V3 loop crystal structure for isolate JR-FL is shown in electron density map. Two cysteine residues C396 and C331 linked by disulphide bridge form

the base of the V3 loop. The 11th and 25th position on the V3 stem, known to associate with coreceptor specificity, are annotated as S306 and E322 respectively. The tip of V3 contains a GPGR motif, annotated by P313 and G314.

(B) Alignment of HIV-1 envelope isolates V3 sequences. The tropism of each V3 loops is annotated on the side. Conserved residues are shown in red, including the starting and ending cysteine, as well as the GPGR motif at the tip. Position 11 and 25 are annotated.

This figure is adapted from Tilton JC, Doms RW, Antiviral Res., 2010, Huang C. *et al*, Science, 2005, and Tan Q et al., Science, 2013.

Figure 1.9.5 Computational models of HIV-1 envelope-coreceptor interaction.

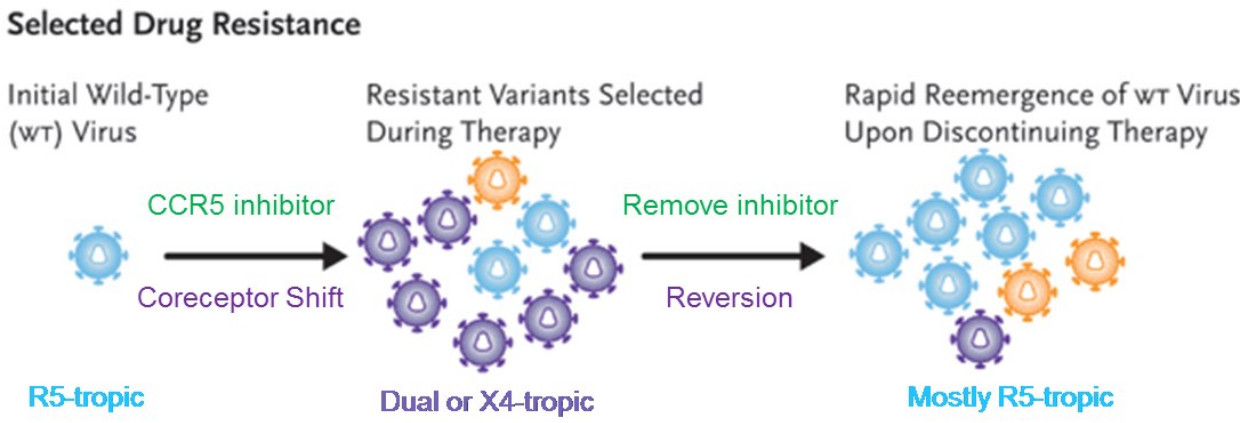


(A) HIV-1 envelope trimer binding to host cell surface CD4 and coreceptor CXCR4. The interaction is a multi-subunit conglomerate, with the stoichiometry of gp120, CD4, and coreceptors not yet clearly defined. Envelope and CD4 are shown as electron density map, while CXCR4 is shown in cartoon format. The parts are colored as the following: gp120, teal; CD4, gold; CXCR4, navy. The V3 loops from gp120 are colored red, and the contact residues on CXCR4 are colored orange.

(B) CCR5-tropic V3 loop (R5-V3) and CXCR4-tropic V3 loop (X4-V3) interaction with CCR5 and CXCR4 extracellular domains respectively are shown superimposed. The V3 loop backbone makes hydrogen bond with the backbone of extracellular loop 2 (ECL2), while the Proline313 at the GPGR motif at the tip of V3 makes hydrophobic interaction with the helices of the coreceptors. The parts are colored as the following: R5-V3, pink; X4-V3, yellow; CCR5, blue; CXCR4, green.

This figure is adapted from Wang B et al., Science, 2010, and Tan Q et al., Science, 2013.

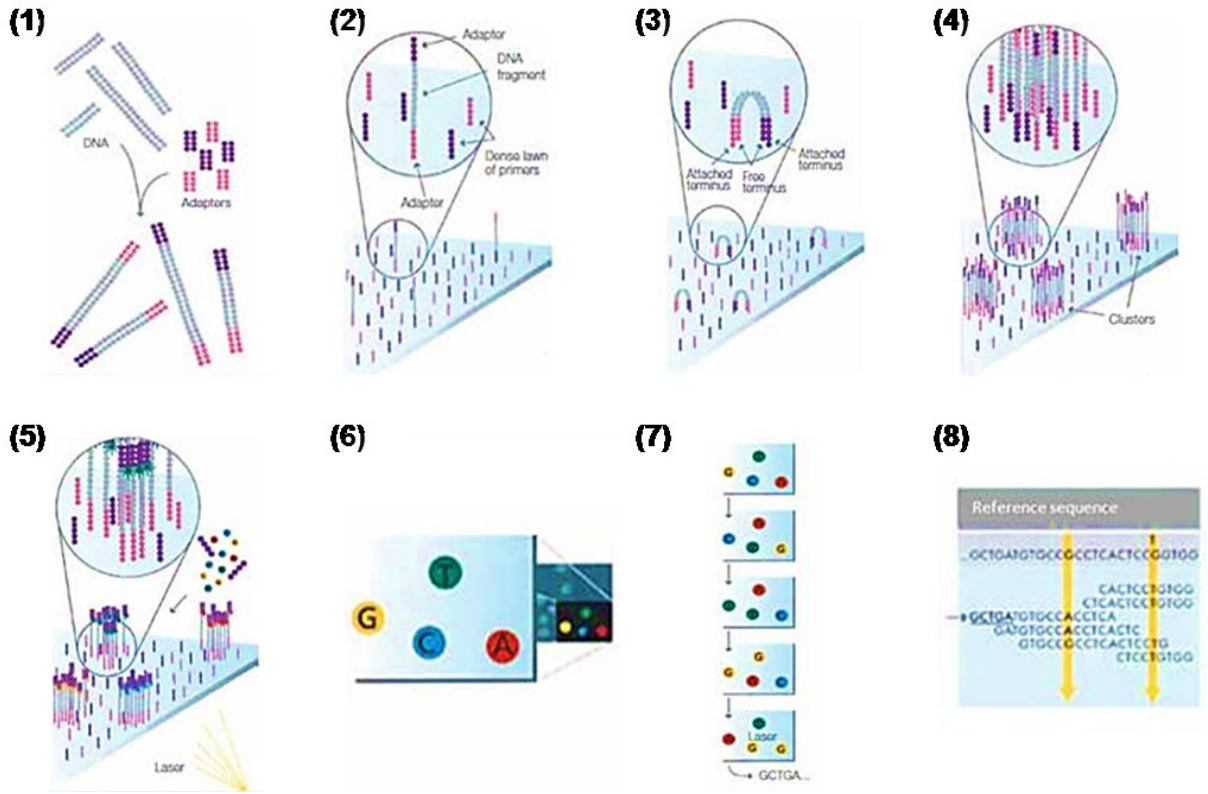
Figure 1.9.6 Drug resistance development process in a quasispecies population.



In the context of tropism shift and resistance to coreceptor inhibitors, the starting founder virus has a R5-topic wildtype phenotype. Upon selection with R5 inhibitors, resistant variants preferentially expand into the majority of the quasispecies population. Upon therapy withdrawal, the quasispecies population may demonstrate a reversion into the original R5-topic variants.

This figure is adapted from Kuritzkes DR, PRN, 2004.

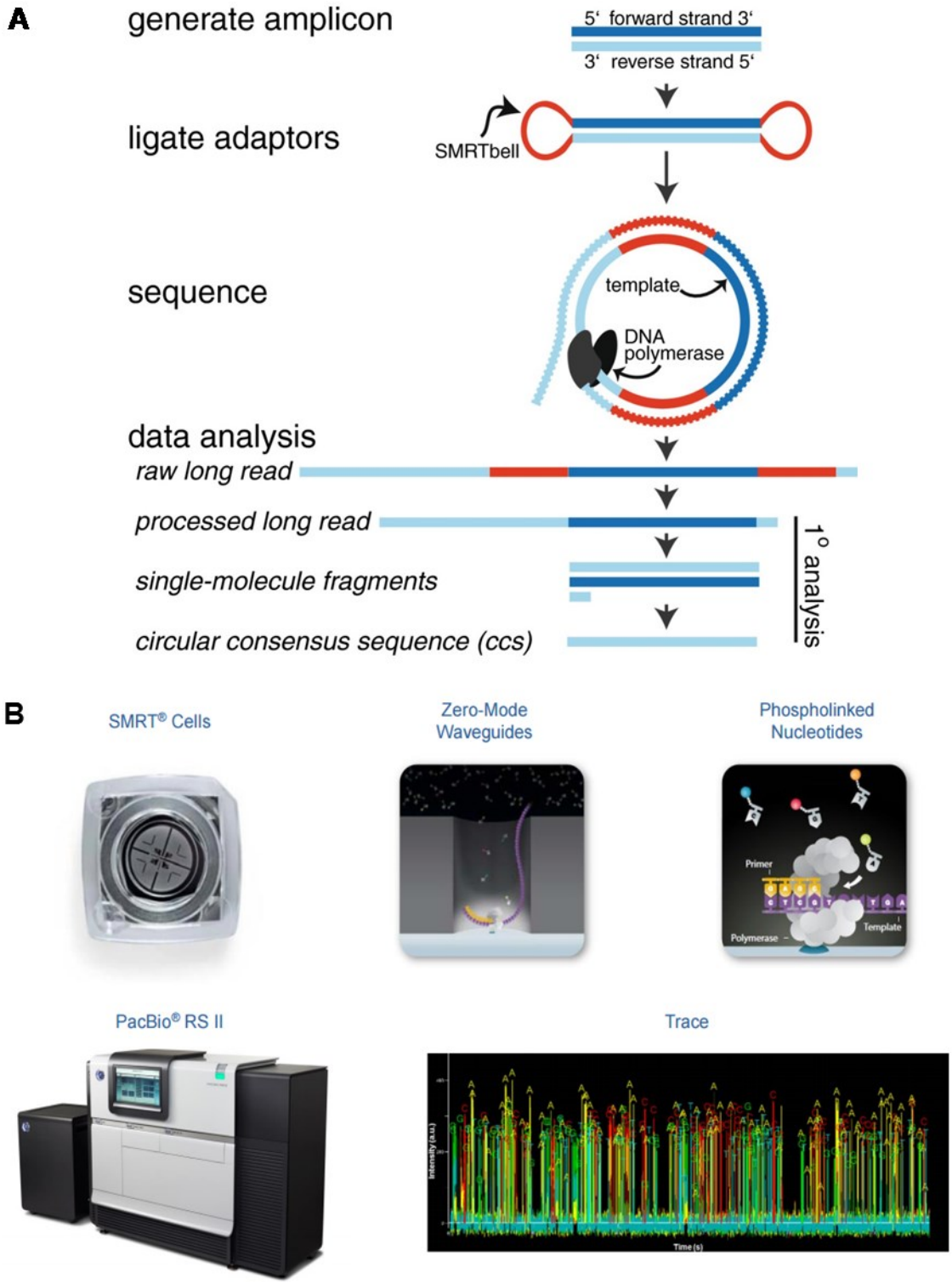
Figure 1.9.7 Illumina next-generation sequencing (NGS) technology work flow.



The steps of Illumina NGS sequencing is the following: (1) random fragmentation of intake DNA samples, and ligation of adaptor to both ends of the fragments; (2) immobilization of single stranded adaptor-DNA fragment to the flow cell surface; (3) Solid-phase bridge amplification using unlabeled nucleotides; (4) formation of several million dense clusters of double-stranded DNA from Step (3); (5) addition of labeled terminators, primers, and DNA polymerase to start sequencing cycles. (6) captures of emitted fluorescence from each cluster upon laser excitation as the DNA elongates; (7) Repeat of Step (5) and (6) to determine the sequence in the DNA fragment, one base at a time; and (8) Sequences of input DNA fragments as short reads to be mapped onto a reference sequence for analysis.

This figure is adapted from www.illumina.com

Figure 1.9.8 Pacific Biosciences third-generation sequencing technology work flow.



(A) Generation of circular consensus sequences (145) on single DNA molecules. Adaptors (SMRTbells) are ligated on both ends of double-stranded DNA fragments. Upon denaturation, a single-stranded circular DNA is obtained for sequencing, DNA polymerase can sequence the circle multiple passes to generate a consensus sequence of high quality.

(B) SMRT cells for PB system contains zero-mode waveguides (ZMW) that houses single DNA molecules during sequencing reaction. The DNA polymerase incorporates labeled nucleotides, and the fluorescence from each ZMW are captured and processed into a DNA sequence (trace).

This figure is adapted from Fichot EB and Norman RS, Microbiome 2013, and http://www.pacb.com/.

1.10 References

1. **Wilen CB, Tilton JC, Doms RW.** 2012. HIV: cell binding and entry. Cold Spring Harb Perspect Med **2**.

2. **Sun SX, Wirtz D.** 2006. Mechanics of enveloped virus entry into host cells. Biophys J **90:**L10-12.

3. **Cosset FL, Lavillette D.** 2011. Cell entry of enveloped viruses. Adv Genet **73:**121-183.

4. **Armstrong RT, Kushnir AS, White JM.** 2000. The transmembrane domain of influenza hemagglutinin exhibits a stringent length requirement to support the hemifusion to fusion transition. J Cell Biol **151:**425-437.

5. **White JM, Delos SE, Brecher M, Schornberg K.** 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit Rev Biochem Mol Biol **43:**189-219.

6. **Engelman A, Cherepanov P.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol **10:**279-290.

7. **Schiller J, Chackerian B.** 2014. Why HIV virions have low numbers of envelope spikes: implications for vaccine development. PLoS Pathog **10:**e1004254.

8. **Peeters M, Toure-Kane C, Nkengasong JN.** 2003. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. AIDS **17:**2547-2560.

9. **Sharp PM, Hahn BH.** 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med **1:**a006841.

10. **Bartolo I, Rocha C, Bartolomeu J, Gama A, Marcelino R, Fonseca M, Mendes A, Epalanga M, Silva PC, Taveira N.** 2009. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: new insights into the origins of the AIDS pandemic. Infect Genet Evol **9:**672-682.

11. **Taylor BS, Hammer SM.** 2008. The challenge of HIV-1 subtype diversity. N Engl J Med **359:**1965-1966.

12. **Hemelaar J, Gouws E, Ghys PD, Osmanov S.** 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS **20:**W13-23.

13. **Giannakopoulou A, Schornack S, Bozkurt TO, Haart D, Ro DK, Faraldos JA, Kamoun S, O'Maille PE.** 2014. Variation in capsidiol sensitivity between Phytophthora infestans and Phytophthora capsici is consistent with their host range. PLoS One **9:**e107462.

14. **Brock JB, Henderson H.** 2015. Antiretroviral Therapy. J Miss State Med Assoc **56:**166, 171-162.

15. **Looney D, Ma A, Johns S.** 2015. HIV therapy-the state of art. Curr Top Microbiol Immunol **389:**1-29.

16. **Gallant JE.** 2002. Initial therapy of HIV infection. J Clin Virol **25:**317-333.

17. **Wirden M, Todesco E, Valantin MA, Lambert-Niclot S, Simon A, Calin R, Tubiana R, Peytavin G, Katlama C, Calvez V, Marcelin AG.** 2015. Low-level HIV-1 viraemia in patients on HAART: risk factors and management in clinical practice. J Antimicrob Chemother **70:**2347-2353.

18. **Mascolini M.** 1999. Salvage solitaire (or, HAART takes a holiday). J Int Assoc Physicians AIDS Care **5:**10-23.

19. **Belec L, Piketty C, Si-Mohamed A, Goujon C, Hallouin MC, Cotigny S, Weiss L, Kazatchkine MD.** 2000. High levels of drug-resistant human immunodeficiency virus variants in patients exhibiting increasing CD4+ T cell counts despite virologic failure of protease inhibitor-containing antiretroviral combination therapy. J Infect Dis **181:**1808-1812.

20. **Geretti AM, Easterbrook P.** 2001. Antiretroviral resistance in clinical practice. Int J STD AIDS **12:**145-153.

21. **Nishizawa M, Matsuda M, Hattori J, Shiino T, Matano T, Heneine W, Johnson JA, Sugiura W.** 2015. Longitudinal Detection and Persistence of Minority Drug-Resistant Populations and Their Effect on Salvage Therapy. PLoS One **10:**e0135941.

22. **Broder CC, Collman RG.** 1997. Chemokine receptors and HIV. J Leukoc Biol **62:**20-29.

23. **Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, Moore LR, Schmidt SD, Stuckey J, Yang Y, Zhou T, Zhu J, Vicic DA, Debnath AK, Shapiro L, Bewley CA, Mascola JR, Sodroski JG, Kwong PD.** 2012. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc Natl Acad Sci U S A **109:**5663-5668.

24. **Abrahamyan LG, Mkrtchyan SR, Binley J, Lu M, Melikyan GB, Cohen FS.** 2005. The cytoplasmic tail slows the folding of human immunodeficiency virus type 1 Env from a late prebundle configuration into the six-helix bundle. J Virol **79:**106-115.

25. **Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K.** 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. RNA Biol **7:**754-774.

26. **Levin JG, Guo J, Rouzina I, Musier-Forsyth K.** 2005. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. Prog Nucleic Acid Res Mol Biol **80:**217-286.

27. **Kuznetsov YG, Victoria JG, Robinson WE, Jr., McPherson A.** 2003. Atomic force microscopy investigation of human immunodeficiency virus (HIV) and HIV-infected lymphocytes. J Virol **77:**11896-11909.

28. **Connell BJ, Lortat-Jacob H.** 2013. Human immunodeficiency virus and heparan sulfate: from attachment to entry inhibition. Front Immunol **4:**385.

29. **Dawson L, Yu XF.** 1998. The role of nucleocapsid of HIV-1 in virus assembly. Virology **251:**141-157.

30. **Shaw GM, Hunter E.** 2012. HIV Transmission. Cold Spring Harbor Perspectives in Medicine **2**.

31. **Doitsh G, Galloway NL, Geng X, Yang Z, Monroe KM, Zepeda O, Hunt PW, Hatano H, Sowinski S, Munoz-Arias I, Greene WC.** 2014. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. Nature **505:**509-514.

32. **Swanstrom R, Coffin J.** 2012. HIV-1 Pathogenesis: The Virus. Cold Spring Harbor Perspectives in Medicine **2**.

33. **Lackner AA, Lederman MM, Rodriguez B.** 2012. HIV Pathogenesis: The Host. Cold Spring Harbor Perspectives in Medicine **2**.

34. **Henrich TJ, Kuritzkes DR.** 2013. HIV-1 entry inhibitors: recent development and clinical use. Curr Opin Virol **3:**51-57.

35. **Huang L, Zhang L, Chen CH.** 2003. Potential drug targets on the HIV-1 envelope glycoproteins, gp120 and gp41. Curr Pharm Des **9:**1453-1462.

36. **Meanwell NA, Kadow JF.** 2003. Inhibitors of the entry of HIV into host cells. Curr Opin Drug Discov Devel **6:**451-461.

37. **Wang T, Zhang Z, Wallace OB, Deshpande M, Fang H, Yang Z, Zadjura LM, Tweedie DL, Huang S, Zhao F, Ranadive S, Robinson BS, Gong YF, Ricarrdi K, Spicer TP, Deminie C, Rose R, Wang HG, Blair WS, Shi PY, Lin PF, Colonno RJ, Meanwell NA.** 2003. Discovery of 4-benzoyl-1-[(4-methoxy-1H- pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2- (R)-methylpiperazine (BMS-378806): a novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. J Med Chem **46:**4236-4239.

38. **Altmeyer R.** 2004. Virus attachment and entry offer numerous targets for antiviral therapy. Curr Pharm Des **10:**3701-3712.

39. **Seibert C, Sakmar TP.** 2004. Small-molecule antagonists of CCR5 and CXCR4: a promising new class of anti-HIV-1 drugs. Curr Pharm Des **10:**2041-2062.

40. **Teissier E, Penin F, Pecheur EI.** 2011. Targeting cell entry of enveloped viruses as an antiviral strategy. Molecules **16:**221-250.

41. **Ryu SE, Kwong PD, Truneh A, Porter TG, Arthos J, Rosenberg M, Dai XP, Xuong NH, Axel R, Sweet RW, et al.** 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature **348:**419-426.

42. **Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393:**648-659.

43. **Xiang SH, Kwong PD, Gupta R, Rizzuto CD, Casper DJ, Wyatt R, Wang L, Hendrickson WA, Doyle ML, Sodroski J.** 2002. Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 gp120 envelope glycoprotein. J Virol **76:**9888-9899.

44. **Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, Kwong PD.** 2005. Structure of a V3-containing HIV-1 gp120 core. Science **310:**1025-1028.

45. **Collman RG, Yi Y, Liu QH, Freedman BD.** 2000. Chemokine signaling and HIV-1 fusion mediated by macrophage CXCR4: implications for target cell tropism. J Leukoc Biol **68:**318-323.

46. **Sambyal V, Manjari M, Sudan M, Uppal MS, Singh NR, Singh H, Guleria K.** 2015. No Association between the CCR5Delta32 Polymorphism and Sporadic Esophageal Cancer in Punjab, North-West India. Asian Pac J Cancer Prev **16:**4291-4295.

47. **O'Connell O, Repik A, Reeves JD, Gonzalez-Perez MP, Quitadamo B, Anton ED, Duenas-Decamp M, Peters P, Lin R, Zolla-Pazner S, Corti D, Wallace A, Wang S, Kong XP, Lu S, Clapham PR.** 2013. Efficiency of bridging-sheet recruitment explains HIV-1 R5 envelope glycoprotein sensitivity to soluble CD4 and macrophage tropism. J Virol **87:**187-198.

48. **Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B.** 2013. Structure of the CCR5 chemokine receptor – HIV entry inhibitor Maraviroc complex. Science (New York, NY) **341:**10.1126/science.1241475.

49. **Melikyan GB, Markosyan RM, Hemmati H, Delmedico MK, Lambert DM, Cohen FS.** 2000. Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. J Cell Biol **151:**413-423.

50. **Skehel JJ, Wiley DC.** 1998. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell **95:**871-874.

51. **Roche J, Louis JM, Grishaev A, Ying J, Bax A.** 2014. Dissociation of the trimeric gp41 ectodomain at the lipid-water interface suggests an active role in HIV-1 Env-mediated membrane fusion. Proc Natl Acad Sci U S A **111:**3425-3430.

52. **Cai L, Gochin M, Liu K.** 2011. Biochemistry and biophysics of HIV-1 gp41 - membrane interactions and implications for HIV-1 envelope protein mediated viral-cell fusion and fusion inhibitor design. Curr Top Med Chem **11:**2959-2984.

53. **Doms RW, Moore JP.** 2000. HIV-1 membrane fusion: targets of opportunity. J Cell Biol **151:**F9-14.

54. **Raja A, Venturi M, Kwong P, Sodroski J.** 2003. CD4 binding site antibodies inhibit human immunodeficiency virus gp120 envelope glycoprotein interaction with CCR5. J Virol **77:**713-718.

55. **Wibmer CK, Moore PL, Morris L.** 2015. HIV broadly neutralizing antibody targets. Curr Opin HIV AIDS **10:**135-143.

56. **Scharf L, Wang H, Gao H, Chen S, McDowall AW, Bjorkman PJ.** 2015. Broadly Neutralizing Antibody 8ANC195 Recognizes Closed and Open States of HIV-1 Env. Cell **162:**1379-1390.

57. **Ahlers JD.** 2014. All eyes on the next generation of HIV vaccines: strategies for inducing a broadly neutralizing antibody response. Discov Med **17:**187-199.

58. **Julien JP, Sok D, Khayat R, Lee JH, Doores KJ, Walker LM, Ramos A, Diwanji DC, Pejchal R, Cupo A, Katpally U, Depetris RS, Stanfield RL, McBride R, Marozsan AJ, Paulson JC, Sanders RW, Moore JP, Burton DR, Poignard P, Ward AB, Wilson IA.** 2013. Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. PLoS Pathog **9:**e1003342.

59. **Julien JP, Lee JH, Cupo A, Murin CD, Derking R, Hoffenberg S, Caulfield MJ, King CR, Marozsan AJ, Klasse PJ, Sanders RW, Moore JP, Wilson IA, Ward AB.** 2013. Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. Proc Natl Acad Sci U S A **110:**4351-4356.

60. **Muhle M, Hoffmann K, Lochelt M, Denner J.** 2013. Construction and characterisation of replicating foamy viral vectors expressing HIV-1 epitopes recognised by broadly neutralising antibodies. Antiviral Res **100:**314-320.

61. **Freeman MM, Seaman MS, Rits-Volloch S, Hong X, Kao CY, Ho DD, Chen B.** 2010. Crystal structure of HIV-1 primary receptor CD4 in complex with a potent antiviral antibody. Structure **18:**1632-1641.

62. **Fessel WJ, Anderson B, Follansbee SE, Winters MA, Lewis ST, Weinheimer SP, Petropoulos CJ, Shafer RW.** 2011. The efficacy of an anti-CD4 monoclonal antibody for HIV-1 treatment. Antiviral Res **92:**484-487.

63. **Pace CS, Fordyce MW, Franco D, Kao CY, Seaman MS, Ho DD.** 2013. Anti-CD4 monoclonal antibody ibalizumab exhibits breadth and potency against HIV-1, with natural resistance mediated by the loss of a V5 glycan in envelope. J Acquir Immune Defic Syndr **62:**1-9.

64. **Kilby JM, Hopkins S, Venetta TM, DiMassimo B, Cloud GA, Lee JY, Alldredge L, Hunter E, Lambert D, Bolognesi D, Matthews T, Johnson MR, Nowak MA, Shaw**

GM, Saag MS. 1998. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. Nat Med **4:**1302-1307.

65. **Markosyan RM, Ma X, Lu M, Cohen FS, Melikyan GB.** 2002. The mechanism of inhibition of HIV-1 env-mediated cell-cell fusion by recombinant cores of gp41 ectodomain. Virology **302:**174-184.

66. **Lawless MK, Barney S, Guthrie KI, Bucy TB, Petteway SR, Jr., Merutka G.** 1996. HIV-1 membrane fusion mechanism: structural studies of the interactions between biologically-active peptides from gp41. Biochemistry **35:**13697-13708.

67. **Imai M, Okada N, Okada H.** 2000. Inhibition of HIV-1 infection by an intramolecular antisense peptide to T20 in gp160. Microbiol Immunol **44:**205-212.

68. **Cohen CJ, Dusek A, Green J, Johns EL, Nelson E, Recny MA.** 2002. Long-term treatment with subcutaneous T-20, a fusion inhibitor, in HIV-infected patients: patient satisfaction and impact on activities of daily living. AIDS Patient Care STDS **16:**327-335.

69. **Shi W, Qi Z, Pan C, Xue N, Debnath AK, Qie J, Jiang S, Liu K.** 2008. Novel anti-HIV peptides containing multiple copies of artificially designed heptad repeat motifs. Biochem Biophys Res Commun **374:**767-772.

70. **Eggink D, Berkhout B, Sanders RW.** 2010. Inhibition of HIV-1 by fusion inhibitors. Curr Pharm Des **16:**3716-3728.

71. **Reeves JD, Piefer AJ.** 2005. Emerging drug targets for antiretroviral therapy. Drugs **65:**1747-1766.

72. **Chelli M, Alizon M.** 2001. Determinants of the trans-dominant negative effect of truncated forms of the CCR5 chemokine receptor. J Biol Chem **276:**46975-46982.

73. **Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, Holmes MC, Gregory PD, Ando DG, Kalos M, Collman RG, Binder-Scholl G, Plesa G, Hwang WT, Levine BL, June CH.** 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N Engl J Med **370:**901-910.

74. **Hoxie JA, June CH.** 2012. Novel cell and gene therapies for HIV. Cold Spring Harb Perspect Med **2**.

75. **Nazari R, Joshi S.** 2008. CCR5 as target for HIV-1 gene therapy. Curr Gene Ther **8:**264-272.

76. **Carrington M, Dean M, Martin MP, O'Brien SJ.** 1999. Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Hum Mol Genet **8:**1939-1945.

77. **Li L, Krymskaya L, Wang J, Henley J, Rao A, Cao L-F, Tran C-A, Torres-Coronado M, Gardner A, Gonzalez N.** 2013. Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. Molecular Therapy **21:**1259-1269.

78. **Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC.** 2010. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nature biotechnology **28:**839-847.

79. **Vandekerckhove L, Verhofstede C, Vogelaers D.** 2009. Maraviroc: perspectives for use in antiretroviral-naive HIV-1-infected patients. J Antimicrob Chemother **63:**1087-1096.

80. **Genebat M, Ruiz-Mateos E, Gonzalez-Serna A, Pulido I, Munoz-Fernandez MA, Ferrando-Martinez S, Leal M.** 2011. Discordance rates between Trofile test and short-term virological response to maraviroc. Antiviral Res **89:**182-185.

81. **Wilkin TJ, Goetz MB, Leduc R, Skowron G, Su Z, Chan ES, Heera J, Chapman D, Spritzler J, Reeves JD, Gulick RM, Coakley E.** 2011. Reanalysis of coreceptor tropism in HIV-1-infected adults using a phenotypic assay with enhanced sensitivity. Clin Infect Dis **52:**925-928.

82. **Woollard SM, Kanmogne GD.** 2015. Maraviroc: a review of its use in HIV infection and beyond. Drug Des Devel Ther **9:**5447-5468.

83. **Panos G, Watson DC.** 2014. Effect of HIV-1 subtype and tropism on treatment with chemokine coreceptor entry inhibitors; overview of viral entry inhibition. Crit Rev Microbiol doi:10.3109/1040841X.2013.867829.

84. **Grande F, Garofalo A, Neamati N.** 2008. Small molecules anti-HIV therapeutics targeting CXCR4. Curr Pharm Des **14:**385-404.

85. **Das D, Maeda K, Hayashi Y, Gavande N, Desai DV, Chang SB, Ghosh AK, Mitsuya H.** 2015. Insights into the mechanism of inhibition of CXCR4: identification of Piperidinylethanamine analogs as anti-HIV-1 inhibitors. Antimicrob Agents Chemother **59:**1895-1904.

86. **Tilton JC, Doms RW.** 2010. Entry inhibitors in the treatment of HIV-1 infection. Antiviral Research **85:**91-100.

87. **Lobritz MA, Ratcliff AN, Arts EJ.** 2010. HIV-1 Entry, Inhibitors, and Resistance. Viruses **2:**1069-1105.

88. **Melby T, Westby M.** 2009. Inhibitors of viral entry. Handb Exp Pharmacol doi:10.1007/978-3-540-79086-0_7**:**177-202.

89. **Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC.** 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature **387:**426-430.

90. **Do Kwon Y, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gorman J, Joyce MG, Guttman M, Ma X, Narpala S, Soto C, Terry DS, Yang Y, Zhou T, Ahlsen G, Bailer RT, Chambers M, Chuang GY, Doria-Rose NA, Druz A, Hallen MA, Harned A, Kirys T, Louder MK, O'Dell S, Ofek G, Osawa K, Prabhakaran M, Sastry M, Stewart-Jones GB, Stuckey J, Thomas PV, Tittley T, Williams C, Zhang B, Zhao H, Zhou Z, Donald BR, Lee LK, Zolla-Pazner S, Baxa U, Schon A, Freire E, Shapiro L, Lee KK, Arthos J, Munro JB, Blanchard SC, Mothes W, Binley JM, et al.** 2015. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. Nat Struct Mol Biol **22:**522-531.

91. **Zhu P, Liu J, Bess J, Jr., Chertova E, Lifson JD, Grise H, Ofek GA, Taylor KA, Roux KH.** 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature **441:**847-852.

92. **Klein JS, Bjorkman PJ.** 2010. Few and far between: how HIV may be evading antibody avidity. PLoS Pathog **6:**e1000908.

93. **Wang W, Nie J, Prochnow C, Truong C, Jia Z, Wang S, Chen XS, Wang Y.** 2013. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. Retrovirology **10:**1-14.

94. **Kong L, Torrents de la Pena A, Deller MC, Garces F, Sliepen K, Hua Y, Stanfield RL, Sanders RW, Wilson IA.** 2015. Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. Acta Crystallogr D Biol Crystallogr **71:**2099-2108.

95. **Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, Stanfield RL, Julien JP, Ramos A, Crispin M, Depetris R, Katpally U, Marozsan A, Cupo A, Maloveste S, Liu Y, McBride R, Ito Y, Sanders RW, Ogohara C, Paulson JC, Feizi T, Scanlan CN, Wong CH, Moore JP, Olson WC, Ward AB, Poignard P, Schief WR, Burton DR, Wilson IA.** 2011. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science **334:**1097-1103.

96. **Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B.** 2013. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science **341:**1387-1390.

97. **Acharya P, Lusvarghi S, Bewley CA, Kwong PD.** 2015. HIV-1 gp120 as a therapeutic target: navigating a moving labyrinth. Expert Opin Ther Targets **19:**765-783.

98. **Yao N, Zhang C, Liu Q, Liu J, Zhang C.** 2015. Polymorphism characteristics of HIV-1 gpl20 and 5 hypervariable regions. Turk J Med Sci **45:**47-54.

99. **Tilton JC, Doms RW.** 2010. Entry inhibitors in the treatment of HIV-1 infection. Antiviral Res **85:**91-100.

100. **Gu R, Westervelt P, Ratner L.** 1993. Role of HIV-1 envelope V3 loop cleavage in cell tropism. AIDS Res Hum Retroviruses **9:**1007-1015.

101. **Westervelt P, Trowbridge DB, Epstein LG, Blumberg BM, Li Y, Hahn BH, Shaw GM, Price RW, Ratner L.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J Virol **66:**2577-2582.

102. **Steidl S, Stitz J, Schmitt I, König R, Flory E, Schweizer M, Cichutek K.** 2002. Coreceptor Switch of [MLV (SIVagm)] pseudotype vectors by V3-loop exchange. Virology **300:**205-216.

103. **Hung CS, Vander Heyden N, Ratner L.** 1999. Analysis of the critical domain in the V3 loop of human immunodeficiency virus type 1 gp120 involved in CCR5 utilization. J Virol **73:**8216-8226.

104. **Henkel T, Westervelt P, Ratner L.** 1995. HIV-1 V3 envelope sequences required for macrophage infection. AIDS **9:**399-401.

105. **Carrillo A, Trowbridge DB, Westervelt P, Ratner L.** 1993. Identification of HIV1 determinants for T lymphoid cell line infection. Virology **197:**817-824.

106. **Ebenbichler C, Westervelt P, Carrillo A, Henkel T, Johnson D, Ratner L.** 1993. Structure-function relationships of the HIV-1 envelope V3 loop tropism determinant: evidence for two distinct conformations. AIDS **7:**639-646.

107. **Svicher V, Alteri C, Artese A, Zhang JM, Costa G, Mercurio F, D'Arrigo R, Alcaro S, Palu G, Clementi M, Zazzi M, Andreoni M, Antinori A, Lazzarin A, Ceccherini-Silberstein F, Perno CF, group Os.** 2011. Identification and structural characterization of novel genetic elements in the HIV-1 V3 loop regulating coreceptor usage. Antivir Ther **16:**1035-1045.

108. **Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, Mosier DE.** 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol **80:**750-758.

109. **Carrillo A, Ratner L.** 1996. Cooperative effects of the human immunodeficiency virus type 1 envelope variable loops V1 and V3 in mediating infectivity for T cells. J Virol **70:**1310-1316.

110. **Monno L, Saracino A, Scudeller L, Punzi G, Brindicci G, Altamura M, Lagioia A, Ladisa N, Angarano G.** 2011. Impact of mutations outside the V3 region on coreceptor tropism phenotypically assessed in patients infected with HIV-1 subtype B. Antimicrob Agents Chemother **55:**5078-5084.

111. **Zhang L, Ma L, Wang Z, Wang Y, Zhang J, Wang H, Shao Y.** 2014. Alterations in HIV-1 gp120 V3 region are necessary but not sufficient for coreceptor switching in CRF07_BC in China. PLoS One **9:**e93426.

112. **Pan R, Gorny MK, Zolla-Pazner S, Kong XP.** 2015. The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel. J Virol **89:**8003-8010.

113. **Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, Kappes JC, Shaw GM, Hunter E.** 2000. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. J Virol **74:**8358-8367.

114. **Derdeyn CA, Decker JM, Sfakianos JN, Zhang Z, O'Brien WA, Ratner L, Shaw GM, Hunter E.** 2001. Sensitivity of human immunodeficiency virus type 1 to fusion inhibitors targeted to the gp41 first heptad repeat involves distinct regions of gp41 and is consistently modulated by gp120 interactions with the coreceptor. J Virol **75:**8605-8614.

115. **Dobrowsky TM, Rabi SA, Nedellec R, Daniels BR, Mullins JI, Mosier DE, Siliciano RF, Wirtz D.** 2013. Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins. Sci Rep **3:**3014.

116. **Zarr M, Siliciano R.** 2015. Stoichiometric parameters of HIV-1 entry. Virology **474:**1-9.

117. **Pancera M, Majeed S, Ban YE, Chen L, Huang CC, Kong L, Kwon YD, Stuckey J, Zhou T, Robinson JE, Schief WR, Sodroski J, Wyatt R, Kwong PD.** 2010. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. Proc Natl Acad Sci U S A **107:**1166-1171.

118. **Sguanci L, Bagnoli F, Lio P.** 2007. Modeling HIV quasispecies evolutionary dynamics. BMC Evol Biol **7 Suppl 2:**S5.

119. **Perelson AS.** 2002. Modelling viral and immune system dynamics. Nat Rev Immunol **2:**28-36.

120. **Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science **271:**1582-1586.

121. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol **69:**5087-5094.

122. **Regoes RR, Hamblin S, Tanaka MM.** 2013. Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. Proc Biol Sci **280:**20122047.

123. **Kamp C, Bornholdt S.** 2002. Coevolution of quasispecies: B-cell mutation rates maximize viral error catastrophes. Phys Rev Lett **88:**068104.

124. **McNearney T, Hornickova Z, Markham R, Birdwell A, Arens M, Saah A, Ratner L.** 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc Natl Acad Sci U S A **89:**10247-10251.

125. **Nankya IL, Tebit DM, Abraha A, Kyeyune F, Gibson R, Jegede O, Nickel G, Arts EJ.** 2015. Defining the fitness of HIV-1 isolates with dual/mixed co-receptor usage. AIDS Res Ther **12:**34.

126. **Coakley E, Reeves JD, Huang W, Mangas-Ruiz M, Maurer I, Harskamp AM, Gupta S, Lie Y, Petropoulos CJ, Schuitemaker H, van 't Wout AB.** 2009. Comparison of human immunodeficiency virus type 1 tropism profiles in clinical samples by the Trofile and MT-2 assays. Antimicrob Agents Chemother **53:**4686-4693.

127. **Su Z, Gulick RM, Krambrink A, Coakley E, Hughes MD, Han D, Flexner C, Wilkin TJ, Skolnik PR, Greaves WL, Kuritzkes DR, Reeves JD, Team ACTGA.** 2009. Response to vicriviroc in treatment-experienced subjects, as determined by an enhanced-sensitivity coreceptor tropism assay: reanalysis of AIDS clinical trials group A5211. J Infect Dis **200:**1724-1728.

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

129. **Schloss JA.** 2008. How to get genomes at one ten-thousandth the cost. Nat Biotechnol **26:**1113-1115.

130. **Portsmouth S, Valluri SR, Daumer M, Thiele B, Valdez H, Lewis M, Craig C, Thielen A, James I, Demarest J, Heera J.** 2013. Correlation between genotypic (V3 population sequencing) and phenotypic (Trofile ES) methods of characterizing co-receptor usage of HIV-1 from 200 treatment-naive HIV patients screened for Study A4001078. Antiviral Res **97:**60-65.

131. **Pou C, Codoner FM, Thielen A, Bellido R, Perez-Alvarez S, Cabrera C, Dalmau J, Curriu M, Lie Y, Noguera-Julian M, Puig J, Martinez-Picado J, Blanco J, Coakley E, Daumer M, Clotet B, Paredes R.** 2013. HIV-1 tropism testing in subjects achieving undetectable HIV-1 RNA: diagnostic accuracy, viral evolution and compartmentalization. PLoS One **8:**e67085.

132. **To SW, Chen JH, Wong KH, Chan KC, Chen Z, Yam WC.** 2013. Determination of the high prevalence of Dual/Mixed- or X4-tropism among HIV type 1 CRF01_AE in Hong Kong by genotyping and phenotyping methods. AIDS Res Hum Retroviruses **29:**1123-1128.

133. **Kagan RM, Johnson EP, Siaw MF, Van Baelen B, Ogden R, Platt JL, Pesano RL, Lefebvre E.** 2014. Comparison of genotypic and phenotypic HIV type 1 tropism assay:

results from the screening samples of Cenicriviroc Study 202, a randomized phase II trial in treatment-naive subjects. AIDS Res Hum Retroviruses **30:**151-159.

134. **Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, Paxinos E, Arts EJ, Robertson DL, Mimms L, Quinones-Mateu ME.** 2012. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One **7:**e49602.

135. **Swenson LC, Daumer M, Paredes R.** 2012. Next-generation sequencing to assess HIV tropism. Curr Opin HIV AIDS **7:**478-485.

136. **Cashin K, Gray LR, Harvey KL, Perez-Bercoff D, Lee GQ, Sterjovski J, Roche M, Demarest JF, Drummond F, Harrigan PR, Churchill MJ, Gorry PR.** 2015. Reliable genotypic tropism tests for the major HIV-1 subtypes. Sci Rep **5:**8543.

137. **van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C.** 2014. Ten years of next-generation sequencing technology. Trends Genet **30:**418-426.

138. **Zhao Q, Shi L, Jiang Y, Wen Y, Pan P, Zhang G, Qiu M.** 2014. [Contact tracing of a possible case of HIV sexual transmission by using Miseq platform]. Zhonghua Yu Fang Yi Xue Za Zhi **48:**471-475.

139. **Rozera G, Abbate I, Ciccozzi M, Lo Presti A, Bruselles A, Vlassi C, D'Offizi G, Narciso P, Giombini E, Bartolini B, Ippolito G, Capobianchi MR.** 2012. Ultra-deep sequencing reveals hidden HIV-1 minority lineages and shifts of viral population between the main cellular reservoirs of the infection after therapy interruption. J Med Virol **84:**839-844.

140. **Bunnik EM, Swenson LC, Edo-Matas D, Huang W, Dong W, Frantzell A, Petropoulos CJ, Coakley E, Schuitemaker H, Harrigan PR, van 't Wout AB.** 2011. Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing. PLoS Pathog **7:**e1002106.

141. **Vandenbroucke I, Van Marck H, Mostmans W, Van Eygen V, Rondelez E, Thys K, Van Baelen K, Fransen K, Vaira D, Kabeya K, De Wit S, Florence E, Moutschen M, Vandekerckhove L, Verhofstede C, Stuyver LJ.** 2010. HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. AIDS Res Ther **7:**4.

142. **Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR.** 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One **4:**e5683.

143. **Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC.** 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330:**1066-1071.

144. **Fichot EB, Norman RS.** 2013. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. Microbiome **1:**10.

145. **Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, Lawrenson K, McGuffog L, Healey S, Lee JM, Spindler TJ, Lin YG, Pejovic T, Bean Y, Li Q, Coetzee S, Hazelett D, Miron A, Southey M, Terry MB, Goldgar DE, Buys SS, Janavicius R, Dorfling CM, van Rensburg EJ, Neuhausen SL, Ding YC, Hansen TV, Jonson L, Gerdes AM, Ejlertsen B, Barrowdale D, Dennis J, Benitez J, Osorio A, Garcia MJ, Komenaka I, Weitzel JN, Ganschow P, Peterlongo P, Bernard L, Viel A, Bonanni B, Peissel B, Manoukian S, Radice P, Papi L, Ottini L, Fostira F, Konstantopoulou I, et al.** 2015. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat Genet **47:**164-171.

146. **Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR.** 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. J Infect Dis **203:**237-245.

147. **McGovern RA, Thielen A, Portsmouth S, Mo T, Dong W, Woods CK, Zhong X, Brumme CJ, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR.** 2012. Population-based sequencing of the V3-loop can predict the virological response to maraviroc in treatment-naive patients of the MERIT trial. J Acquir Immune Defic Syndr **61:**279-286.

148. **Prosperi MC, Bracciale L, Fabbiani M, Di Giambenedetto S, Razzolini F, Meini G, Colafigli M, Marzocchetti A, Cauda R, Zazzi M, De Luca A.** 2010. Comparative determination of HIV-1 co-receptor tropism by Enhanced Sensitivity Trofile, gp120 V3-loop RNA and DNA genotyping. Retrovirology **7:**56.

149. **Svicher V, D'Arrigo R, Alteri C, Andreoni M, Angarano G, Antinori A, Antonelli G, Bagnarelli P, Baldanti F, Bertoli A, Borderi M, Boeri E, Bonn I, Bruzzone B, Callegaro AP, Cammarota R, Canducci F, Ceccherini-Silberstein F, Clementi M, Monforte AD, De Luca A, Di Biagio A, Di Gianbenedetto S, Di Perri G, Di Pietro M, Fabeni L, Fadda G, Galli M, Gennari W, Ghisetti V, Giacometti A, Gori A, Leoncini F, Maggiolo F, Maserati R, Mazzotta F, Micheli V, Meini G, Monno L, Mussini C, Nozza S, Paolucci S, Parisi S, Pecorari M, Pizzi D, Quirino T, Re MC, Rizzardini G, Santangelo R, Soria A, et al.** 2010. Performance of genotypic tropism testing in clinical practice using the enhanced sensitivity version of Trofile as reference assay: results from the OSCAR Study Group. New Microbiol **33:**195-206.

150. **Saliou A, Delobel P, Dubois M, Nicot F, Raymond S, Calvez V, Masquelier B, Izopet J, Group AARS.** 2011. Concordance between two phenotypic assays and ultradeep pyrosequencing for determining HIV-1 tropism. Antimicrob Agents Chemother **55:**2831-2836.

151. **Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, Ruiz-Mateos E, Leal M, Robertson DL, Schmotzer CL, Quinones-Mateu ME.** 2014. Sensitive deep-

sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. Antimicrob Agents Chemother **58:**2167-2185.

152. **Pandit A, de Boer RJ.** 2014. Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. Retrovirology **11:**56.

153. **Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R.** 2007. Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol **25:**1407-1410.

154. **Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI.** 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol **77:**13376-13388.

155. **Brumme ZL, Dong WW, Yip B, Wynhoven B, Hoffman NG, Swanstrom R, Jensen MA, Mullins JI, Hogg RS, Montaner JS, Harrigan PR.** 2004. Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. AIDS **18:**F1-9.

156. **Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI.** 2006. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. J Virol **80:**4698-4704.

157. **Montagna C, De Crignis E, Bon I, Re MC, Mezzaroma I, Turriziani O, Graziosi C, Antonelli G.** 2014. V3 net charge: additional tool in HIV-1 tropism prediction. AIDS Res Hum Retroviruses **30:**1203-1212.

158. **Pillai S, Good B, Richman D, Corbeil J.** 2003. A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses **19:**145-149.

159. **Thielen A, Lengauer T.** 2012. Geno2pheno[454]: a Web server for the prediction of HIV-1 coreceptor usage from next-generation sequencing data. Intervirology **55:**113-117.

160. **Garrido C, Roulet V, Chueca N, Poveda E, Aguilera A, Skrabal K, Zahonero N, Carlos S, Garcia F, Faudon JL, Soriano V, de Mendoza C.** 2008. Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. J Clin Microbiol **46:**887-891.

161. **Cashin K, Gray LR, Jakobsen MR, Sterjovski J, Churchill MJ, Gorry PR.** 2013. CoRSeqV3-C: a novel HIV-1 subtype C specific V3 sequence based coreceptor usage prediction algorithm. Retrovirology **10:**24.

162. **Rupp D, Geuenich S, Keppler OT.** 2010. Poor performance of bioinformatics programs for genotypic prediction of coreceptor usage of HIV-1 group O isolates. J Acquir Immune Defic Syndr **53:**412-413.

163. **Vandekerckhove LP, Wensing AM, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A, Dressler S, Garcia F, Geretti AM, Klimkait T, Korn K, Masquelier B, Perno CF, Schapiro JM, Soriano V, Sonnerborg A, Vandamme AM, Verhofstede C, Walter H, Zazzi M, Boucher CA, European Consensus Group on clinical management of tropism t.** 2011. European guidelines on the clinical management of HIV-1 tropism testing. Lancet Infect Dis **11:**394-407.

164. **Svicher V, Cento V, Rozera G, Abbate I, Santoro MM, Armenia D, Fabeni L, Bruselles A, Latini A, Palamara G, Micheli V, Rizzardini G, Gori C, Forbici F, Ippolito G, Andreoni M, Antinori A, Ceccherini-Silberstein F, Capobianchi MR, Perno CF.** 2013. The genotypic false positive rate determined by V3 population sequencing can predict the burden of HIV-1 CXCR4-using species detected by pyrosequencing. PLoS One **8:**e53603.

165. **Rodriguez C, Soulie C, Marcelin AG, Calvez V, Descamps D, Charpentier C, Flandre P, Recordon-Pinson P, Bellecave P, Pawlotsky JM, Masquelier B, Group AAS.** 2015. HIV-1 Coreceptor Usage Assessment by Ultra-Deep Pyrosequencing and Response to Maraviroc. PLoS One **10:**e0127816.

166. **Pfeifer N, Lengauer T.** 2012. Improving HIV coreceptor usage prediction in the clinic using hints from next-generation sequencing data. Bioinformatics **28:**i589-i595.

167. **Soulie C, Descamps D, Grude M, Schneider V, Trabaud MA, Morand-Joubert L, Delaugerre C, Montes B, Barin F, Ferre V, Raymond S, Jeulin H, Alloui C, Yerly S, Pallier C, Reigadas S, Signori-Schmuck A, Guigon A, Fafi-Kremer S, Haim-Boukobza S, Mirand A, Maillard A, Vallet S, Roussel C, Assoumou L, Calvez V, Flandre P, Marcelin AG, Group ARA.** 2015. Antiretroviral-naive and -treated HIV-1 patients can harbour more resistant viruses in CSF than in plasma. J Antimicrob Chemother **70:**566-572.

Chapter 2

Single Clone Analysis of HIV-1 Envelope Quasispecies

Reveals Dynamic Intra-Subject Evolution

with CCR5 Antagonist Therapy

2.1 Preface

Chapter 2 is adapted from a manuscript prepared for submission. The author contributions are the following: Lee Ratner and Jie Zhang conceived all the experiments; Jie Zhang independently performed all the experiments, analyzed all the data, and prepared all the tables and figures except for Figure 2.11.4(C) and Figure 2.11.7, which are prepared with help from Xiang Gao; Timothy Henrich and Daniel Kuritzkes kindly provided the clinical samples used in this study. The manuscript was written and revised by primarily by Jie Zhang and Lee Ratner.

Analysis of HIV-1 Envelope Quasispecies Reveals Dynamic Intra-Subject Evolution with CCR5 Antagonist Therapy

Jie Zhang[a], Xiang Gao[a], Timothy Henrich[b], Daniel Kuritzkes[b], Lee Ratner[a]#

Division of Molecular Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA[a]; Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, MA, USA[b]

Corresponding author: Lee Ratner

Department of Medicine, Washington University School of Medicine, 660 S. Euclid, Campus Box 8069, St. Louis, MO 63110, USA.

lratner@dom.wustl.edu

Running head: Single clone analysis of HIV-1 variants in coreceptor shift

2.2 Abstract

HIV-1 envelope interacts with coreceptors CCR5 and/or CXCR4 before gaining entry into the host cells through a dynamic, multi-step process. The tropism determinants on the envelope protein are not clearly defined despite recent advances in structural understanding of the coreceptors as well as the HIV-1 native envelope trimer spike. CCR5 inhibitor treatment in patients harboring CCR5-tropic virus frequently leads to coreceptor shift and the emergence of CXCR4-using virus. In order to understand the evolutionary pathways of coreceptor shift, we examined a panel of single clones from patients before and after they failed CCR5 antagonist therapy, using a combination of tissue-culture based phenotypic tropism assays and sequencing-based genotypic analysis. We found distinct quasispecies compositions in two study subjects after the treatment failure. In one of the two subjects, we observed a rapid, complete shift of tropism and large variations in the envelope variable regions within a two-week period of drug selection. This study reveals the rapid nature of HIV-1 envelope evolution *in vivo,* and confirms key structure-function relationships in envelope coreceptor specificity.

2.3 Importance

Preventing resistance development to current and future antiretroviral therapeutic agents is a top priority for global public health against HIV/AIDS. Utilizing a combination of phenotypic and genotypic assays on 97 single clones isolated from 2 subjects, we confirmed the previously reported sequence signature for CXCR4 tropism, and cross-validated the structure-function relationship of existing tropism prediction algorithms. In addition, we demonstrated the rapid and complete nature of coreceptor shift and treatment escape in one patient, and a slow and incomplete coreceptor shift in the other patient, under the specific selection pressure of a CCR5 antagonist. This finding suggests different evolutionary pathways about coreceptor shift, and has implications in clinical administration of coreceptor inhibitors and the need to closely monitor quasispecies tropism under such therapy.

2.4 Introduction

C-C chemokine receptor type 5 (CCR5) and C-X-C chemokine receptor type 4 (CXCR4) are host cell targets for small molecule inhibitors against HIV-1 infection (1-3). After engaging the primary receptor CD4, HIV virions can utilize either CCR5 (R5), CXCR4 (X4), or both (dual or R5X4) coreceptors to enter a susceptible cell (4, 5). Blocking virion-coreceptor interaction stops an infection at the first step of viral replication cycle. Currently, Maraviroc (MRV) is the first and only FDA approved small molecule CCR5 antagonist against HIV-1 infection, though a number of other small molecules have been developed and tested as coreceptor inhibitors to both CCR5 and CXCR4 (6-12). Eligibility for clinical application of MRV requires that a patient harbors undetectable X4-tropic virus at the commencement of treatment, typically determined by a tissue-culture based phenotypic method called Enhanced Sensitivity Trofile Assay (ESTA) (13). Despite the stringent admission criteria, resistance to MRV has been shown to develop *in vivo* through multiple potential mechanisms, such as coreceptor shift from R5- to X4-tropism, and binding to different CCR5 confirmations including the coreceptor-drug complex (14-18). Resistance development to CCR5 antagonists increases the risk of treatment failure and hampers the utilization of current and future coreceptor antagonist therapeutic agents (19).

One important factor in the development of resistance to antiretroviral agents lies in the fact that HIV-1 exists in the host as quasispecies, or a group of genetically related but non-identical variants (20). An infection typically starts with a few if not a single transmission/founder virus crossing the mucosal barrier, mostly likely a R5 isolate, and expansion to more than 10^6 copies per milliliter of blood a few weeks after infection. During this rapid course of replication, the virus accumulates mutations as a result of its error-prone RNA polymerase and high turnover rate, and evolves into a community of quasispecies that is

adaptable to host and external selection pressures (21). With the memory T lymphocytes being depleted over the course of the infection, fifty percent of HIV-1 Clade B patients spontaneously develop X4-tropic virus late in the disease course to more efficiently infect naïve T cells and bypass inhibitory chemokines produced by HIV-specific cytotoxic T lymphocytes (22). Administration of a CCR5 antagonist often expedites the coreceptor shift process (23). In some cases, the development of resistance to a CCR5 antagonist occurred at the cost of reduced viral fitness, as shown by observed reversion to pre-treatment strains upon the withdrawal of therapy (24-28). An understanding of the mechanism of coreceptor shift will guide rational design of better coreceptor inhibitors by blocking the potential paths to resistance development.

Despite recent advances in our understanding of HIV-1 envelope trimer structure and the structures of both coreceptors through high-resolution crystallography studies, details of molecular interactions between the envelope and coreceptors remain to be defined (29-35). Numerous studies have shown that the Variable loop 3 (V3) region of HIV-1 envelope is critical in coreceptor specificity by making direct contact with extracellular loops of the coreceptors (36-38). Regions outside V3, including the C2-V5 regions, that have been shown to regulate coreceptor tropism include specific residues at position 440 in C4, and regions in gp41 (Chapter 3) (39-42). Study of coreceptor antagonist resistant variants allows us to identify key envelope determinants for resistance development and coreceptor shift, especially in the context of a defined selection pressure (17, 43-47).

To better understand the evolutionary process of HIV-1 envelope quasispecies that led to drug resistance and coreceptor shift *in vivo* under the selection of a CCR5 antagonist, we analyzed a panel of 97 HIV-1 envelope single clones from 4 samples from 2 subjects who experienced treatment failure in a clinical trial for a CCR5 antagonist Vicriviroc (VCV). Using a

combination of phenotypic tropism assays and genotypic analysis, we showed a rapid change of tropism upon VCV treatment and detected signature changes in envelope V3 regions that correspond to the phenotypic coreceptor shift.

2.5 Materials and Methods

2.5.1 Patient samples

HIV-1 envelope amplicons were obtained from participants in a Vicriviroc (VCV) Phase II clinical trial (AIDS Clinical Trials Group [ACTG] A5211; NCT00082498) (48). Participant selection and sample preparation were described previously (Manuscript Zhang et al) (49). Plasma samples were collected from study participants who experienced virologic failure and had tropism change by the phenotypic Trofile assay (Monogram Biosciences). Pelleted virion RNA was used to prepared *env* cDNA. Envelope amplicons were generated using previously reported primers (50):

Env1Atopo (5'-CACCGGCTTAGGCATCTCCTATGGCAGGAAGAA-3')

FLenv2.2 (5'-AGCTGGATCCGTCTCGAGATACTGCTCCCACCC-3')

2.5.2 Single Clone Isolation

Patient envelope amplicons were cloned into expression vectors in frame. Two approaches were taken to generate single clones from patient samples. The first approach cloned patient envelope amplicon PCR products into pcDNA3.3-TOPO, a TA cloning vector containing a CMV promoter, by following the manufacturer's protocol (Thermal Fisher Scientific). The TA ligation reaction was prepared using QiaQuick PCR purification columns following the manufacturer's protocol (Qiagen) and electroporated into ElectroMAX Stbl4 competent cells (Thermo Fisher) in 1 mm cuvette at 1.2 kV, 25 μF, 200 Ω mA on BioRad MicroPulser (BioRad). Transformants were recovered in 1 mL of S.O.C. medium by shaking at 30 °C for 90 minutes at 225 RPM, and 20 – 100 μL was plated on Luria-Broth (LB) agar plates supplemented with 150

μg/mL of ampicillin (Sigma-Aldrich). Single colonies were separately transferred and grown in 3 mL of LB liquid media containing 150 μg/mL of ampicillin (Sigma-Aldrich) by shaking at 30 °C at 225 RPM for 18 – 20 hours. The resultant culture was used to prepare plasmid DNA using PureLink Quick Plasmid Miniprep Kit (Thermal Fisher Scientific). The single clone plasmids were screened for the presence or absence of the insert patient envelope gene and for the orientation of the insert using PCR using the following combinations of directional primers.

CMV-forward: 5'-CGCAAATGGGCGGTAGGCGTG-3'

TK polyA-reverse: 5´-CTTCCGTGTTTCAGTTAGC-3´

Env1Atopo-forward: 5'-CACCGGCTTAGGCATCTCCTATGGCAGGAAGAA-3'

FL-Env2.2-reverse: 5'-AGCTGGATCCGTCTCGAGATACTGCTCCCACCC-3'

A 3kb product with CMV-forward and TK polyA-reverse, and Env1Atopo-forward and FL-Env2.2-reverse primers confirmed the presence of HIV-1 envelope amplicon insert in pcDNA3.3 vector. A 3kb product with CMV-forward and FL-Env2.2-reverse primers confirmed that the envelope gene is inserted in the forward orientation downstream of the CMV promoter, whereas products with TK polyA-reverse and FL-Env2.2-reverse indicated a reverse orientation, unsuitable for follow-up expression experiments. Only pcDNA3.3-TOPO-CMV-Env clones containing an envelope gene insert in the correct orientation were used in pseudotyping experiments to test the functionality of individual envelope isolates.

The second approach directly cloned patient amplicon samples in frame into HIV-1 molecular clone pNL4-3.Luc.R^-.E^- (Dr. Nathaniel Landau, the NIH AIDS Reagent Program) to generate replication competent viral molecular clones in a directional manner, using Gibson Assembly® Master Mix (New England Biolabs) according to the manufacturer's protocol (51,

52). An AfeI site (AGCGCT) that does not alter the amino acid sequence was introduced at nucleotide 5954 of pNL4-3.Luc.R^{-}.E^{-} by site directed mutagenesis. A NotI site is present in the vector at the 4th codon of *nef*, the position in which the luciferase gene was inserted. The vector was linearized using AfeI and NotI. The insert was prepared by PCR amplification of the patient envelope amplicons using the following primers:

NL4-AfeI-EnvF:

5'-TTGTTTCATGACAAA AGCGCT AGGCATCTCCTATGGCAGGAAG -3'

NL4-NotI-EnvR:

5'- TTTTGGCGTCTTCAGCGGCCGCGCCACCCATCTTATAGCAAAATCCTTTC -3'

The PCR was performed with Q5 High-Fidelity 2× Master Mix (New England Biolabs) following the manufacturer's protocol, and run on BioRad T100 thermal cycler (BioRad) using the following conditions: 98°C for 30 sec; 15 cycles of 98 °C for 10 sec, 55 °C for 20 sec, and 72 °C for 2 min; and 72°C for 5 min. The PCR product included flanking regions of pNL4-3.Luc.R^{-}.E^{-}. The Gibson assembly reaction was performed with an insert to vector ratio of 3:1 at 50 °C for 60 minutes. The assembly product was diluted 1:3 and electroporated into ElectroMAX™ Stbl4™ competent cells as described above, and single clone were selected in the same manner for the first approach. Colony PCR and diagnostic restriction digestion were performed to screen for clones containing an insert. Plasmid DNA prepared as described above were used in tissue-culture based phenotypic functional assays as well as genotype sequencing.

2.5.3 Cell Lines

HEK 293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1x antibacterial, antimycotic solution (containing 100 U/mL penicillin, 100 μg/mL streptomycin, and 250 ng/mL amphotericin B). U87.CD4 cells were stably transfected with pBABE-CCR5-GFP or pBABE-CXCR5-GFP, constructed as previously described, and maintained in DMEM supplemented with 15% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1x antibacterial, antimycotic solution, 0.2 mg/mL G418, and 1 μg/mL puromycin (53).

2.5.4 Virus production

Pseudotyped HIV-1 virions containing individual envelope variants were produced by co-transfection of 5 μg of HIV-1 molecular clone pNL4-3.Luc.R^-.E^- plasmid DNA and 0.5 μg of pcDNA3.3-TOPO-CMV-Env into 4×10^5 HEK 293T cells in a 6-well plate format using 10 μL of *Trans*IT®-LT1 Transfection Reagent (Mirus). Replication-competent HIV-1 containing individual envelope isolates were produced by transfecting 5 μg of the molecular clone plasmid DNA as described above. The viral supernatant was harvested at 48 to 72 hours post transfection and passed through a 0.44 μm syringe filter to remove cell debris.

2.5.5 Luciferase-based infection assay

Freshly prepared viral stocks were used to infect U87.CD4.CCR5 and U87.CD4.CXCR4 cells in the presence of 8 μg/mL DEAE dextran, as previously described (53). U87.CD4.CCR5 and U87.CD4.CXCR4 cells were plated in 96-well plates at 10^4 cells in 100 μL of media per

well 24 hours before infection. During infection, 100 μL of freshly harvested viral supernatant was added to each well in triplicates, and incubated with cells at 37 °C, 5% CO_2, and 95% humidity for 12 hours. The virus was removed 12 hours after the start of infection, and the infection was continued in fresh media for another 36 hours to allow for expression of luciferase in the infected cells. Cells were harvested 48 hours post-infection and lysed with 0.2% Triton-X100 (Sigma-Aldrich) in PBS. The cell lysates were read for luciferase activity on an Optocomp I luminometer (MGM Instruments). A result was scored as positive if the relative light unit readout was more than 2 standard deviations over the mean of mock infected control wells. Cells infected with pNL4-3.Luc.R^-.E^- pseudotyped with vesicular stomatitis virus G glycoprotein envelope was used as the positive control, and cells infected with pNL4-3.Luc.R^-.E^- alone was used as the negative control.

2.5.6 Flow cytometry

The coreceptor-expressing U87-CD4 cells were stained with anti-CD4 antibody (clone Q4120) conjugated with phycoerythrin (Sigma-Aldrich), anti-CXCR4 antibody (clone 12G5) conjugated with allophycocyanin (BioLegend), and anti- CCR5 antibody (clone 2D7) conjugated with allophycocyanin at 4 °C in the dark for 20 minutes according to manufacturer's protocol (BD Pharmingen, BD Biosciences), washed with PBS three times, resuspended in 5% FBS in PBS, fixed with 2% paraformaldehyde, and run on FACSCalibur (BD Biosciences). Flow cytometry data were visualized with Flowing Software version 2.5 (http://www.flowingsoftware.com/).

2.5.7 Sequence alignment and phylogenetic analysis

Pairwise and multiple sequence alignments were performed using Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) and Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) on the European Bioinformatics Institute website. Phylogenetic trees of envelope variants were generated using PhyML (http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html) or Clustal Omega with Pearson/FASTA output and iTOL (http://itol.embl.de/). Evolutionary analyses of specific envelope regions via calculating dN/dS ratio were performed using Synonymous Non-synonymous Analysis Program (54) on the HIV database website (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html).

2.5.8 Tropism prediction

Tropism prediction was performed using the online version of Geno2Pheno[coreceptor] 2.5 following the developer's instructions at http://coreceptor.geno2pheno.org/. Only the Variable loop 3 (V3) region sequences were used for tropism prediction.

2.6 Results

2.6.1 Evaluation of HIV-1 patient samples by single clone analysis

HIV-1 patient samples were collected from eight study subjects at two time points on ACTG A5211 Phase II study of vicriviroc (VCV) (Zhang et al, manuscript). Based on the original phenotypic Trofile assay results, at study initiation, participants harbored viruses that used exclusively or preferentially CCR5 (Time point 1 at Week 0). At the second time point, these individuals experienced protocol-defined virologic failure, defined as less than 1 $\log_{10}$ decrease of HIV-1 RNA level. At this time point, plasma *env* amplicons demonstrated the ability to utilize CXCR4 for entry, and their tropism was designated as dual-mixed (DM). The time elapsed between the first and second time points varied among the study subjects from 2 to 32 weeks.

In order to define the phenotypes of quasispecies present in these samples, we isolated single clones from the library of HIV-1 envelope patient amplicons and performed in-depth phenotypic and genotypic characterization of individual envelope-expression clones (Figure 1). The TOPO TA cloning technology offered an approach to clone a mixture of similar-sized envelope amplicon PCR products into a CMV-promoter driven expression vector and to isolate bacterial single colonies after transformation. We assumed that one transformed bacterial colony contains a single ligated vector that contains one inserted HIV-1 *env* gene from the original patient amplicon samples. Therefore, the single colonies represent independent sampling of the pool of quasispecies in the original patient sample, offering a glimpse of the viral variants in the bloodstream of the patients at different times, during the course of VCV treatment. By sampling sufficient numbers of single clones, we defined the distribution of quasispecies variants, their

respective tropisms, and sequence similarities or differences, in order to define structure-function correlations of the envelope protein and its role in coreceptor specificity.

From the sixteen samples coming from eight study subjects, we selected four samples from two subjects for the single clone analysis. We chose to focus on these two subjects because their samples demonstrated different lengths of time for coreceptor shift. Subject 1 showed high R5-tropic viral activity at baseline, and comparable R5- and X4-tropic activity in only two weeks. In contrast, samples from Subject 2 remained predominantly R5-tropic at Week 0 and Week 8, but also showed significantly increased X4-tropism at the second time point based on the luciferase assay readings in the Trofile assay (Subjects 1 and 2, Figure 4A). The difference in times required for coreceptor-shift might relate to distinct evolutionary pathways to switch from R5-tropism to dual- or X4-tropism.

2.6.2 Phenotypic analysis of single clones consistent with Trofile results

We established a tissue-culture based luciferase-reporter infection assay that is similar to the bulk-screening Trofile assay. The luciferase-based infection assay consisted of producing pseudotyped or replication competent virions that expressed luciferase in place of Nef by transfecting 293T cells, and reading luciferase activity as an indicator of viral entry from infected U87.CD4.CCR5 or U87.CD4.CXCR4 cells (Figure 1). We generated U87.CD4 cell lines stably transfected and expressing CCR5-GFP or CXCR4-GFP fusion proteins on their surface (53). These cell lines expressed comparable levels of CD4 and coreceptor based on surface staining and GFP expression (Figure 2). The activity of envelope expression clones to utilize CD4 and one of the two coreceptors was quantified by luciferase measurements reported as relative light

units (RLUs). A representative luciferase-based infection assay result is shown in Figure 3. In this experiment, six out of eight clones from Sample 1.2 were functional and dual-tropic, while the other two clones were non-functional, defined as being unable to enter either X4- or R5-expressing U87.CD4 cells.

Using the aforementioned phenotypic assay, we characterized a total of 97 single clones from four samples, including 6 clones from Sample 1.1, 46 clones from Sample 1.2, 10 clones from Sample 2.1, and 35 clones from Sample 2.2 (Figure 4B-C). The outcome was consistent with the Trofile assay results which examined plasma viral samples in bulk. At the first time point, only R5-tropic quasispecies for Subjects 1 and 2 were found. At the second time point, 65% of single clones from Sample 1.2 were R5X4-tropic, and the remaining clones were non-functional. In contrast, in Sample 2.2, 17% of clones were R5-tropic, 20% of clones were R5X4-tropic, and 63% of clones were non-functional. .

2.6.3 Quantitation of molecular cloning artifact using laboratory adapted molecular clones

To control for mutations generated through the molecular cloning process, two lab adapted HIV-1 molecular clone strains p102 (pNL4.3) and p125 were processed in parallel with the patient amplicon samples. With the TOPO cloning approach, p102 and p125 envelope PCR products were ligated into pcDNA3.3-TOPO-CMV vector. Single clones were isolated and tested for phenotypic functionality in the luciferase-based tropism assay, as well for genotypic variation via sequencing. We observed on average 12.5% of single clones being non-functional, and there were minimal numbers of sequence variations among the non-functional control single clones (Table 1). With the Gibson Assembly (GA) clone approach, we performed three sets of

GA ligation using envelopes from p102, p120, and 1:1 mixture of p102 and p120 as inserts. After quality control using colony PCR and diagnostic restriction digestion, the clones that were confirmed to have an insert with intact restriction sites were used in phenotypic functional testing. Out of all 23 clones tested from three sets combined, 3 clones were non-functional, resulting in a 13.0% non-functional rate due to cloning artifacts (Table 1). The level of non-functional clones and sequence diversity observed in the controls was 3-5 folds lower comparing to the level of non-functional clones and sequence diversity present in the single clones isolated from patient amplicon samples. Therefore, the striking sequence variations observed in Sample 1.1 and 1.2 single clones were unlikely an artifact introduced by molecular cloning, but most likely reflects the sequence variation in viral quasispecies present *in vivo* within the patients.

2.6.4 Single clones within one time point were highly similar in sequence

To understand the structural determinants that affect coreceptor specificity, envelope clones were subjected to sequence analysis. Alignment of clones from Sample 1.2 by Clustal Omega showed a strong selection pressure towards two major variants (Figure 5). Among the six clones, four of them exhibited identical amino acid sequences. One of the four contained two synonymous mutations. Spanning the length of the partial variable loop 2 (V2) region, and the complete V3 and V4 regions, there were in total four positions that differed among the six clones. Examination of the V3 region revealed that all of six clones from Sample 1.2 bear signature of X4-utilization, including a glycine at the 11th position, a lysine at the 25th position, and overall positive charge of +8. Prediction of tropism, using the Geno2Pheno[coreceptor] algorithm, correctly identified all six clones to be X4-tropicwith a low false positive rate. Overall, the variants from Sample 1.2 have all been shown to be dual-tropic, and the V3 signatures were

consistent with functional characterization, based on previous studies. Moreover, there was a lack of diversity among the variable regions of the six clones from Sample 1.2, suggesting a strong selection pressure towards one particular genotype. These suggested that it might be advantageous for this virus subtype for replication in the presence of CCR5 inhibitor VCV, resulting in rapid change to an abundant circulating genotype within just two weeks of treatment.

2.6.5 Genotypic analysis of single clones over time shows great diversity

Longitudinal analysis of variant sequences from before and after VCV treatment of the same subject revealed significant quasispecies changes over time (Figure 6). Examination of the alignment showed that amino acid substitutions mostly concentrated in variable regions V2, V3, and V4. Insertions and deletions were found exclusively in the V2 and V4 region. The conserved regions also contained amino acid substitutions which were located more sparsely than those in the variable regions. The V3 regions from samples before and after treatment showed sequence signatures that corresponded to R5- and X4-usage. In particular, the 25th position in the V3 region was a negatively charged glutamic acid in the R5-tropic Sample 1.1 clone we sampled, whereas it was converted to a positively charged lysine in the X4-tropic Sample 1.2 clone. This observation was consistent with previous reports that the variant is more likely to use CXCR4 if there is at least one positively charged residue at the 11th and 25th position (glycine and glutamic acid in the 1.1 clones; glycine and lysine in the 1.2 clone; highlighted in Figure 6) of V3. Genotypic analysis of the V3 region of these two clones demonstrated sequence signatures corresponding to a change of coreceptor specificity and was consistent with the luciferase-based phenotypic tropism assay results for these two clones, respectively.

In addition, a quantitative analysis of the alignment revealed a high level of sequence diversity in the V2-C3-V3-C4-V4 regions between a major clone from Sample 1.2 and one representative clone from Sample 1.1 (Table 1). On average, 19 – 20% of sequences were different in these two clones through the above mentioned regions. The variable regions showed over 20% of sequence differences, with the V4 region being the most variable (36 – 41% of sequences being different). The conserved regions C3 and C4 were less variable as expected, with 13% of sequence diversity between the two clones from before and after treatment. Typically, the intra-clade *env* sequence variability was reported to be 10 – 15% among all circulating strains. One longitudinal study also showed that viral V3 sequences in chronically infected individuals stayed largely constant over time, with limited within-person evolution. Therefore, it was surprising to discover a much higher percentage of sequence variation in Subject 1 from single clones that arose over a relatively short time period of two weeks.

2.6.6 Phylogenetic analysis of envelope sequences consistent with their phenotypic function

To dissect the evolutionary relationship between the HIV-1 *env* single clones, we generated a maximal parsimony (MP) tree using the V3 regions from the six single clones from Sample 1.2 and the single clone from Sample 1.1 (Figure 7). The lab-adapted strain HXB2 that uses CXCR4 exclusively for entry was included as a reference. The MP tree showed clustering of dual-tropic clones from Sample 1.2 together on a single branch and away from the R5-tropic clones from Sample 1.1. HXB2 sequence was furthest away from the patient sample clones. Overall, the sequence based phylogenetic analysis was consistent with the phenotypically validated tropism of these single clones, suggesting sequence-based analysis of envelope could be a valuable tool in predicting functional tropism. In addition, single clones that were close on a

phylogenetic tree but potentially with discordant tropisms could offer insights in important determinants in coreceptor shift.

2.7 Discussion

In this study, we utilized a rapid, reliable sampling method to isolate a large number of single clones from clinical trial subjects who developed resistance to VCV therapy. By performing tissue-culture based phenotypic assay and sequencing-based genotypic analysis on single clones, we were able to probe the viral quasispecies before and after VCV treatment *in vivo,* and understand how the HIV-1 envelope evolves in the face of a specific CCR5 antagonist selection. The advantage of the single clone analysis rested in the ability to fully characterize each individual clone and interrogate the structure-function relationship of HIV-1 envelope by making associations between key sequence determinants and its specific tropism status. In addition, the availability of the full-length envelope sequence allowed for investigation of non-V3 determinants in the process of coreceptor shift, and phylogenetic studies looking at clones before and after coreceptor shift would potentially shed light on the evolutionary pathway for this process.

Single clone isolation and analysis from patient samples is an established and widely used approach in studying HIV-1 quasispecies. Depending on the cloning and screening methodology, single clone analysis with a large enough sampling size provides a survey of the quasispecies composition *in vivo* at different time points. A caveat with our approach using the pcDNA3.3-TOPO-CMV was a potential bias in the directional TOPO cloning strategy. We observed approximately 70 – 95% of reverse-orientation insert, much higher and statistically significant ($p<0.01$ by Chi square test), than the predicted 50% chance if the insertion orientation were truly random. This could potentially be due to a selection against envelope gene inserted in the correct orientation downstream of the CMV promoter. A plausible speculation is that expression of envelope protein, even at a low level from a eukaryotic promoter, in Stbl4

bacterial cell leads to toxicity or cell death. There were previous reports on the leakiness of CMV promoter in prokaryotic systems (55). This would also result in an enrichment of non-functional clones that harbor premature stop codon and do not produce full-length, functional envelope protein. To mitigate this potential pitfall, we adopted a Gibson Assembly ligation system that directly cloned patient envelope amplicons into a HIV-1 molecular clone. Since the HIV-1 LTR is a much weaker promoter in *E. coli*, the problem from toxicity of aberrantly expressed viral gene products should be largely alleviated if not completely eliminated. Single clones from Subject 2 were isolated using the latter approach.

In the phenotypic analysis, we observed a significant proportion of non-functional single clones, ranging from 50% in Sample 1.1, 35% in Sample 1.2, to 60% in Sample 2.1 and 63% in Sample 2.2. Given that we observed a 12.5% non-functional rate in the lab adapted p102 and p125 control experiments, the large percentage of non-functional single clones cannot be attributed to cloning artifacts alone. A plausible explanation is that these single clones are a natural product of viral replication. Since these patients did not have effective suppression of viral replication, the virus was actively replicating despite of the antiretroviral therapy. The high mutation rate of viral reverse transcriptase combined with a rapid viral turnover rate could produce numerous mutations in the progeny virus, many of these mutations being deleterious in nature by chance. In addition, a recent latency study showed that the viral quasispecies present in the peripheral blood are more likely to be defective and non-infectious in nature, comparing to the quasispecies in the latently infected compartment in the lymphoid organs (56). This is based on the observation that in chronically infected patients on HAART, there are more clonally expanded, dividing T cells that harbor defective provirus and release them to the circulation, in contrast to quiescent, non-dividing memory T cells where the replication-competent virus

mostly reside. Therefore, the large fraction of non-functional single clones we detected is likely a true reflection of the viral quasispecies landscape in the plasma of the treatment-experienced study participants.

From the phenotypic and genotypic analysis of the single clones from Sample 1.1 and 1.2, we observed a complete shift of tropism from R5-tropic to dual-tropic clones, as well as two representative sequences that were, on average, 20% different. This dramatic change in tropism and sequences of quasispecies suggested rapid expansion of one or a several related dual-tropic quasispecies during the two weeks of VCV treatment. In addition, the six clones sequenced from Sample 1.2 were highly similar, suggesting a strong selection pressure towards this homogenous genotype. Overall, a phylogenetic analysis of the 1.1 and 1.2 clones showed that clones with similar functional tropism status were also evolutionarily closer to each other than clones with a different tropism. By analyzing the functions and sequences of individual clones, we were able to gather important information on the evolution of HIV-1 envelope quasispecies in the context of coreceptor inhibitor treatment.

On the other hand, the single clones from Subject 2 showed mixed tropism even after VCV treatment. More detailed analysis of the 2.1 and 2.2 clones, such as dN/dS analysis on the full-length, variable and conserved regions of the envelope sequence, and maximal parsimony or neighbor joining tree analysis, are needed to understand the evolutionary relationship between these clones, and the potential drug-resistance mechanism of the R5-tropic clones from Sample 2.2. A pair of evolutionarily close clones with discordant tropism status would potentially reveal the critical difference that leads to coreceptor shift from R5- to X4-usage.We have not been able to detect such a pair in the single clone analysis, likely due to the limited sample size. It is possible that we could detect such pairs of sequences using deep sequencing approaching.

Alternatively, such tropism-shift intermediate clones might not be detected at all in the patient plasma samples, if the viral quasispecies undergone the mutational process largely reside in latently infected reservoir or lymph tissues that are not sampled in this study. This is due to the fact that

One disadvantage of the single clone analysis lies in its power of this analysis, in particular when the number of single clones sampled is insufficient to provide an accurate representation of the pool of quasispecies. In this study, we analyzed a relatively large number of single clones from the samples of the second time point, post VCV treatment failure. Additionally, we performed next generation sequencing (NGS) experiments that systematically examined the quasispecies community (Chapter 3), which complemented and validated the single clone analysis. The combination of depth from the single clone analysis and breadth from the deep sequencing analysis provides a comprehensive understanding of envelope evolution *in vivo* and the key determinants associated with coreceptor shift. The new information of the structure-function relationship of HIV-1 envelope and the mechanism of coreceptor shift could potentially provide new prognostic tools for HIV-1 infected patients, as well as new insights applicable to coreceptor inhibitor design.

2.8 Funding sources

This work was supported by National Institutes of Health grant R01 AI106361 to L.R., UM1 AI068636 and R37 AI055357 to D.R.K.

2.9 Acknowledgements

We thank Dr J. Fay for guidance in dN/dS analysis, Drs. Rauch, McCulley, Leibel, and Huang for intellectual discussions, and Ms N. Campbell, Mr J. Zhang and Mr. M. Madans for technical assistance. We also thank Dr. Roy M. Gulick and members of the ACTG A5211 protocol team, as well the participating ACTG Clinical Research Sites, for providing the plasma samples used in this study.

2.10 Tables

Table 2.10.1. Summary of phenotypic and genotypic analysis for control single clones

Strains	Tropism	Number of Clones	Non-functional	Mutations*
TOPO TA cloning with pcDNA3.3-TOPO-CMV vector				
p102**	X4	12	2	5 substitutions in 2 clones
p125***	R5	12	1	3 substitutions in 1 clone
Gibson Assembly cloning into pNL4.3 E-R-Luc vector				
p120	X4	8	2	N.D.****
p125	R5	8	1	N.D.
p102 / p125 (1:1)	X4 and R5	7	0	N.D.

*Mutations: all the mutations observed in the non-functional full-length envelope clones.

** p102:NL4.3 laboratory adapted strain

***p125: NL-HXADA-GG, 2.7 kb of SalI-BamHI (5785-8474) HXADA sequence into p102 NL4.3 backbone; HXADA comprises of BglII-BglII (7040-7620) ADA sequence into HXB2 backbone.

****N.D.: not determined

Table 2.10.2. Sequence comparison between one major clone from 1.2 and one clone from 1.1

Regions*	**Length****	**Substitutions**	**Diversity*****	**Indels**[+]
V2 through V4	272 – 279	53	19 - 19.5%	6
V2	57 - 59	13	22 – 22.8%	2 (-4, +6)
V3	35	8	22.9%	0
V4	29 - 33	12	36.4 – 41.4%	3 (+1, -1, -4)
C3 and C4 combined	151 - 152	20	13.2%	1

*Regions: Variable loop 2, 3 and 4 (V2, V3 and V4) as shown in Figure 6. Conserved regions 3 and 4 (C3 and C4) are located between V2/V3, and V3/V4 respectively.

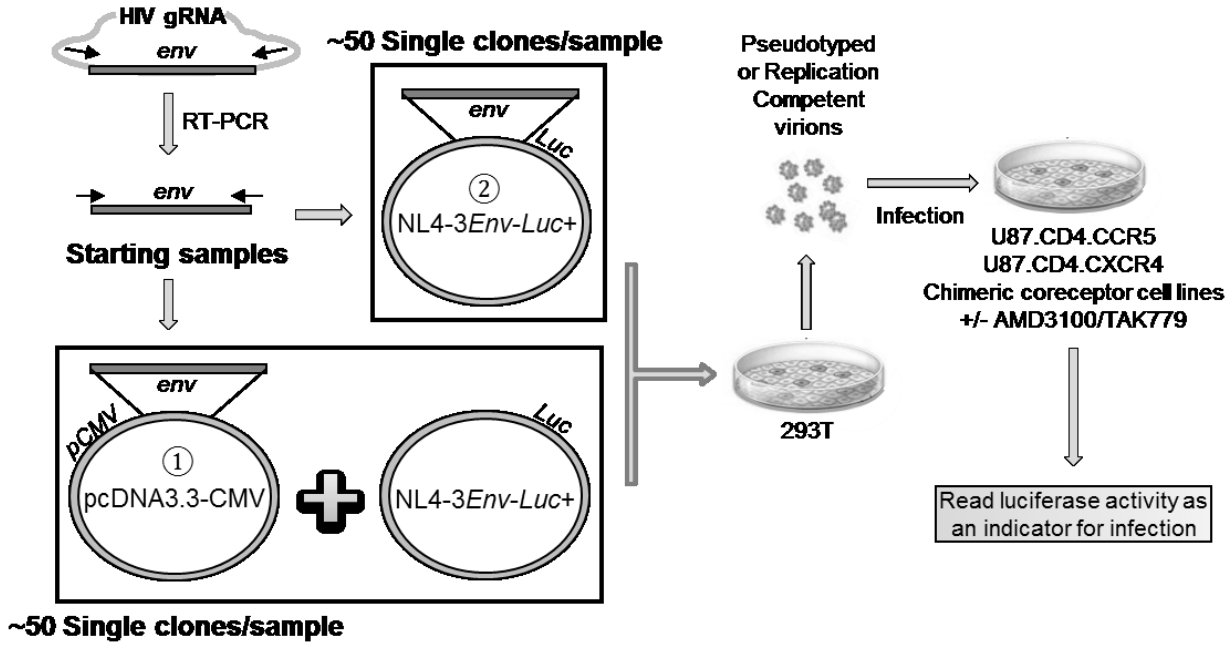
** Length (Lackner, #2553): length of the specified regions in amino acid. In the case of insertions and deletions, the length of the particular region is shown as a range.

***Diversity: calculated as number of amino acid substitutions divided by the length of the region, presented in percentage.

[+]Indels: insertions and deletions. The number of indels and their nature are presented. For example, in V2 region, there are 2 indels; one is a 4 amino acid deletion, the other is a 6 amino acid insertion.

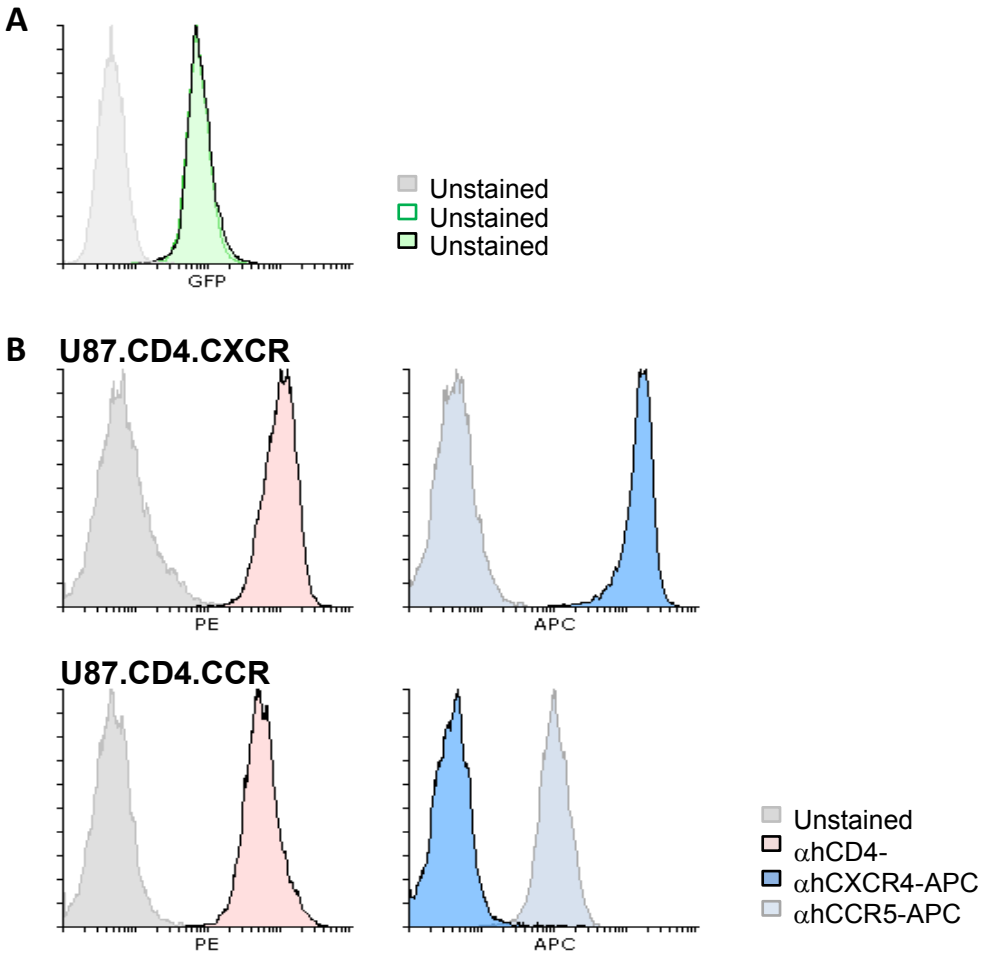
2.11 Figures

Figure 2.11.1 Schematic diagram of the experimental design.



Patient derived envelope quasispecies amplicons were cloned into (1) pcDNA3.3-TOPO-CMV expression vectors, or (2) NL4.3*Env-luc+* reporter vector (plasmid libraries) and expressed in 293T cells as (1) pseudotyped or (2) replication competent virions that were used to infect coreceptor expressing cell lines in the presence or absence of coreceptor inhibitors. Infected cells were lysed and read for luciferase activity as an indicator for infection. Approximately 50 clones were screened for each patient sample. The envelope genes from the single clones were also analyzed by Sanger sequencing.

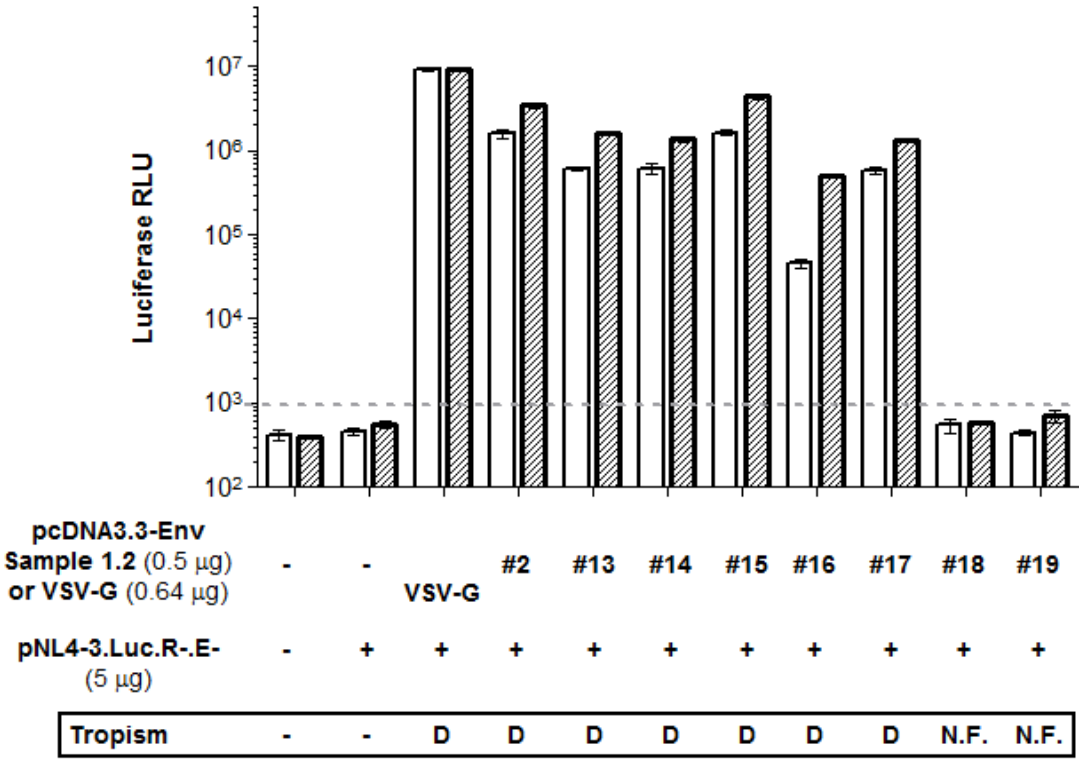
Figure 2.11.2 Coreceptor-GFP cell lines express comparable levels of CD4 and coreceptors on the surface.



(A) Unstained U87.CD4.CXCR4 (X4) and U87.CD4.CCR5 (R5) cell lines show similar level of GFP expression comparing to the parental U87.CD4 cell line.

(B) X4 and R5 cell lines show comparable levels of surface CD4 staining as well as positive CXCR4 and CCR5 surface expression by flow cytometry comparing to unstained or negative controls. Different levels of mean fluorescence intensity for CXCR4 staining on X4 cells and CCR5 staining on R5 cells are likely due to the inherent difference between the two conjugated antibodies against these surface antigens, and not a reflection of the coreceptor surface expression levels.

Figure 2.11.3 A representative luciferase-based phenotypic tropism assay for envelope single clones.



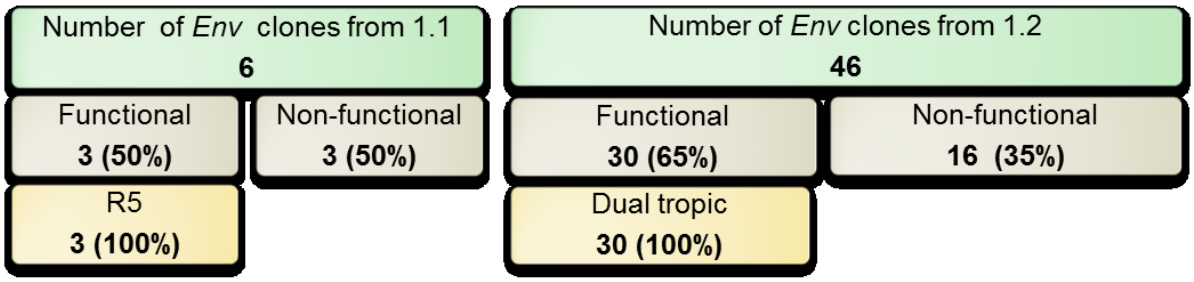
Out of 46 envelope single clone from Sample 1.2 tested, 8 clones are shown here. RLU: relative light units. D: dual-tropic, ability to enter using both X4 and R5 coreceptors. N.F.: non-functional, the inability to use either X4 or R5 for entry. Negative controls include mock transfection without DNA and mock infection with non-pseudotyped NL4.3-Luc Env knockout virus. Positive control was NL4.3-Luc.R-.E- pseudotyped with VSV-g envelope. Out of the eight clones shown in this experiment, six of them are functional and can infect both CXCR4 and CCR5 expressing cells.

Figure 2.11.4 Summary of single clone tropism analysis.

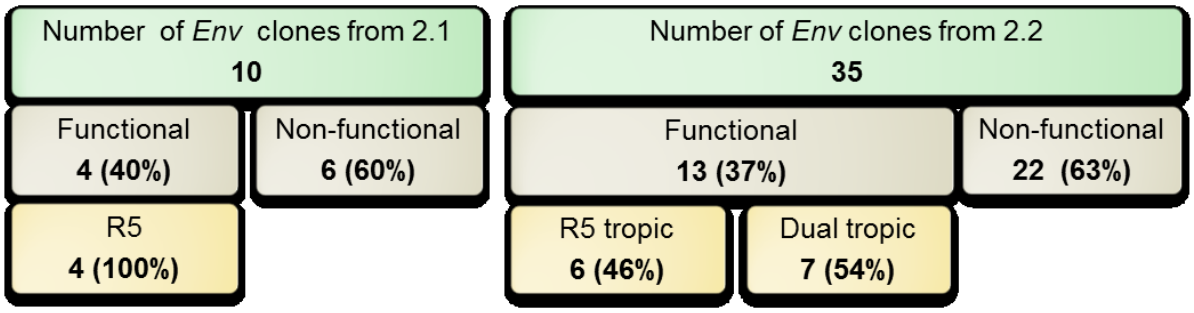
A

Subject ID	Sample ID	Time (weeks)	R5-RLU	X4-RLU	Tropism
1	1.1	0	593008	86	R5
	1.2	2	139491	185312	DM
2	2.1	0	7360	129	R5
	2.2	8	6260	2034	DM

B



C



(A) Tropism analysis from bulk patient sample. Subject: HIV-1 patients who experienced virological failure (defined as smaller than 1 $\log_{10}$ viral load decrease at or after week 16 by protocol) in a Vicriviroc phase IIb clinical trial. RLU: Relative Light Units by the original luciferase-based Trofile assay on CCR5 (R5) or CXCR4 (X4) expressing cell lines. Tropism: the ability of the virus to utilize R5 and/or X4 for viral entry. DM: dual/mixed tropism. Dual tropism: the ability to utilize both R5 and X4 for entry; mixed tropism: coexistence of viral species that can use either R5 or X4 coreceptor for entry.

(B) Single clone tropism analysis for Sample 1.1 and 1.2. **(C)** Single clone tropism analysis for Sample 2.1 and 2.2.

Figure 2.11.5 Variable loop region sequence analysis among 6 clones from Sample 1.2.

```
V2
#7   TYALFYKLDIVPVNDTDETSHNKNNSTNGASYNFTSYRLISC
#4   TYALFYKLDIVPVNDTDETSHNKNSSTNGASYNFTSYRLISC
#2   TYALFYKLDIVPVNDTDKTSHNKNNSTNGASYNFTSYRLISC
#13  TYALFYKLDIVPVNDTDKTSHNKNNSTNGASYNFTSYRLISC
#14  TYALFYKLDIVPVNDTDKTSHNKNNSTNGASYNFTSYRLISC
#22  TYALFYKLDIVPVNDTDKTSHNKNNSTNGASYNFTSYRLISC
                     *      *

V3
#7   CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC
#4   CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIKKAYC
#2   CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC
#13  CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC
#14  CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC
#22  CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC
                                   *

V4
#7   CDTTELFNHTRWPNGTWNTDESENITLPC
#4   CDTTKLFNHTRWPNGTWNTDESENITLPC
#2   CDTTELFNHTRWPNGTWNTDESENITLPC
#13  CDTTELFNHTRWPNGTWNTDESENITLPC
#14  CDTTELFNHTRWPNGTWNTDESENITLPC
#22  CDTTELFNHTRWPNGTWNTDESENITLPC
         *
```

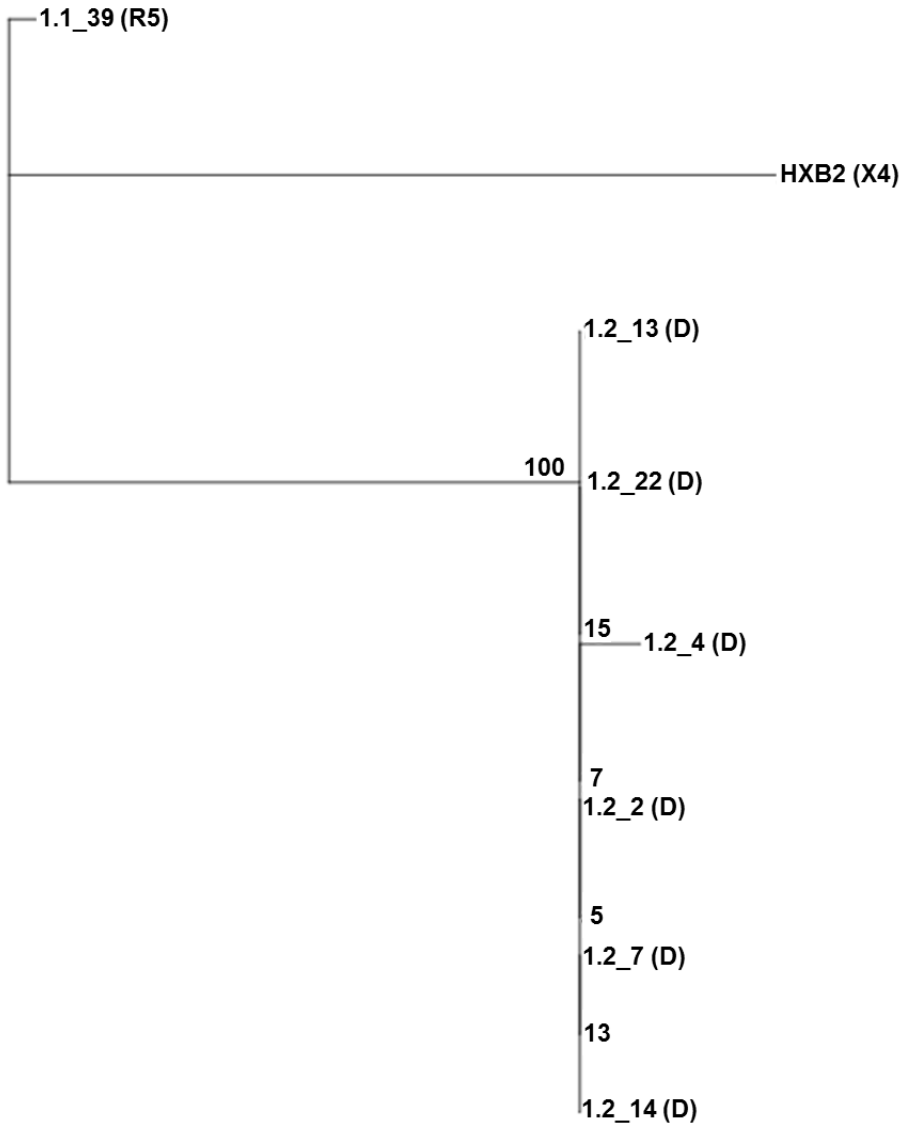
Partial sequences of V2, complete sequences of V3 and V4 from six 1.2 clones were aligned at both nucleotide and amino acid level. Four clones (2, 13, 14, and 22) are identical at amino acid level, though two synonymous mutations were found in clone #2. Clone #4 and #7 contains four non-synonymous mutations total (annotated by asterisk * and minor variants labelled with squares). The 11th and 25th position in the V3 loop were highlighted and contain a glutamic acid and lysine residue respectively.

Figure 2.11.6 V2 through V4 region sequence analysis between a major clone from Sample 1.2 and a single clone from Sample 1.1.

```
                                 V2 region
1.2 csfnvtttmrdkmQRTYALFYKLDIVPV------NDTDKTSHNKNNSTNGASYNFTSYRLISCNTSVLTQACPKVS
1.1 csfnvtttmrdkmQRTYALFYKLDIVPIdnstnnNNTSCTTHNDTSN----NCNYTSYRLVSCDTSVVTQACPKVS
                              ******* * ** * *********** *     *  *

1.2 FEPIPIHYCAPAGFAILKCNNKTFDGKGTCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSDNFSDNVKTIIV
1.1 FEPIPIHYCAPAGFAILKCNNKTFSGKGSCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEIVIRSANFSDNAKTIIV
                            *  *                                *    *    *

                V3 region
1.2 QLNESVEINCTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYCIINRTRWNDTLKKIAVKLIEQYGNTTII-FNRSS
1.1 QLNESVKIECIRPNNNTRKGIHIGPGRAFYATGEIIGDIRQAHCNLNRTAWNNTLKQIATKLREQYGNTTIIaFNKSA
       * * *                 * ****      * * **   *  *   * *  *        *  * *

                          V4 region
1.2 GGDPEIATHSFNCGGEFFYCDTTELFNHTRWP-NGTWN----TDESENITLPCRIRQIVNRWQEVGKAMYAPPINGII
1.1 GGDLEIVTHSFNCGGEFFYCNTSQLFNST-WPaNGAVSysirNNTADNITLPCRIKQIIKRWQEVGTAMYAPPISGII
    * *             * **   * *  *  ************        *  **      *      *
```

Non-synonymous mutations are annotated with asterisk (*). The 11th and 25th position in the V3 loop were labelled by black squares. Numerous substitutions, deletions, and insertions were detected between the single clone from 1.1 and the major clone form 1.2

Figure 2.11.7 Maximal parsimony tree based on V3 sequences for six Sample 1.2 clones, one Sample 1.1 clone and HXB2



The analysis was performed with 100 iterations. Boot-strapping values are displayed next to the nodes between the neighboring leaves or branches. Phenotypically validated tropism was annotated on the side. D: dual-tropism. R5: CCR5-tropism. HXB2: a lab adapted reference strain that utilizes X4 for entry. The tree was built and visualized using PhyML on HIV database website.

2.12 References

1. **Clapham PR, McKnight A.** 2001. HIV-1 receptors and cell tropism. Br Med Bull **58:**43-59.

2. **Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR.** 1997. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med **185:**621-628.

3. **Bieniasz PD, Cullen BR.** 1998. Chemokine receptors and human immunodeficiency virus infection. Front Biosci **3:**d44-58.

4. **Gorry PR, Ancuta P.** 2011. Coreceptors and HIV-1 pathogenesis. Curr HIV/AIDS Rep **8:**45-53.

5. **Tagliamonte M, Tornesello ML, Buonaguro FM, Buonaguro L.** 2011. Conformational HIV-1 envelope on particulate structures: a tool for chemokine coreceptor binding studies. J Transl Med **9 Suppl 1:**S1.

6. **Baba M.** 2006. Recent advances of CCR5 antagonists. Curr Opin HIV AIDS **1:**367-372.

7. **de la Tribonniere X, Yazdanpanah Y, Reynes J.** 2008. [CCR5 antagonists: a new class of antiretrovirals]. Med Mal Infect **38 Suppl 1:**S1-6.

8. **Parra J, Portilla J, Pulido F, Sanchez-de la Rosa R, Alonso-Villaverde C, Berenguer J, Blanco JL, Domingo P, Dronda F, Galera C, Gutierrez F, Kindelan JM, Knobel H, Leal M, Lopez-Aldeguer J, Marino A, Miralles C, Molto J, Ortega E, Oteo JA.** 2011. Clinical utility of maraviroc. Clin Drug Investig **31:**527-542.

9. **Scholten DJ, Canals M, Maussang D, Roumen L, Smit MJ, Wijtmans M, de Graaf C, Vischer HF, Leurs R.** 2012. Pharmacological modulation of chemokine receptor function. Br J Pharmacol **165:**1617-1643.

10. **Woollard SM, Kanmogne GD.** 2015. Maraviroc: a review of its use in HIV infection and beyond. Drug Des Devel Ther **9:**5447-5468.

11. **Henrich TJ, Kuritzkes DR.** 2013. HIV-1 entry inhibitors: recent development and clinical use. Curr Opin Virol **3:**51-57.

12. **Carter PH, Brown GD, Cherney RJ, Batt DG, Chen J, Clark CM, Cvijic ME, Duncia JV, Ko SS, Mandlekar S, Mo R, Nelson DJ, Pang J, Rose AV, Santella JB, 3rd, Tebben AJ, Traeger SC, Xu S, Zhao Q, Barrish JC.** 2015. Discovery of a Potent and Orally Bioavailable Dual Antagonist of CC Chemokine Receptors 2 and 5. ACS Med Chem Lett **6:**439-444.

13. **Low AJ, McGovern RA, Harrigan PR.** 2009. Trofile HIV co-receptor usage assay. Expert Opin Med Diagn **3:**181-191.

14. **Roche M, Jakobsen MR, Sterjovski J, Ellett A, Posta F, Lee B, Jubb B, Westby M, Lewin SR, Ramsland PA, Churchill MJ, Gorry PR.** 2011. HIV-1 escape from the CCR5 antagonist maraviroc associated with an altered and less-efficient mechanism of gp120-CCR5 engagement that attenuates macrophage tropism. J Virol **85:**4330-4342.

15. **Moore JP, Kuritzkes DR.** 2009. A piece de resistance: how HIV-1 escapes small molecule CCR5 inhibitors. Curr Opin HIV AIDS **4:**118-124.

16. **Lobritz MA, Ratcliff AN, Arts EJ.** 2010. HIV-1 Entry, Inhibitors, and Resistance. Viruses **2:**1069-1105.

17. **Berro R, Klasse PJ, Jakobsen MR, Gorry PR, Moore JP, Sanders RW.** 2012. V3 determinants of HIV-1 escape from the CCR5 inhibitors Maraviroc and Vicriviroc. Virology **427:**158-165.

18. **Sterjovski J, Roche M, Churchill MJ, Ellett A, Farrugia W, Gray LR, Cowley D, Poumbourios P, Lee B, Wesselingh SL, Cunningham AL, Ramsland PA, Gorry PR.** 2010. An altered and more efficient mechanism of CCR5 engagement contributes to macrophage tropism of CCR5-using HIV-1 envelopes. Virology **404:**269-278.

19. **Delgado R.** 2008. [Mechanisms of resistance and failure of treatment with maraviroc]. Enferm Infecc Microbiol Clin **26 Suppl 11:**28-33.

20. **Moir S, Chun TW, Fauci AS.** 2011. Pathogenic mechanisms of HIV disease. Annu Rev Pathol **6:**223-248.

21. **Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, Deymier MJ, Ende ZS, Klatt NR, DeZiel CE, Lin TH, Peng J, Seese AM, Shapiro R, Frater J, Ndung'u T, Tang J, Goepfert P, Gilmour J, Price MA, Kilembe W, Heckerman D, Goulder PJ, Allen TM, Allen S, Hunter E.** 2014. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science **345:**1254031.

22. **Schuitemaker H, van 't Wout AB, Lusso P.** 2011. Clinical significance of HIV-1 coreceptor usage. J Transl Med **9 Suppl 1:**S5.

23. **Sguanci L, Bagnoli F, Lio P.** 2007. Modeling HIV quasispecies evolutionary dynamics. BMC Evol Biol **7 Suppl 2:**S5.

24. **Madsen TV, Gerstoft J, Nielsen C, Jorgensen LB.** 2008. Reappearance of an 11-year-old sequence in an HIV-1 infected patient during treatment interruption. Scand J Infect Dis **40:**174-182.

25. **Westby M, van der Ryst E.** 2010. CCR5 antagonists: host-targeted antiviral agents for the treatment of HIV infection, 4 years on. Antivir Chem Chemother **20:**179-192.

26. **Tsibris A, Leopold K, Paredes R, Coakley E, Su Z, Hughes M, Greaves W, Flexner C, Gulick R, Kuritzkes D.** Reduced fitness of vicriviroc-resistant HIV-1 leads to rapid post-treatment reversion in vivo, p 3-6. *In* (ed),

27. **Van der Ryst E, Westby M.** Changes in HIV-1 co-receptor tropism for patients participating in the maraviroc MOTIVATE 1 and 2 clinical trials, p. *In* (ed),

28. **Anastassopoulou CG, Marozsan AJ, Matet A, Snyder AD, Arts EJ, Kuhmann SE, Moore JP.** 2007. Escape of HIV-1 from a small molecule CCR5 inhibitor is not associated with a fitness loss. PLoS Pathog **3:**e79.

29. **Harris A, Borgnia MJ, Shi D, Bartesaghi A, He H, Pejchal R, Kang YK, Depetris R, Marozsan AJ, Sanders RW, Klasse PJ, Milne JL, Wilson IA, Olson WC, Moore JP, Subramaniam S.** 2011. Trimeric HIV-1 glycoprotein gp140 immunogens and native HIV-1 envelope glycoproteins display the same closed and open quaternary molecular architectures. Proc Natl Acad Sci U S A **108:**11440-11445.

30. **Munro JB, Mothes W.** 2015. Structure and Dynamics of the Native HIV-1 Env Trimer. J Virol **89:**5752-5755.

31. **Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, Stewart-Jones GB, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L, Munro JB, Blanchard SC, Mothes W, Connors M, Kwong PD.** 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature **514:**455-461.

32. **Do Kwon Y, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gorman J, Joyce MG, Guttman M, Ma X, Narpala S, Soto C, Terry DS, Yang Y, Zhou T, Ahlsen G, Bailer RT, Chambers M, Chuang GY, Doria-Rose NA, Druz A, Hallen MA, Harned A, Kirys T, Louder MK, O'Dell S, Ofek G, Osawa K, Prabhakaran M, Sastry M, Stewart-Jones GB, Stuckey J, Thomas PV, Tittley T, Williams C, Zhang B, Zhao H, Zhou Z, Donald BR, Lee LK, Zolla-Pazner S, Baxa U, Schon A, Freire E, Shapiro L, Lee KK, Arthos J, Munro JB, Blanchard SC, Mothes W, Binley JM, et al.** 2015. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. Nat Struct Mol Biol **22:**522-531.

33. **Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC.** 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330:**1066-1071.

34. **Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B.**

2013. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science **341:**1387-1390.

35. **Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, Fenalti G, Wu H, Han GW, Cherezov V, Abagyan R, Stevens RC, Handel TM.** 2015. Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. Science **347:**1117-1122.

36. **Gu R, Westervelt P, Ratner L.** 1993. Role of HIV-1 envelope V3 loop cleavage in cell tropism. AIDS Res Hum Retroviruses **9:**1007-1015.

37. **Hartley O, Klasse PJ, Sattentau QJ, Moore JP.** 2005. V3: HIV's switch-hitter. AIDS Res Hum Retroviruses **21:**171-189.

38. **Cardozo T, Kimura T, Philpott S, Weiser B, Burger H, Zolla-Pazner S.** 2007. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. AIDS Res Hum Retroviruses **23:**415-426.

39. **Ogert RA, Wojcik L, Buontempo C, Ba L, Buontempo P, Ralston R, Strizki J, Howe JA.** 2008. Mapping resistance to the CCR5 co-receptor antagonist vicriviroc using heterologous chimeric HIV-1 envelope genes reveals key determinants in the C2-V5 domain of gp120. Virology **373:**387-399.

40. **Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, Mosier DE.** 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol **80:**750-758.

41. **Cashin K, Sterjovski J, Harvey KL, Ramsland PA, Churchill MJ, Gorry PR.** 2014. Covariance of charged amino acids at positions 322 and 440 of HIV-1 Env contributes to coreceptor specificity of subtype B viruses, and can be used to improve the performance of V3 sequence-based coreceptor usage prediction algorithms. PLoS One **9:**e109771.

42. **Dobrowsky TM, Rabi SA, Nedellec R, Daniels BR, Mullins JI, Mosier DE, Siliciano RF, Wirtz D.** 2013. Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins. Sci Rep **3:**3014.

43. **Berro R, Sanders RW, Lu M, Klasse PJ, Moore JP.** 2009. Two HIV-1 variants resistant to small molecule CCR5 inhibitors differ in how they use CCR5 for entry. PLoS Pathog **5:**e1000548.

44. **Garcia-Perez J, Staropoli I, Azoulay S, Heinrich JT, Cascajero A, Colin P, Lortat-Jacob H, Arenzana-Seisdedos F, Alcami J, Kellenberger E, Lagane B.** 2015. A single-residue change in the HIV-1 V3 loop associated with maraviroc resistance impairs CCR5 binding affinity while increasing replicative capacity. Retrovirology **12:**50.

45. **McNicholas PM, Mann PA, Wojcik L, Phd PQ, Lee E, McCarthy M, Shen J, Black TA, Strizki JM.** 2011. Mapping and characterization of vicriviroc resistance mutations from HIV-1 isolated from treatment-experienced subjects enrolled in a phase II study (VICTOR-E1). J Acquir Immune Defic Syndr **56:**222-229.

46. **McNicholas P, Vilchez RA, Greaves W, Kumar S, Onyebuchi C, Black T, Strizki JM.** 2012. Detection of HIV-1 CXCR4 tropism and resistance in treatment experienced subjects receiving CCR5 antagonist-Vicriviroc. J Clin Virol **55:**134-139.

47. **Roche M, Salimi H, Duncan R, Wilkinson BL, Chikere K, Moore MS, Webb NE, Zappi H, Sterjovski J, Flynn JK, Ellett A, Gray LR, Lee B, Jubb B, Westby M, Ramsland PA, Lewin SR, Payne RJ, Churchill MJ, Gorry PR.** 2013. A common mechanism of clinical HIV-1 resistance to the CCR5 antagonist maraviroc despite divergent resistance levels and lack of common gp120 resistance mutations. Retrovirology **10:**43.

48. **Gulick RM, Su Z, Flexner C, Hughes MD, Skolnik PR, Wilkin TJ, Gross R, Krambrink A, Coakley E, Greaves WL, Zolopa A, Reichman R, Godfrey C, Hirsch M, Kuritzkes DR, Team ACTG.** 2007. Phase 2 study of the safety and efficacy of vicriviroc, a CCR5 inhibitor, in HIV-1-Infected, treatment-experienced patients: AIDS clinical trials group 5211. J Infect Dis **196:**304-312.

49. **Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR.** 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One **4:**e5683.

50. **Kirchherr JL, Lu X, Kasongo W, Chalwe V, Mwananyanda L, Musonda RM, Xia SM, Scearce RM, Liao HX, Montefiori DC, Haynes BF, Gao F.** 2007. High throughput functional analysis of HIV-1 env genes without cloning. J Virol Methods **143:**104-111.

51. **Connor RI, Chen BK, Choe S, Landau NR.** 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology **206:**935-944.

52. **He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR.** 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol **69:**6705-6711.

53. **Pontow S, Ratner L.** 2001. Evidence for common structural determinants of human immunodeficiency virus type 1 coreceptor activity provided through functional analysis of CCR5/CXCR4 chimeric coreceptors. J Virol **75:**11503-11514.

54. **Sefik E, Geva-Zatorsky N, Oh S, Konnikova L, Zemmour D, McGuire AM, Burzyn D, Ortiz-Lopez A, Lobera M, Yang J, Ghosh S, Earl A, Snapper SB, Jupp R, Kasper D, Mathis D, Benoist C.** 2015. MUCOSAL IMMUNOLOGY. Individual intestinal symbionts induce a distinct population of RORgamma(+) regulatory T cells. Science **349:**993-997.

55. **Lewin A, Mayer M, Chusainow J, Jacob D, Appel B.** 2005. Viral promoters can initiate expression of toxin genes introduced into Escherichia coli. BMC Biotechnol **5:**19.

56. **Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M, Nussenzweig MC.** 2015. HIV-1 integration landscape during latent and active infection. Cell **160:**420-432.

Chapter 3

Deep Sequencing Analysis show Evolution of Coreceptor Utilization

to Escape CCR5 Antagonist Therapy

3.1 Preface

Chapter 3 is adapted from a manuscript submitted for publication on October 20th, 2015. The author contributions are the following: Lee Ratner and Jie Zhang conceived all the experiments and analysis; Jie Zhang and Xiang Gao generated all the plasmid libraries; Xiang Gao generated all the functional libraries; Jie Zhang and Zheng Chen developed approaches for sequencing data alignment, reads extraction, and entropy analysis; John Martin performed all the sequencing data quality control, alignment, and reads extraction. Jie Zhang analyzed all the processed data, and prepared all the tables and figures except for Figure 3.11.4 and Figure 3.11.5, which were prepared by John Martin and Bruce Rosa; Makedonka Mitreva provided intellectual suggestion and support for the sequencing process; Timothy Henrich and Daniel Kuritzkes kindly provided the clinical samples used in this study. The manuscript was written and revised by primarily by Jie Zhang and Lee Ratner; Zheng Chen, Makedonka Mitreva, and Daniel Kuritzkes provided useful comments.

Evolution of Coreceptor Utilization to Escape CCR5 Antagonist Therapy

Jie Zhang[a], Xiang Gao[a], John Martin[b], Bruce Rosa[b], Zheng Chen[c]*, Makedonka Mitreva[b], Timothy Henrich[d], Daniel Kuritzkes[d], Lee Ratner[a]#

Division of Molecular Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA[a]; The Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA[b]; Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri, USA[c]; Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, MA, USA[d]

Corresponding author: Lee Ratner

Department of Medicine, Washington University School of Medicine, 660 S. Euclid, Campus Box 8069, St. Louis, MO 63110, USA.

lratner@dom.wustl.edu

* Present address: Zheng Chen, Facebook, Inc., Menlo Park, CA, USA

Running head: non-V3 determinants of HIV-1 gp160 in CXCR4 tropism

3.2 Abstract

HIV-1 envelope interacts with coreceptors CCR5 and CXCR4 in a dynamic, multi-step fashion. Lack of detailed understanding of the molecular mechanism of envelope-coreceptor interactions hampers efforts to design more effective and precise entry inhibitors. Clinical application of CCR5 antagonists often results in tropism shift and subsequent therapeutic failure in patients. Using plasma from individuals whose virus escaped a CCR5 antagonist therapy, we constructed libraries of replication competent proviral clones that contain full-length patient-derived *env* amplicons. Tropism of quasispecies was determined by cell-based functional assays *in vitro*. Deep sequencing analysis revealed novel non-V3 region determinants that correlated with differential use of CCR5 and CXCR4. In particular, we identified important residues located on the gp120-gp41 interaction surface within one gp160 subunit, and on the gp41-gp41 contact surface between two neighboring gp160 subunits in a trimeric spike structure. Through in-depth analysis of functionally validated full-length envelope sequences, this study identifies important structure-function relationship within HIV-1 envelope and provides mechanistic insight into coreceptor specificity.

3.3 Importance

Preventing resistance development to current and future antiretroviral therapeutic agents is a top priority for global public health against HIV/AIDS. Utilizing next-generation sequencing approaches and phenotypic assays, we show a detailed picture of quasispecies evolution and treatment escape *in vivo* under the specific selection pressure of a CCR5 antagonist. We identified important regions in full-length gp160 that correlate with CXCR4 usage, in particular the regions at gp120-gp41 and gp41-gp41 interaction surfaces. This new line of evidence suggests a new hypothesis about coreceptor shift, and potentially has implications in coreceptor inhibitor design and tropism predictive algorithm development.

3.4 Introduction

Infection by human Immunodeficiency Virus Type 1 (HIV-1) is initiated by viral envelope trimer binding to the host cell receptor, CD4, and subsequently one of the two coreceptors C-C Chemokine Receptor Type 5 (CCR5) or C-X-C Chemokine Receptor Type 4 (CXCR4). A series of step-wise, dynamic conformational changes occurs following the two binding events which culminates in viral and cell membrane fusion, and release of the nucleocapsid core into the host cell (1, 2). The ability of a virion to bind either one or both chemokine receptors for entry has significant implications in disease pathogenesis, and affects treatment options (3). Although nearly all founder / transmitter viruses are strictly CCR5-tropic (termed R5 virus), close to 50% of HIV-1 subtype B patients will spontaneously develop CXCR4-using (X4) virus as the disease progresses, and the presence of X4 virus correlates with a worse clinical prognosis (4). This observation is thought to reflect the bottleneck nature of HIV-1 transmission, a preferential targeting of previously activated memory T lymphocytes in the early phase of infection, and the push to infect naïve T cells late in the disease course when the majority of memory T cells have been depleted (5). The process of coreceptor shift that otherwise naturally occurs with disease progression can take an expedited course upon the administration of a CCR5 antagonist antiretroviral drug (6). This limits the therapeutic application of CCR5 antagonists in patient management, and requires meticulous tropism determination prior to administration of the drug (7, 8).

Numerous studies examined the structural aspects of HIV-1envelope and coreceptor interaction (9-12). There has been a continuing effort to determine the viral envelope protein structure in its ligand-free, mature trimer configuration, as well as in complex with neutralizing antibodies (13-15). Recently, high-resolution crystal structures of CCR5 and CXCR4, both in

complexes with a small molecular ligand, have been also determined (16, 17). These findings shed light on the molecular interaction between the envelope and coreceptor. However, because the viral entry process is highly dynamic and involves large conformational changes of a complex macromolecular conglomeration, there are still large gaps in our understanding at a molecular level. The Variable loop 3 (V3) region has been shown to play a critical role in coreceptor specificity, by making direct contact with the extracellular loops of CCR5 or CXCR4 (18). Regions outside V3 have been found to affect tropism as well, but it is less clear how they affect the entry process (19-21).

Currently, tropism determination approaches can be broadly categorized as phenotypic or genotypic approaches. Phenotypic testing, such as the Trofile assay, is primarily based on *ex vivo* determination of tropism in established coreceptor expressing cell lines, and has been the gold standard for tropism determination (22). Recent advances in Next-Generation Sequencing (NGS) technologies have made genotypic tropism testing more feasible and potentially a cheaper, faster, and more accurate alternative to phenotypic methods (23-27). The current challenge with the wide application of genotypic tropism testing lies in the ability to accurately predict tropism based on amino acid sequences. Current algorithms, such as Geno2Pheno and PSSM, focus exclusively on the V3 region sequence for tropism prediction; and the outcomes, though accurate for CXCR4 utilization, have a high false positive rate, and limits the application of CCR5 antagonist drug to patients who could potentially benefit from the therapy (28-30). Therefore, it is important to better understand the molecular determinants for CXCR4 usage and the structure-function relationship of HIV-1envelope, thus enhancing the predictive power of the current methods (20, 31). In addition, new determinants critical for CXCR4 usage can be attractive drug targets for development of future entry inhibitors.

Given the challenge that there are a limited number of phenotypically validated envelope sequences, in which most of the available sequences include only the V3 region, we devised a novel comprehensive approach that combines phenotypic tropism determination with NGS technologies, and used it to study a panel of samples from patients who failed a CCR5 inhibitor clinical trial. By analyzing NGS data of phenotypically validated full-length HIV-1 gp160 sequences, we identified key determinants outside V3 that correlate with coreceptor use and CCR5 inhibitor resistance.

3.5 Materials and Methods

3.5.1 Patient samples

HIV-1 envelope amplicons were obtained from participants in a Vicriviroc (VCV) Phase II clinical trial (AIDS Clinical Trials Group [ACTG] A5211; NCT00082498) (32). Participant selection and sample preparation were described previously (27). Plasma samples were collected from study participants who experienced virologic failure and had tropism change by the phenotypic Trofile assay (Monogram Biosciences). Pelleted virion RNA was used to prepared *env* cDNA. Envelope amplicons were generated using previously reported primers (33):

Env1Atopo (5'-CACCGGCTTAGGCATCTCCTATGGCAGGAAGAA-3')

FLenv2.2 (5'-AGCTGGATCCGTCTCGAGATACTGCTCCCACCC-3')

3.5.2 Plasmid library generation

Patient envelope amplicons were cloned in frame into HIV-1 molecular clone pNL4-3.Luc.R$^-$.E$^-$ (Dr. Nathaniel Landau, the NIH AIDS Reagent Program) using Gibson Assembly® Master Mix (New England Biolabs) following the manufacturer's protocol (34, 35). An AfeI site (AGCGCT) that does not alter the amino acid sequence was introduced at nucleotide 5954 of pNL4-3.Luc.R$^-$.E$^-$ by site directed mutagenesis. A NotI site is present in the vector at the 4th codon *nef*, the position in which the luciferase gene was inserted. The vector was linearized using AfeI and NotI. The insert was prepared by PCR amplification of the patient envelope amplicons using the following primers:

NL4-AfeI-EnvF:

(5'-TTGTTTCATGACAAA AGCGCT AGGCATCTCCTATGGCAGGAAG -3')

NL4-NotI-EnvR:

(5'- TTTTGGCGTCTTCAGCGGCCGCGCCACCCATCTTATAGCAAAATCCTTTC -3').

The PCR was set up with Q5 High-Fidelity 2× Master Mix (New England Biolabs) following manufacturer's protocol, and run on BioRad T100 thermal cycler (BioRad) using the following conditions: 98°C for 30 sec; 15 cycles of 98°C for 10 sec, 55°C for 20 sec, and 72°C for 2 min; and 72°C for 5 min. The PCR product included flanking regions of pNL4-3.Luc.R^-.E^-. The Gibson assembly reaction was performed with an insert to vector ratio of 3:1 at 50 °C for 60 minutes. The assembly product was diluted 1:3 and electroporated into ElectroMAX™ Stbl4™ competent cells (Thermo Fisher) in 1 mm cuvette at 1.2 kV, 25 μF, 200 Ω mA on BioRad MicroPulser™ (BioRad). Transformants were recovered in 1 mL of S.O.C. medium by shaking at 30 °C for 90 minutes at 225 RPM, and plated on Luria-Broth agar plates supplemented with 150 μg/mL of ampicillin (Sigma-Aldrich).

Plasmid libraries of envelope amplicons were generated by harvesting 10^6 plate-grown transformant colonies and extracting plasmid DNA using QIAGEN Plasmid Plus Mega Kit (Qiagen). Transformed stbl4 cells were plated on 100 mm ampicillin selection plates at 10^4 colonies per plate. The colonies were incubated at 30 °C for 36 hours before being harvested using a scraper. The harvested bacterial colonies were centrifuged at 6,000 g at 4 °C for 15 minutes in JA-14 rotor in a J2-HS centrifuge (Beckman). Endotoxin-free plasmid DNA libraries were prepared according to the Qiagen kit protocol. Plasmid library DNA samples for Illumina sequencing were prepared by restriction enzyme digestion using NotI and AfeI, followed by agarose gel purification of the 3 kb band corresponding to the size of the insert envelope gene.

3.5.3 Cell lines

HEK 293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, and 1x antibacterial, antimycotic solution (containing 100 U/mL penicillin, 100 μg/mL streptomycin, and 250 ng/mL amphotericin B). U87.CD4 cells were stably transfected with pBABE-CCR5-GFP or pBABE-CXCR5-GFP, constructed as previously described, and maintained in DMEM supplemented with 15% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1x antibacterial, antimycotic solution, 0.2 mg/mL G418, and 1 μg/mL puromycin (12).

3.5.4 Virus production and passage

Replication-competent HIV-1 was produced in culture by transfecting 15 μg of the molecular clone plasmid DNA into 4×10^6 HEK 293T cells using *Trans*IT®-LT1 Transfection Reagent (Mirus). The viral supernatant was harvested at 48 to 72 hours post transfection and passed through a 0.44 μm syringe filter to remove cell debris. Freshly prepared viral stocks were passaged on U87.CD4.CCR5 and U87.CD4.CXCR4 cells at low multiplicity of infection (MOI equal to or less than 0.1), in the presence of 8 μg/mL DEAE dextran, as previously described (12). U87.CD4.CCR5 and U87.CD4.CXCR4 cells were incubated with viral supernatant at 37 °C, 5% CO_2, and 95% humidity for 12 hours. After removing the virus and applying fresh media, infection proceeded for another 24 hours to allow for a single cycle of virus replication. The viral supernatant was then harvested from the infected U87 cells, number of infectious particles quantified, and virus then used in the next round of passage.

3.5.5 HIV-1 luciferase reporter-based titration assay

Viral stocks harvested from 293T or U87 cells were quantified for infectivity on U87.CD4.CCR5 and U87.CD4.CXCR4 cells using serial dilutions in culture media from 5 to 5^{10} -fold (36). The diluted viral stocks were used to infect 10^4 U87 cells in 6 replicates in a 96-well plate in the presence of 8 μg/mL of DEAE dextran. Infection was carried out by incubating the cells with dilutions of viral stocks for 12 hours, removing the virus, and incubating for another 24 hours in fresh media. Cells were harvested 24 hours post-infection and lysed with 0.2% Triton-X 100 (Sigma-Aldrich) in PBS. The cell lysates were read for luciferase activity on Optocomp I luminometer (MGM Instruments). A result was scored as positive if the relative light unit readout was more than 2 standard deviations over the mean of mock infected control wells. The number of positive and negative wells were counted, and used to compute tissue culture infectious units using the Spearman-Karber formula.

3.5.6 Functional library DNA preparation

Functional libraries were prepared by PCR amplification of the envelope gene from integrated proviral DNA using genomic DNA of infected U87.CD4.CCR5 and U87.CD4.CXCR4 cells, harvested 24 hours post infection using the DNeasy Blood & Tissue Kits (Qiagen) according to the manufacturer's protocol for nucleated cells. To include maximum number of quasispecies and to minimize PCR introduced founder effects, the entire genomic DNA preparation was used in multiple independent PCR runs using NL4-AfeI-EnvF and NL4-NotI-EnvR primers, and the resultant PCR products were pooled. To minimize amplification bias introduced by varying template copy number, between 400-500 ng of genomic DNA was used as PCR template for each run, resulting in approximately 6,000 copies of proviral genome per

reaction, after adjusting for MOI and cell number. The final pooled PCR products were purified on agarose gels and quantified on TAKE3 plates using Synergy H4 Hybrid Multi-Mode Microplate Reader (Bio-Tek).

3.5.7 Next generation sequencing (NGS)

DNA samples were submitted to McDonnell Genome Institute at Washington University in St. Louis, MO for library construction and sequencing using HiSeq 2000 and/or Pacific Biosciences (Pac Bio) RS II. For amplicon and plasmid library sequencing with Illumina HiSeq2000, a minimum of 100 ng DNA per sample was used for library construction. The yield and size of input DNA was determined by a HS Qubit assay for quantitation (Invitrogen) and run on an Agilent Bioanalyzer 2100 (Agilent Technologies). Small insert dual indexed Illumina paired end libraries were constructed with the KAPA LTP Library Prep kit (KAPA Biosystems). 100ng of genomic DNA was fragmented using the Covaris LE220 DNA Sonicator (Covaris) to a size range between 200-800bp using the following settings: volume = 50µL, temperature = 4°C, duty factor = 15%, peak incident power = 450, cycle burst = 200, time = 130 seconds. The Illumina ligations were amplified in eight 50uL reactions and were amplified with 10 PCR cycles. Libraries were fractionated on the LabChip XT using the DNA 750 chip (Perkin Elmer) collecting a 575bp fraction with a +/- 5% covariance, followed by an AMPure XP bead purification to remove residual small fragments. Each fraction/library was assessed for concentration and size to determine molarity using the HS Qubit assay and the Agilent BioAnalyzer High Sensitivity DNA Assay. The concentration of each library fraction was verified through qPCR according to the manufacturer's protocol (Kapa Biosystems) to produce cluster counts appropriate for the Illumina HiSeq2000 platform. The libraries were pooled in

equal molar ratios and loaded on 1 lane of the HiSeq2000 platform utilizing a 2x101bp recipe according to the manufacturer's recommendations (Illumina).

Functional libraries of 3 kb in length were sequenced with HiSeq2000 as well as Pac Bio RS II. A minimum of 2 μg of functional library *env* PCR product per sample was submitted for Pac Bio library construction using SMRTBell™ Template Preparation Kit (Pacific Biosciences) following the manufacturer's protocol. Library construction input was 750ng per library. Each sample was run on a single SMRT cell on the PacBio RS II platform using P6v2/C4 chemistry and 240 minute movie length.

3.5.8 NGS data analysis

Illumina datasets were quality controlled by performing adaptor trimming, quality trimming, complexity screening, and length filtering using in-house codes. The amplicon and plasmid library Illumina datasets were mapped, using BWA (37) or Bowtie2 (38), to sample-specific full-length reference sequences, which were pre-determined by Sanger-sequencing of single clones. Pac Bio datasets of the functional libraries were quality trimmed using the SMRT Portal system to extract the high-quality Reads-of-Inserts (ROIs). The ROIs were aligned to sample-specific references using BWA-MEM. Reads that span the entire V3 region were extracted and numerated, assuming every individual read represents a single DNA molecule in the sequencing process. Rare sequences that were occurring at single-digit read level, visibly different from V3 and likely arising from mapping error and frame-shift translation were manually removed. Single Nucleotide Polymorphism (SNP) analysis was performed using GATK Haplotype Caller (https://www.broadinstitute.org/gatk/).

3.5.9 Phylogenetic and statistical analysis

Hierarchical clustering was performed based on Euclidean distance matrices and visualized using GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/) or iTOF (http://itol.embl.de/). V3 loop consensus sequences from two functional clusterings and their significantly different amino acid positions were computed and visualized using IceLogo (http://iomics.ugent.be/icelogoserver/index.html). The statistical significance of the differences in entropy at every nucleotide position between two samples were determined by Student's t-test.

3.5.10 Deep Sequence Data Sets

Raw sequence files from this study will be deposited in NCBI Sequence Read Archive (SRA) upon manuscript acceptance. Accession numbers is TBD.

3.6 Results

3.6.1 Evaluation of HIV-1 patient samples by deep sequencing

HIV-1 patient samples were collected from eight study subjects at two time points on ACTG A5211 Phase II study of vicriviroc (VCV) (Table 1). Based on the original Tropile *in vitro* phenotypic assay, most study participants (all but Subject 3 and 6) harbored only R5 viruses before VCV treatment (Time point 1 at Week 0). At the second time point where the subject experienced protocol-defined virologic failure (less than 1 log10 decrease of HIV-1 RNA level at or after Week 16), plasma *env* amplicons demonstrated the ability to utilize CXCR4 for entry, and their tropism was designated as dual-mixed (DM). The time elapsed between the first and second time points varied among the study subjects from 2 to 32 weeks.

To understand HIV-1 *env* sequence characteristics with respect to their function, we devised a novel experimental scheme that allowed high-throughput phenotypic tropism determination followed by deep sequencing of functionally validated *env* libraries (Fig 1A). By generating a heterogeneous library of replication competent HIV-1 virions that contained a diverse set of *env* quasispecies from patient samples, we physically separated the *env* quasispecies based on their tropism by passaging the virus on CCR5- or CXCR4-expressing U87.CD4 cells. The subset of *env* variants extracted from the proviral DNA in the infected CCR5- or CXCR4-expressing cells was termed a functional library, and their sequences were determined on both Illumina and PacBio platforms. To confirm that the sequence heterogeneity was not lost during the process of molecular cloning, we analyzed the original patient amplicon samples (named amplicon libraries) and the intermediate plasmid library containing one million colonies using Illumina sequencing.

A clustering analysis using single nucleotide polymorphism (SNP) information of the sixteen amplicon libraries positioned the pair of samples from each study subject on a single branch of a Euclidean-distance tree (Fig 1B). The clustering pattern was consistent with the known sample identities, and confirmed that deep sequencing data can be used to accurately categorize HIV-1 *env* samples based on their sequence signatures.

3.6.2 Sequence diversity of patient amplicon libraries over time

We first analyzed the Variable loop 3 (V3) coding sequence, as it is the primary determinant of coreceptor use. We found the patient samples were diverse and heterogeneous, to a similar degree as previous studies had reported, by counting the number and frequencies of different variants in the sample (24, 27). An in-depth population level variant analysis was performed by extracting and enumerating reads that span the entire 33 codons between the two cysteine residues at the beginning and the end of V3. The assumption was that one read comes from one DNA molecule, which comes from one quasispecies in the infected individual. Thus, by enumerating the non-redundant reads, we captured a snapshot of the quasispecies landscape in the patient at the time of sampling (27). To control for PCR introduced mutations and sequence errors, we performed PCR amplification and deep sequencing of a clonal control *env* (data not shown). We detected 233 distinct non-redundant variants out of 31,989 extracted V3 reads, with the most abundant variant that was identical to the pre-determined sequence occurring 30,967 times (96.8%). None of the remaining 232 variants exceeded 0.05% of the total population. Therefore, we concluded that the error rate was low and applied a cut-off filter at 0.05% level to the sixteen amplicon library datasets to remove possible artifact reads.

From the sixteen samples coming from eight study subjects, we selected four samples from two subjects (Samples 1 and 2, Table 1) for the population level variant analysis (Figure 2). We chose to focus on these two subjects because their samples demonstrated different lengths of time for coreceptor shift. Subject 1 showed high R5-tropic viral activity at baseline and comparable R5- and X4-tropic activity in only two weeks, whereas samples from Subject 2 remained predominantly R5-tropic at Week 0 and Week 8, but also showed significantly increased X4-tropism at the second time point, based on the luciferase assay readings (Table 1). Examination of the variant profile revealed distinct quasispecies compositions in these two subjects (Figure 2A-D).

The viral population in Subject 1 before VCV treatment existed as one dominant variant that accounted for 62.6% of the total population (Figure 2A Variant 1), a less dominant variant at 17.2% level (Figure 2A Variant 2), and numerous rare variants that occurred in single-digit counts and less than 1% of all V3 reads. On the other hand, the second time point sample (1.2) contained 2 major variants, at 46.1% and 38.8% abundance, respectively (Figure 2B Variant 1 and 2), followed by a minor variant at 1.4% level, and rare variants at less than 1% level. Contrary to previous studies and our expectation that a pre-existing CXCR4-using variant at Week 0 expanded under drug selection to become a major variant at Week 2, we did not detect the major variants from Sample 1.2 (Figure 2B Variant 1 and 2) in Sample 1.1. Instead, a completely different and CXCR4-using population emerged and rapidly took over in two weeks, resulting in escape to the VCV therapy and virologic failure in Subject 1.

Different from Subject 1, Subject 2 showed dynamic expansion and contraction of variants over the 8 weeks of VCV treatment. At Week 0, the subject harbored multiple major variants, each at 47.6%, 20.6%, and 8.1% of the total, respectively (Figure 2C Variant 1-3), all of

which were predicted to use CCR5 exclusively for viral entry based on the Geno2Pheno$_{[coreceptor]}$ predictive algorithm. Some minor R5-tropic variants from Sample 2.1, present at abundances of 5.9%, 4.7% and 1.4% (Figure 2C Variant 4-6), respectively, persisted and expanded through the treatment and were detected again at Week 8 at relative abundances of 15.7%, 20.9%, and 9.9% (Figure 2D Variant 3, 2, and 4). The most abundant variant in Sample 2.2, accounting for 25.7% of total reads (Figure 2D Variant 1), arose from a rare variant at 0.47% level in Sample 2.1. In contrast, the two most abundant variants present at 47.6% and 20.6% frequency in Sample 2.1 were completely eliminated from the circulation. Overall, Subject 2 largely retained its population diversity over 8 weeks, with expansion of pre-existing CXCR4-using variants and maintenance of several CCR5-only variants in face of VCV treatment.

3.6.3 Evaluation of plasmid and functional library variants

To phenotypically test a large number of *env* quasispecies, we generated plasmid libraries of complete NL4.3.Luc+ molecular clones containing *env* sequences from the two samples of Subject 1. The plasmid libraries were sequenced with Illumina HiSeq to ensure that the original quasispecies diversity was retained by the cloning process. We conducted the population level variant analysis on the plasmid library sequencing datasets and discovered almost identical variant distributions, with a maximal change of major variant abundance at 3.3% of the total V3 reads (Figure 2E compared to 2A, 2F compared to 2B, and 2G).

Functional libraries were generated by passaging plasmid libraries 1.1 and 1.2, through U87.CD4.CCR5 or U87.CD4.CXCR4 cells at low MOI (≤0.1) for two or three passages each consisting of a single-cycle infection. A function library control was also generating by passage of the single *env* control in parallel. The four functional libraries obtained through this procedure

included two independent passages of Sample 1.1 on CCR5-expressing cells (thus named 1.1_R5_FL1 and 1.1_R5_FL2), passage of Sample 1.2 on CXCR4-expressing cells (named 1.2_X4_FL), and a control passage of the single clone isolate from Sample 1.2 on CXCR4-expressing cells (named 1.2_Ctrl_X4). The functional libraries were sequenced by PacBio RSII in addition to Illumina, so that the linkage information between SNPs longer than 100 bases apart could be studied.

High-quality Reads-of-Inserts (ROIs) of the 4 functional libraries from the PacBio platform were generated. A total of 26,046 ROIs were obtained passed a quality filter of 99% per base accuracy, including 4667 ROIs from 1.1_R5_FL1, 8707 ROIs from 1.1_R5_FL2, 6783 ROIs from 1.2_X4_FL, and 5889 ROIs from 1.2_Ctrl_X4. The average length for ROIs is 2260 bp, with the majority of ROI at 2.9 kb, consistent with the input fragment size. All ROIs were generated from circular consensus sequencing of 10 passes or more, resulting in low error rate per base and high quality output sequencing reads.

Analysis of variants at population level in the four functional libraries was performed to understand the quasispecies profile across the envelope variable regions (Figure 3). Insufficient reads from the V1 region were extracted to permit meaningful analysis, likely due to the incompatibility between our stringent read extraction criteria and the presence of large insertion and deletions in this region. Therefore, we focused on the V2 through V5 regions instead. To control for the errors introduced by PCR, virus passaging, and sequencing procedures, we analyzed the clonal control 1.2_Ctrl_X4 functional library at V2 through V5, and detected minimal amounts of deviant variants across the five regions (data not shown). Overall, in the three remaining libraries, the analysis revealed distinct quasispecies landscapes for R5 versus X4 variants. The two biological replicates, 1.1_R5_FL1 and 1.1_R5_FL2, were highly consistent in

terms of the major variants and their abundances across V3 – V5 regions, confirming that the passaging method was a reliable and reproducible approach to capture functionally validated quasispecies (Figure 3, Columns 1 and 2). Comparing the variant composition across the variable regions, the variants from V2 through V4 were more closely correlated, consistent with one major and one minor clone in 1.1_R5_FL1 and FL2, and two major clones in 1.2_X4_FL1. The V5 region appeared more variable in R5-using populations than in X4-using populations, and showed less correlation with the other three regions.

Examination of the quasispecies profiles in the V3 region of the functional libraries show that they were highly similar to those of the starting amplicon libraries (Figure 2A-B, Figure 3). The most abundant variants in Sample 1.1 and 1.2 amplicon libraries were present in the functional libraries, suggesting the V3 variants observed in the amplicon libraries were part of full-length, functional *env* quasispecies. The two most abundant variants in Sample 1.1 (Figure 2A Variant 1 and 2), as well as the three most abundant variants in Sample 1.2 (Figure 2B Variant 1-3) were detected at comparable proportions in the functional libraries (Figure 3, Row 2). Interestingly, a rare variant in the original amplicon library of Sample 1.1 at only 0.13% level (Figure 2A and B) expanded through the passaging to become the third most abundant variant in 1.1_R5_FL1 at 3.5%, and the second most abundant variant in 1.1_R5_FL2 at 7.1% (Figure 3, Row 2). This suggested that the enrichment and contraction of variants in the functional library were not stochastic events. The ability of certain variants but not the others to persist through the passaging process was directly dependent on their capacity to facilitate cell entry using CD4 and one of the two coreceptors at a given CD4 and coreceptor concentration of the reporter cell lines.

3.6.4 Characterization of V3 variants based on sequence signature shows consistency with functional tropism

To dissect the sequence signature of HIV-1 *env* isolates and its relationship with the coreceptor binding function, we generated a hierarchical cluster, based on a Euclidean-distance matrix, using 167 V3 variants from all 4 functional libraries (Figure 4). The sequence-based clustering separated the variants into two groups that coincided almost perfectly with their function, except for three variants from CCR5 functional libraries that were grouped with variants from CXCR4 libraries (Figure 4, circled in green). These three variants were predicted by the Geno2Pheno algorithm to be able to also use CXCR4 for entry, suggesting that these were rare variants before treatment that could utilize CXCR4 for entry but did not expand upon VCV treatment likely due to low fitness.

We generated consensus sequences of all the variants in each cluster with IceLogo2, taking into consideration their respectively abundance, and identified amino acid positions differentiating cluster 1 and cluster 2 at a significance level of $P < 0.001$ (Figure 5). The variants in CCR5 functional libraries were highly similar to each other based on the Logo plots, and were distinct from those in CXCR4 functional libraries. Cluster 1 showed a high degree of conservation, which was reflected by a major variant that accounted for over 80% of the population. Cluster 2 showed a clear divergence at Position 31, which was representative of two major variants at 43.5% and 50.2%, respectively (Figure 5A). Altogether, we identified nine positions (Positions 2, 20, 22, 23, 24, 25, 31, 32, and 34) in the V3 region with significantly different amino acid residues, that were likely responsible for the differential functionality to engage CCR5 versus CXCR4 as entry coreceptor (Figure 5B). Consistent with our observation, these positions and amino acid residues have been previously reported in HIV-1 Subtype B and

C patient samples to correlate with CCR5 or dual-tropism (39, 40). In addition, at the 25th position in Cluster 1 was a negatively charged glutamic acid, whereas occupying the same position in Cluster 2 was a positively charged lysine, which was indicative of CXCR4-usage by the 11/25 Rule (41). Through the clustering and consensus analysis, we demonstrated that categorizing HIV-1 V3 variants by sequence similarity created clusters that were consistent with their function, and confirmed that the V3 region is a critical modality for coreceptor tropism determination.

3.6.5 Important regions outside V3 that correlate with envelope tropism

We analyzed the full *env* gene to examine other determinants of tropism beyond the V3 region using Shannon's entropy approach and SNP analysis. Shannon's entropy analysis examines cross-sectional diversity at a given position, with higher entropy values responding to more diversity. We utilized entropy analysis to compare the entropy values at each nucleotide position between amplicon libraries of Sample 1.1 and 1.2 to explore regions that were undergoing an increase or decrease in diversity due to VCV selection.

Overall, Sample 1.1 had higher entropy across the length of the *env* gene than Sample 1.2, as demonstrated by 120 positions in Sample 1.1 and only 50 positions in Sample 1.2 that have statistically above-background entropy values when compared with the clonal control dataset (Figure 6A, $p<0.0001$). This suggested that the viral quasispecies population before VCV treatment was more heterogeneous than the quasispecies present after treatment. The treatment likely selected for a dominant drug-resistant variant, or a group of closely related resistant variants, thus, driving down the overall diversity in the population. The decrease in entropy from

1.1 to 1.2 occurred mostly in the V1, C2, and C5 regions, as represented by the entropy difference map (Figure 6B). However, specific locations showed an increase in entropy, in particular at HR1/HR2 regions and the cytoplasmic domain of gp41, suggesting diversifying selection at those loci (Table 2).

SNP analysis on the four functional library datasets using GATK program identified 106 SNPs in the entire *env* gene with reference to the control single clone sequence from Sample 1.2. Clustering of the four functional library datasets based on their SNP frequency patterns reflects the functional categorization of the four libraries, with 1.1_R5_FL1 and 1.1_R5_FL2 on a single branch, and away from 1.2_X4_FL and 1.2_Ctrl_X4 on a different branch (Figure 7A). To identify SNPs that could differentiate CCR5 versus CXCR4 coreceptor usage, we focused on 23 non-synonymous loci (annotated in the red boxes) that were completely different in CCR5-using groups than in CXCR4-using groups, including 22 loci outside the V3 region. Among the 23 positions identified, 4 positions (N325, K421, I424, and S440) were previously known to strongly affect coreceptor binding and specificity, and 3 positions (N425, K432, and R476) were reported to be CD4 binding residues (13, 42, 43). Sixteen positions identified through the SNP analysis have not previously been identified to have a function in receptor or coreceptor binding, although many were found to be involved with gp120 – gp41 interaction (Table 2) (42-45). This finding suggests a role of gp120-gp41 interaction in coreceptor specificity, potentially by affecting entry dynamics and thus fusion efficiency following coreceptor binding.

For visualization of the critical regions that correlated with coreceptor specificity, annotation was performed on the ligand-free native envelope gp140 trimer crystal structure (PDB ID: 4MJZ) (Figure 7B and 7C). A group of residues (colored red) that were previously reported to be coreceptor-specific is located near the envelope-coreceptor binding interface and likely to

affect binding via direct contact with extracellular regions of the coreceptors. A second group of residues (colored magenta) located near the CD4 binding site are likely to affect coreceptor usage via interaction with CD4 and subsequent induction of a conformational change into an open trimer. A third group (colored blue and green) that is located at gp120-gp41 contact interface and the interface between the three subunits is likely to affect tropism via global mechanisms, such as alteration of the overall configuration of envelope trimer or its dynamic conformation change upon CD4 and/or coreceptor engagement.

3.7 Discussion

In this study, we developed a comprehensive phenotypic assay platform in combination with NGS technologies to analyze HIV-1 quasispecies diversity and tropism before and after emergence of drug resistance *in vivo*. This novel approach employed a tissue-culture based library passaging method and allowed us to study in unprecedented depth and breadth the quasispecies population and their structure-function relationship by generating sequences from phenotypically validated *env* variants. The output data provided new insights into important features of HIV-1 env that are implicated in coreceptor interaction, and could shed light on new drug targets for HIV-1 entry inhibitor development. In addition, the experimental approach can potentially be adapted for the study of drug resistance development in other clinical situations, especially for development of unusual combinations of drug resistant mutations that requires phenotypic characterization of new and complex genotypes.

Quality control of the sequencing data is extremely important in NGS-based studies to ensure the accuracy and relevance of the final conclusion. In our study, we performed all the experimental procedures including PCR, cloning, infection, passaging, Illumina and PacBio sequencing, reads filtering, and the final analysis with a clonal control of known nucleotide sequence in parallel. Therefore, we could calculate the cumulative error rate of PCR and cloning errors, passaging artifacts, and sequencing errors at every step of the process by analyzing the single clone control dataset and apply an appropriate filter level based on the control error rate. Notably, the combined error rate from PCR and sequencing in our experience were markedly lower than previously reported, likely reflecting improved accuracy of the higher-fidelity PCR polymerase and of the newer generation sequencing method (27). A potential source of error that we could not control for using the single clone method is strand-switching (recombination)

during the PCR process. Based on previously published information, the recombined clones accounted for equal to or less than 0.15% in the total population, and did not significantly affect the proportions of truly existing variants (27). In addition, we also utilized the amplicon and plasmid library dataset as baseline for the functional library variant analysis, and removed variants that did not exist in the first two libraries, therefore eliminating the majority of recombination errors and passaging artifacts. Biological replicates of Sample1.1 on CCR5-expressing cell lines served as another layer of control and confirmed that our approach is robust and reproducible.

The quasispecies landscape at the two snapshots through time in two study subjects alluded to different evolutionary pathways. Through analysis of V3 variants in amplicon libraries from Subject 1, we were unable to detect a minor variant at Week 0 that expanded into a major variant at Week 2 despite multiple rounds of sequencing and extensive depths of coverage at 600× to 6000×. Though a minute possibility remains that the major CXCR4-using variants in 1.2 were present in1.1 at a level too low to allow detection, a more plausible hypothesis is that the virus had undergone *de novo* mutations that conferred CXCR4-utilization, and achieved a complete population shift and coreceptor switch within 2 weeks. Consistent with the "fitness valley" model of coreceptor shift, we detected few intermediate variants from CCR5-only to CXCR4-using genotypes, as they are thought to be less fit and are quickly selected out (46, 47). On the contrary, Subject 2 demonstrated a typical dynamic quasispecies population shift under the selection pressure from the CCR5 antagonist, during which rare pre-existing CXCR4-using clones expanded and susceptible clones declined. Interestingly, several major variants predicted to be R5-tropic only in Sample 2.1 had unexpectedly persisted and expanded over treatment,

perhaps by acquiring resistance through an alternative mechanism, such as binding to the CCR5-VCV complex for entry (6).

One limitation of this study is the sampling method. We were unable to access quasispecies residing in lymph nodes and latently infected cells, as the patient samples were collected and prepared from free circulating virions in the plasma. Since HIV-1 quasispecies at these sanctuary sites are often sheltered from antiretroviral therapy, or are exposed to a lower dose and sub-optimal combination of antiretroviral drugs, the virus may continue to replicate at low levels and develop resistance progressively in a step-wise fashion, even when the treatment appears fully suppressive (48). Some of the viral quasispecies that eventually escape treatment and migrate to the blood are predicted to harbor multiple mutations, consistent with our observations. A recent study also shows that proviral populations in peripheral blood mononuclear cells are different from those in lymphoid organs and in latently infected memory cells; the variants in circulation are often defective or less fit (49). Since we are just beginning to understand these tissue-localized HIV quasispecies, more studies are needed to examine the evolutionary process in local lymphoid tissues and latently infected reservoirs to elucidate how they contribute to coreceptor shift and resistance development.

Finally, this study has identified three regions outside V3 that strongly associate with coreceptor shift in the context of VCV selection, including the C4/V4/C5 region, CD4 binding site, and the interaction surface between gp120-gp41 and between gp41 subunits. A majority of the residues identified in the first two regions are consistent with and confirm previously published data on coreceptor specificity and viral entry. In contrast, a new subset of mutations at the gp120-gp41 interaction site alludes to interesting new possibilities for the mechanism of coreceptor shift. Since the gp120-gp41 interaction surface is located away from gp120-

coreceptor interaction surface, mutations in this region are unlikely to directly affect gp120 binding to coreceptor by charge or binding surface steric hindrance. They may, however, affect the trimer conformation and alter the affinity of envelope trimer to one coreceptor versus the other. These mutations may also compensate for the decrease of viral fitness from changes in other regions of *env* and promote viral entry. The new line of evidence demonstrated in this study contributes to our understanding of the dynamic macromolecular interactions during HIV-1 entry, warrants further investigation via mutagenesis studies into the coreceptor shift process, and potentially sheds light on a new therapeutic target for entry inhibitors.

3.8 Funding sources

This work was supported by National Institutes of Health grant R01 AI106361 to L.R., UM1 AI068636 and R37 AI055357 to D.R.K.

3.9 Acknowledgements

We thank Dr H. Huang for assistance with PyMOL, Dr J. Fay for guidance in dN/dS analysis, Dr K. Wylie and Dr J. Gu for assistance with NGS data analysis, Drs. Kyei, Rauch, McCulley, and Leibel for intellectual discussions, Dr D. Wang and Dr. G. Zhao for diversity analysis, and Ms N. Campbell and Mr J. Zhang for technical assistance. We also thank Dr. Roy M. Gulick and members of the ACTG A5211 protocol team, as well the participating ACTG Clinical Research Sites, for providing the plasma samples used in this study.

3.10 Tables

Table 3.10.1 Longitudinal changes of HIV-1 envelope tropism in VCV treated patients

*Subject ID	Sample ID	Time (weeks)	R5-RLU**	X4-RLU**	Tropism***
1	1.1	0	593008	86	R5
	1.2	2	139491	185312	DM
2	2.1	0	7360	129	R5
	2.2	8	6260	2034	DM
3	3.1	0	774210	388	DM****
	3.2	24	223785	1178193	DM
4	4.1	0	434886	76	R5
	4.2	32	6627	117574	DM
5	5.1	0	37210	66	R5
	5.2	8	50461	78889	DM
6	6.1	0	511349	1920	DM
	6.2	8	124983	173363	DM
7	7.1	0	90578	63	R5
	7.2	8	8211	9836	DM
8	8.1	0	391164	73	R5
	8.2	2	1789	151040	DM

* Subject: HIV-1 patients who experienced virological failure (defined as smaller than 1 $\log_{10}$ viral load decrease at or after week 16 by protocol) in a Vicriviroc phase IIb clinical trial.

** RLU: Relative Light Units by luciferase-based Trofile® assay on CCR5 (R5) or CXCR4 (X4) expressing cell lines.

*** Tropism: the ability of the virus to utilize R5 and/or X4 for viral entry.

**** DM: dual/mixed tropism. Dual tropism: the ability to utilize both R5 and X4 for entry; mixed tropism: coexistence of viral species that can use either R5 or X4 coreceptor for entry.

Table 3.10.2 Positions within and outside the V3 region that correlate with coreceptor usage

	Position *	SNP +	ΔEntropy		Significance	Citation
			value**	p value++		
SP	24		-0.955	5.62E-05		
	26		-0.901	2.99E-05		
	29		0.988	2.43E-05		
C1	86	L/M			gp120-gp41 within one protomer	47
	87	E/G			gp120-gp41 interaction surface within single protomer; N88 glycosylation	
V1	134a		-0.956	5.73E-05		
C2	240		-0.716	1.98E-05		
	256		-0.757	1.38E-05		
V3	325	D/N			Coreceptor binding site within V3	44
	327		0.988	6.62E-05	Coreceptor binding site within V3	
C3	351	E/K	-0.714	6.36E-04		
V4	412	E/D				
C4	421	R/K			Coreceptor binding site outside V3	44, 13
	424	V/I			Coreceptor specific site (R5X4)	
	425	N/K			CD4 contact site, main & side chain	45
	432	K/T			CD4 contact site main chain only	
	440	N/S			Coreceptor binding outside V3; R5/X4	44
	442		-0.871	3.03E-05	Coreceptor binding site outside V3	
	444	S/R			Cys 378 linked to Cys 445	
V5	465d	R/K	-0.931	8.74E-04		
C5	476	K/R			CD4 contact site, side chain only	45
	490	Q/K				
	496	L/V	-0.707	7.71E-05	gp120-gp41 interaction surface within single protomer; RRE stem I start	46, 47
	500	K/M	-0.918	4.28E-05		
	503		-0.757	5.76E-05		
FP	514a	M/T/I				
	518	V/M				
	525	T/A			gp120-gp41 interaction surface within single protomer	47
FPPR	534	A/S				
	535		-0.819	4.70E-05	gp41-gp41 interaction site & surface between two protomers	
	541	T/A				
	543	R/Q			gp120-gp41 interaction surface within single protomer	44, 47
HR1	555		0.495	7.18E-04		
	596		-0.989	1.23E-05		
	621	E/D	-0.239	1.57E-04		
HR2	633		0.311	9.52E-05	gp120-gp41 interaction surface within single protomer	47
CD	724		0.983	6.38E-05		
	749		0.991	2.72E-05		
	802		-0.498	7.12E-05		

* Positions are numbered based on alignment to HXB2.

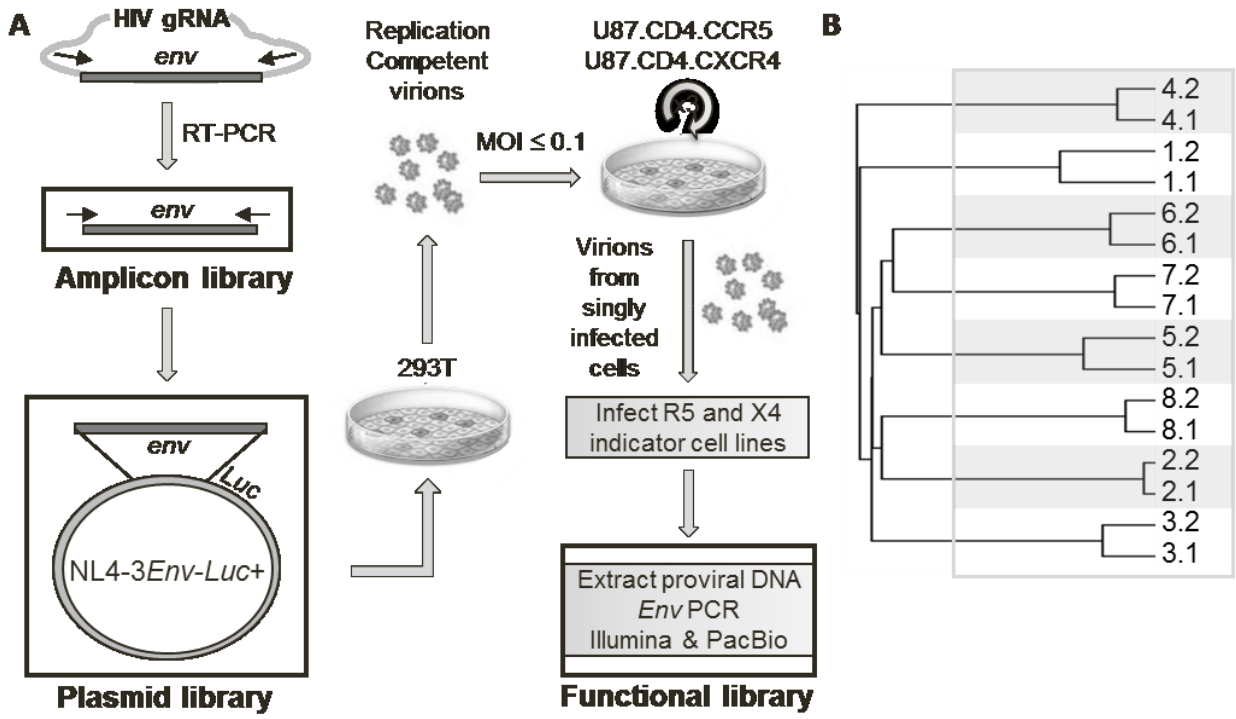
[+] Positions showing > 70% switch from reference to alternative allele in SNP analysis are annotated as ref/alt/alt. Reference allele is based on 1.2 single clone control sequence.

** ΔEntropy value = entropy value of Sample 1.2 AL – entropy value of Sample 1.1 AL Positive value: increase in entropy/diversity; negative value: decrease in entropy.

++ Positions with a significant change of entropy have p values <0.0001, by two-tailed t-test

3.11 Figures

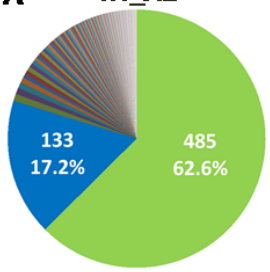
Figure 3.11.1 Experimental design and clustering of 16 samples



(A) Schematic diagram of the experimental design. Patient derived envelope quasispecies amplicons (amplicon libraries) were cloned into replication competent NL4.3*Env-luc+* reporter vector (plasmid libraries) and expressed in 293T cells as replication competent virions that were used to infect CCR5 and CXCR4 expressing cell lines. The functionally validated *env* quasispecies (functional libraries) were prepared by polymerase-chain reaction of the proviral DNA using *env*-specific primers from the genomic DNA of the infected cells. The three sets of libraries were sequenced by Illumina HiSeq2000; functional libraries were also sequenced by Pacific Biosciences RS II. **(B)** SNP based clustering analysis of 16 amplicon libraries based on Illumina sequencing of the full-length *env* gene. SNPs were called with HXB2 *env* as the reference sequence using GATK. A Euclidean-distance based clustering of 16 amplicon libraries was constructed using 843 SNPs present in all samples.

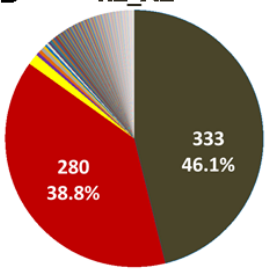
Figure 3.11.2 Population analysis of V3 loop sequences and frequencies in amplicon and plasmid libraries.

A 1.1_AL



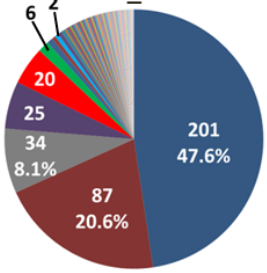
Variant #	V3 sequences	# of times	Abundance (%)	Use X4? Geno2Pheno
Sample 1.1: 93 unique V3 reads out of 775 total V3 reads				
1	CIRPNNNTRKGIHIGPGRAFYATGEIIGDIRQAHC	485	62.58	N
2	----------G----------	133	17.16	N
3	----------E----------	7	0.90	N
4	--G----------	6	0.77	N
5	----D----------	6	0.77	N

B 1.2_AL



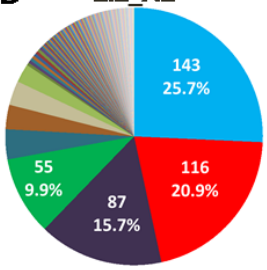
Sample 1.2: 88 unique V3 reads out of 722 total V3 reads				
1	CTRPNNNTRKGIHIGPGRAVYVAEKIIGDIRKAYC	333	46.12	Y
2	----------K----	280	38.78	Y
3	--G----------	10	1.39	Y
4	----------S----------	3	0.42	Y
5	-----S----------	3	0.42	Y
6	-------A----------	3	0.42	Y

C 2.1_AL



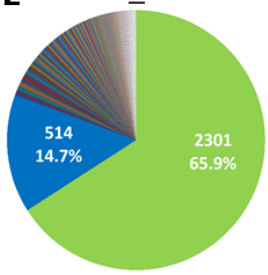
Sample 2.1 AL: 84 unique V3 reads out of 422 total V3 reads				
1	CVRPNNNTRRSINMGPGRVFHTSEIIGDIRQAHC	201	47.63	N
2	---------K----------	87	20.62	N
3	---------K----------N------	34	8.06	N
4	----S----------	25	5.92	N
5	----S----K----------N------	20	4.74	N
6	----S----------N------	6	1.42	N
7	----------N------	3	0.71	N

D 2.2_AL

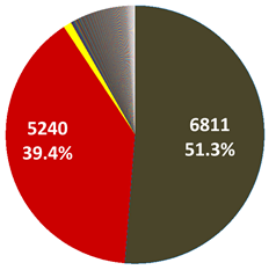


Sample 2.2 AL: 111 unique V3 reads out of 556 total V3 reads				
1	CVRPNNNTSKRVGIGPGMAFRATRIIGDIRQAHC	143	25.72	Y
2	----S---R-SINM---RV-HTSE---N------	116	20.86	N
3	----------A----------	87	15.65	Y
4	----S---RRSINM---RV-HTSE---N------	55	9.89	N
5	--------R-SINM---RV-HTSE----V------	21	3.78	5.4%*
6	----S---R-SINM---RV-HTSE----------	17	3.06	N
7	----------K----------	17	3.06	Y
8	--------RRSINM---RV-HTSE---N------	13	2.34	N

E 1.1_PL



F 1.2_PL

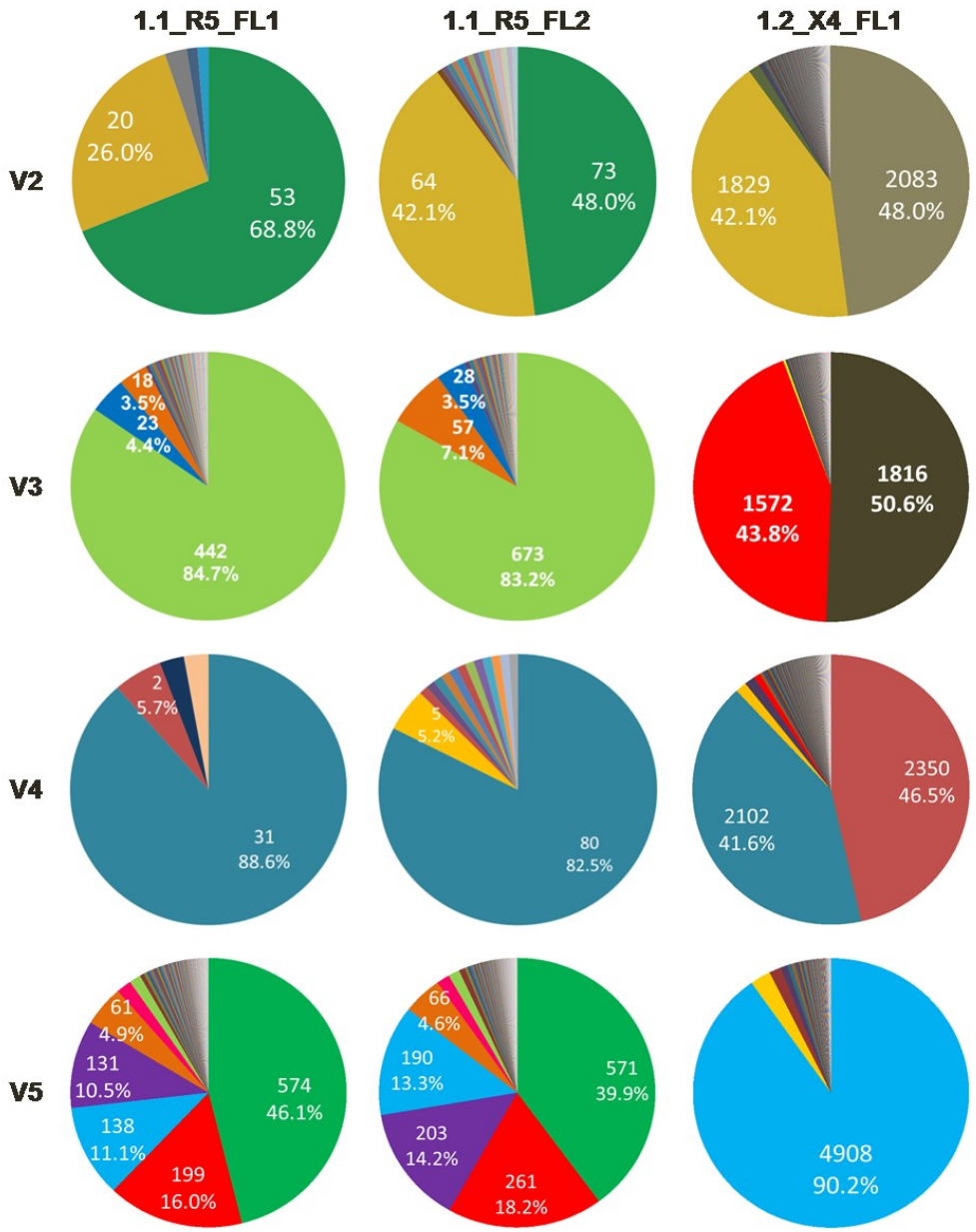


G

Sample ID	Library type	Total V3 reads	Unique V3 reads
1.1	Amplicon	775	93
	Plasmid	3490	215
1.2	Amplicon	6256	236
	Plasmid	13284	329
	Clonal ctrl	31138	8

(A) Sample 1.1 amplicon library (AL), **(B)** Sample 1.2 AL, **(C)** Sample 2.1 AL, **(D)** Sample 2.2 AL, **(E)** Sample 1.1 plasmid library (PL), and **(F)** Sample 1.2 PL. **(G)** A summary table compares amplicon libraries versus plasmid libraries for Sample 1.1 and 1.2. Total V3 reads are representative of the depth of sequencing. A clonal control *env* from Sample 1.2 was prepared and sequenced in parallel. For amplicon and plasmid libraries in (A) – (F), unique V3 variants and their respective frequencies are represented as slices on a pie chart. The same color in different libraries represents a common V3 variant found in both libraries. For amplicon libraries in (A) – (D), the sequences and proportions of the most abundant V3 loop variants are listed. The coreceptor usage predicted by Geno2Pheno[corecptor] is shown for every sequence. The significance level of the prediction was set at 2% and 5.75% False-Positive Rate (FPR) as the optimized cut-offs based on clinical data from the MOTIVATE study (FPR<0.2%: X4-capable; FPR>5.75%, X4-incapable.) *For the 5th most abundant variant in Sample 2.2, the V3 sequence is predicted to be X4-capable with a FPR of 5.4%, which falls in between the optimized cut-off values.

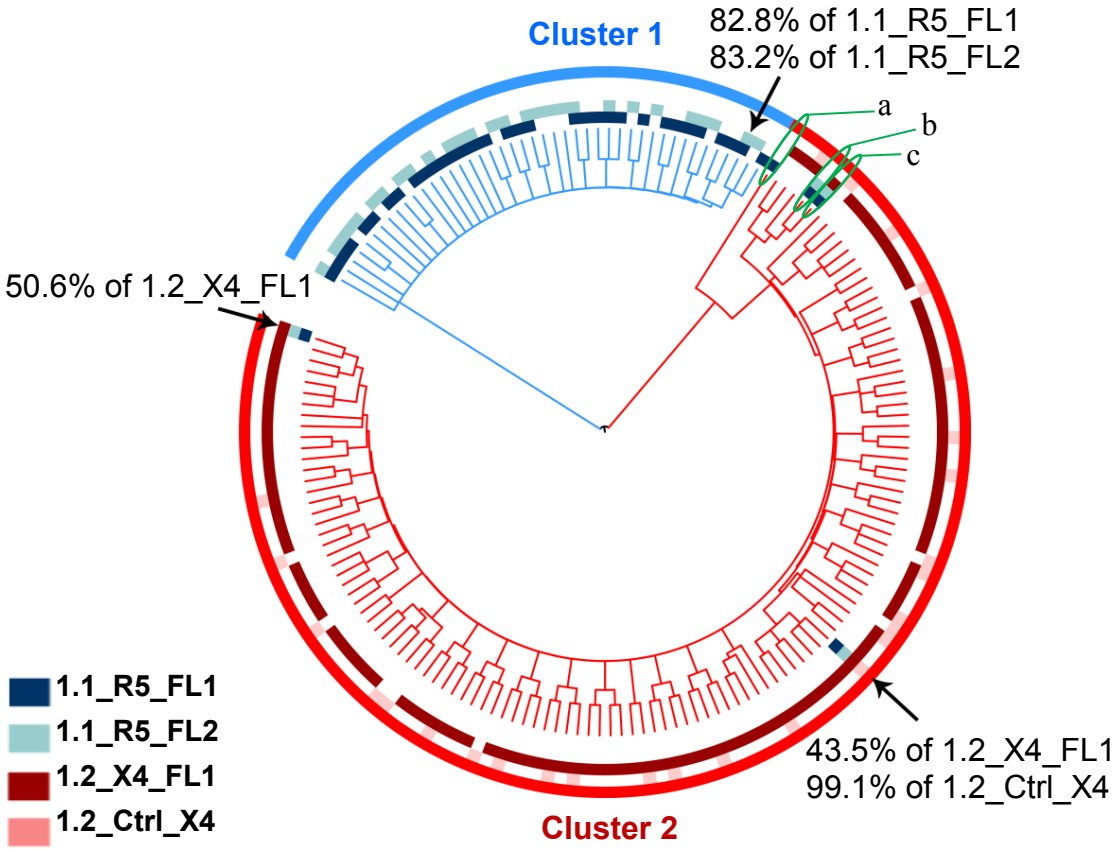
Figure 3.11.3 Functional library population level variant analysis at *env* variable regions.



Reads extracted from variable regions V2 (HXB2 gp160 amino acid 158 – 186d), V3 (297 – 330), V4 (385 – 418) and V5 (460 – 469) from functional libraries 1.1_R5_FL1, 1.1_R5_FL2, and 1.2_X4_FL1 are visualized in pie charts with the most abundant variants annotated. 1.1_R5_FL1 and 1.1_R5_FL2 are biological replicates prepared from Sample 1.1 passaged on

U87.CD4.CCR5 cells. 1.2_X4_FL1 was prepared from Sample 1.2 passaged on U87.CD4.CXCR4 cells. Sequencing data shown were generated on the PacBio RS II system. Individual reads that span through the entire region between the given coordinates were selected, trimmed, and translated into amino acid sequences. Quasispecies are represented by non-redundant amino acid sequences and are color-coded across libraries in the same row. V3 variants are represented in same colors in Row 2 as in Figure 2A-B, and E-F.

Figure 3.11.4 Hierarchical clustering analysis of 167 V3 variants from four functional libraries (1.1_R5_FL1, 1.1_R5_FL2, 1.2_X4_FL1, and 1.2_Ctrl_X4).



The presence and absence of variants is indicated using the color coded squares. Highly abundant variants are indicated by arrows. Green circles denote variants a, b, and c from 1.1_R5_FL1 and/or 1.1_R5_FL2, which were grouped in Cluster 2 and predicted to use CXCR4 by Geno2Pheno[coreceptor].

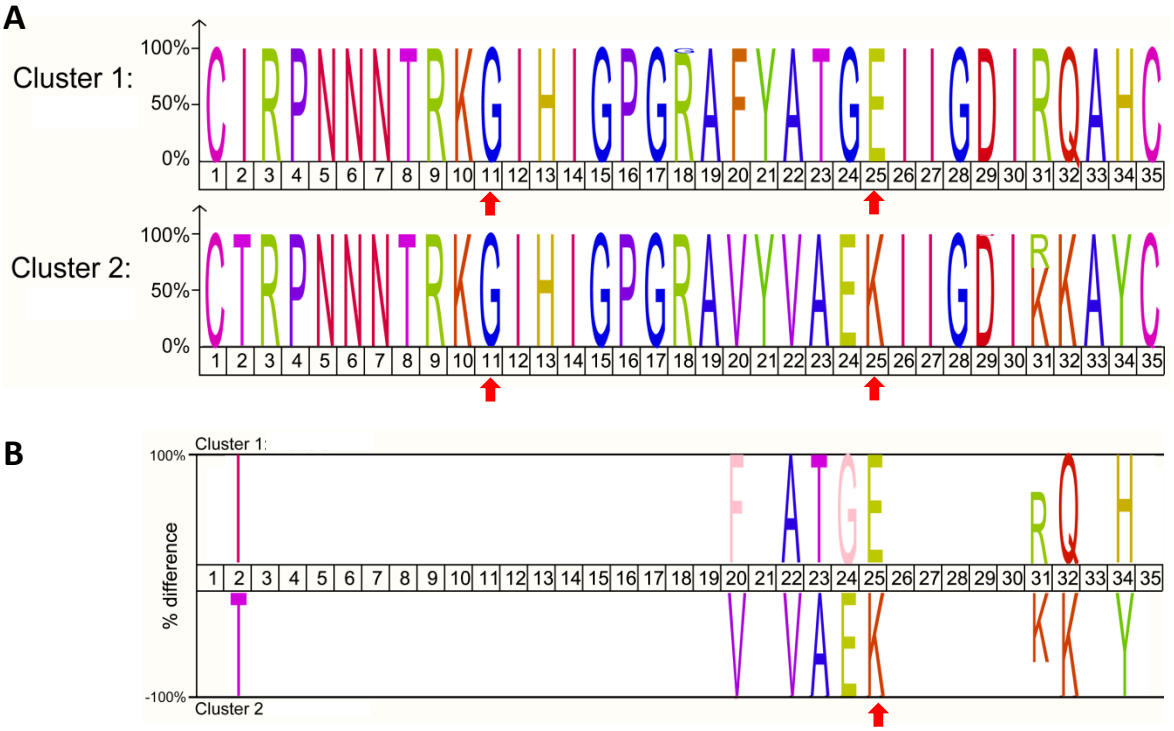
Sequences for the three circled variants are:

a – CTRPNNNTRKGIHIGPGRAVYVAEKIIGNIRQAHC;

b – CIRPNNNTRKGIHIGPGRAVYVAEKIIGNIRKAYC;

and c – CTRPNNNTRKGIHIGPGRAVYVAEKIIGNIRKAYC.

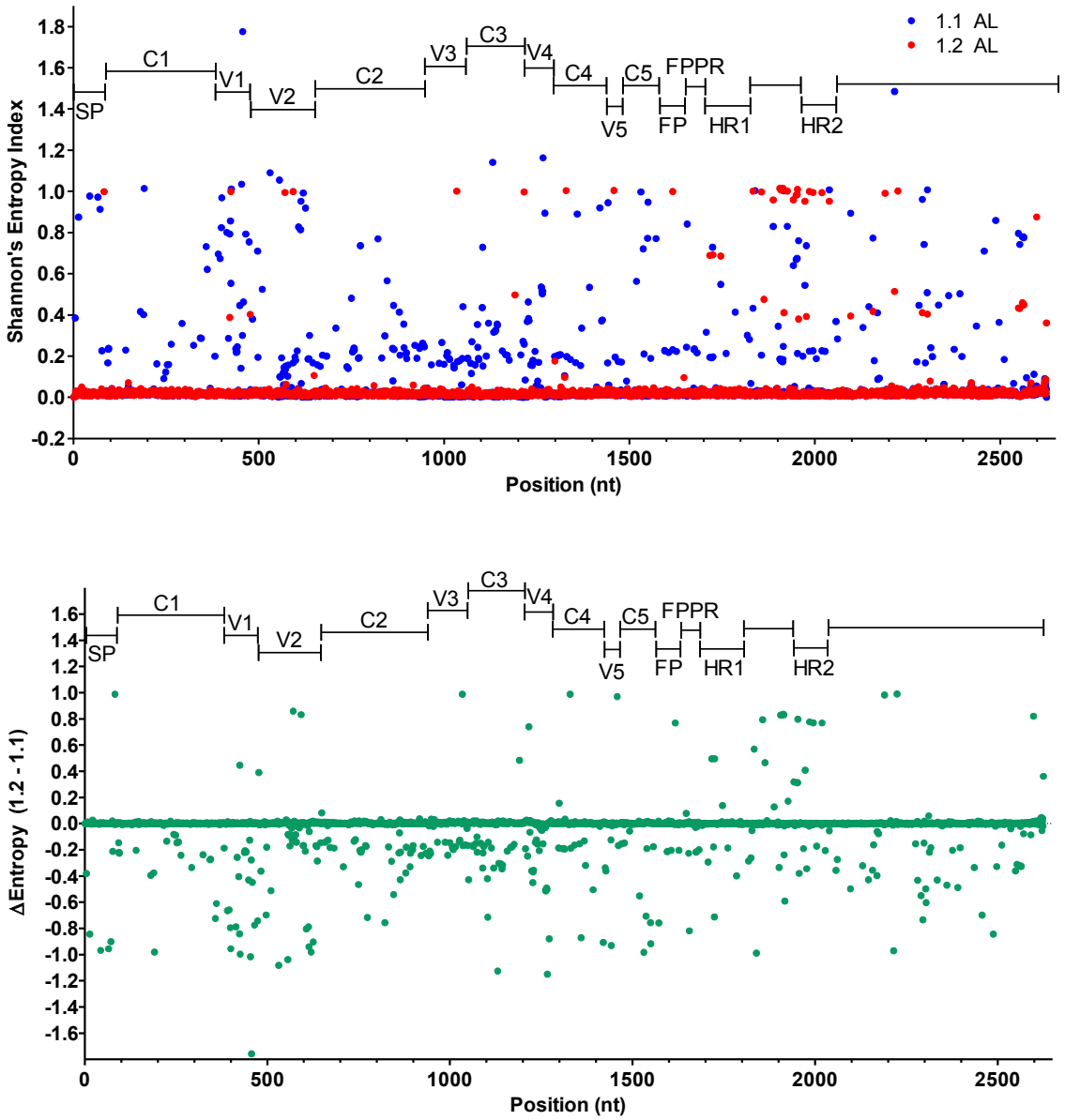
Figure 3.11.5 V3 Logo Plot for Cluster 1 and 2.



(A) ConsensusV3 sequences of Cluster 1 and 2.

(B) Significantly different amino acids ($P < 0.001$) between the V3 loop sequences from cluster 1 and cluster 2. Critical residues at the 11th and 25th positions are highlighted with red arrows.

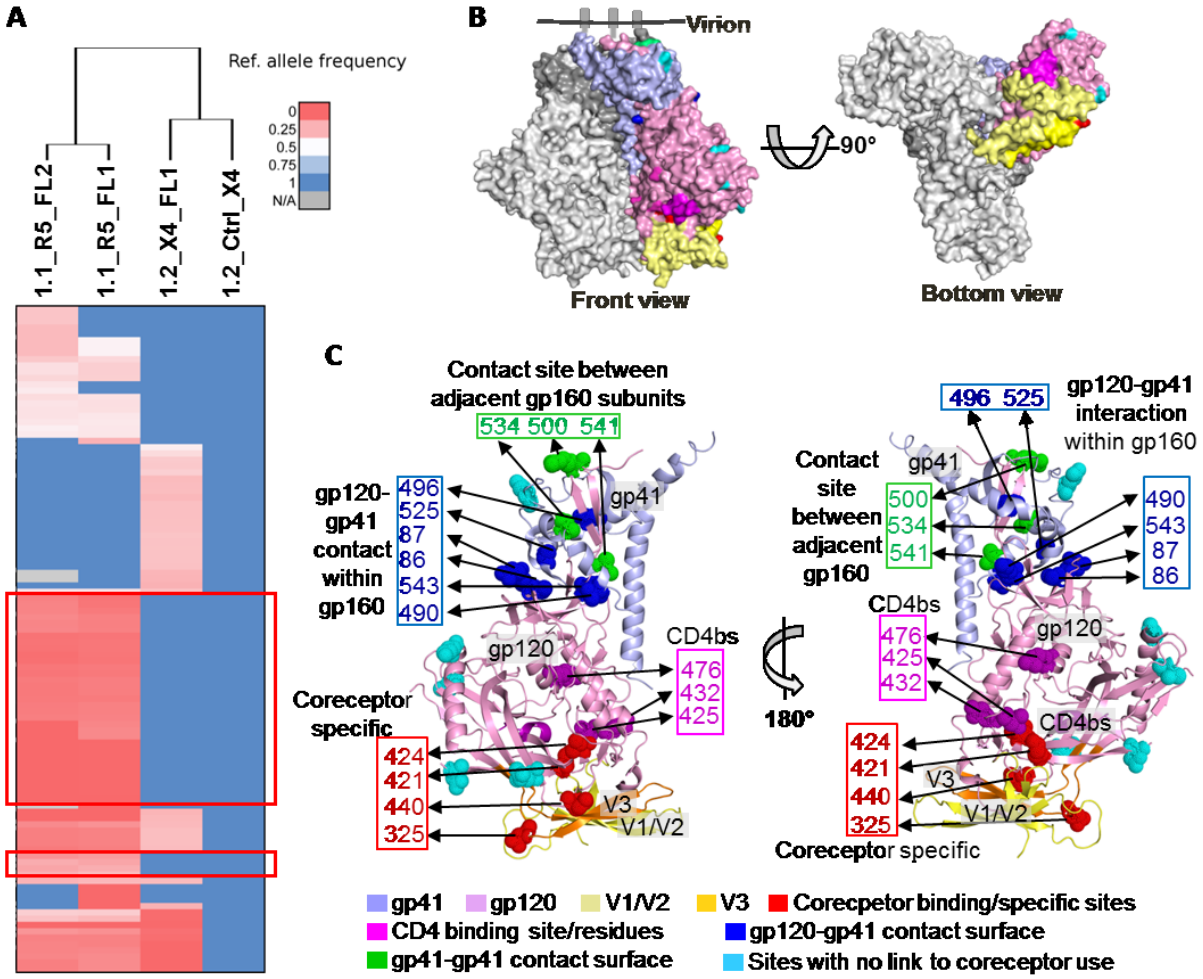
Figure 3.11.6 Shannon's entropy analysis of amplicon libraries (AL) from Sample 1.1 and 1.2.



(A) Absolute entropy values at each nucleotide position along the *env* genes for 1.1 AL and 1.2 AL. A total of 120 positions in Sample 1.1 (blue) and 51 positions in Sample 1.2 (red) have entropy values significantly over background entropy values in the 1.2 clonal control dataset (not shown), by one-tailed Student's t-test, $p<0.0001$.

(B) Entropy difference at each nucleotide position along the *env* genes, by subtracting the values of 1.1 AL from that of 1.2 AL. At 17 positions, the entropy values in Sample 1.1 differ significantly from those in Sample 1.2, by two-tailed Student's t-test, $p<0.0001$.

Figure 3.11.7 Residues outside the V3 loop differentiate R5- vs. X4- usage.



(A) Clustering analysis based on SNP loci allele frequency of the full-length *env* in four functional libraries. A total of 106 SNP loci were identified by GTAK using the nucleotide sequence of the single clone control (1.2_Ctrl_X4) as reference. The reference allele frequency at each locus was calculated in every functional library and is represented as a colored block along the vertical axis of the plot. Euclidean distance based matrix was used to generate the clustering using GENE-E. The branching over the horizontal axis represents genotype similarity based on similar patterns of SNPs found in all four samples. A total of 37 differentiating SNPs are found in the regions outline by the red boxes, among which 24 SNPs lead to non-synonymous amino acid substitution. Only 1 out of the 24 SNPs falls within the V3 loop region.

(B) Representation on the gp140 SOSIP ligand-free trimer structure (PDB ID: 4ZMJ) of 21 loci differentiating CCR5 from CXCR4 utilization. The front view shows a trimeric spike with gp41 on the top and gp120 at the bottom. The bottom view shows the perspective looking up from the host cell membrane. The subunits are shown in surface model, with gp120 in light grey and gp41 in dark grey in two of the three protomers. The third protomer is colored by region as the following: light blue, gp41 ectodomain; pink, gp120; pale-yellow, V1/V2 loop; orange, V3 loop; and magenta, CD4 binding site (CD4bs). The differentiating loci are color-coded based on their known functional status previously reported in the literature. Red: coreceptor binding or coreceptor specific site; magenta: CD4 contact residues; blue: intra-protomer gp120-gp41 interaction site; green: inter-protomer gp120-gp41 and gp41-gp41 interaction site; teal: no previously reported function in receptor or coreceptor interaction. **(C)** Enlarged front and back views of the 21 differentiating residues on a cartoon representation of gp140 subunit. The residues are annotated and grouped based on their functional status and color coded in the same manner as in (B).

3.12 References

1. **Wilen CB, Tilton JC, Doms RW.** 2012. HIV: cell binding and entry. Cold Spring Harb Perspect Med **2**.

2. **Berger EA.** 1997. HIV entry and tropism: the chemokine receptor connection. AIDS **11 Suppl A:**S3-16.

3. **Gorry PR, Ancuta P.** 2011. Coreceptors and HIV-1 pathogenesis. Curr HIV/AIDS Rep **8:**45-53.

4. **Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR.** 1997. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med **185:**621-628.

5. **Schuitemaker H, van 't Wout AB, Lusso P.** 2011. Clinical significance of HIV-1 coreceptor usage. J Transl Med **9 Suppl 1:**S5.

6. **Moore JP, Kuritzkes DR.** 2009. A piece de resistance: how HIV-1 escapes small molecule CCR5 inhibitors. Curr Opin HIV AIDS **4:**118-124.

7. **Henrich TJ, Kuritzkes DR.** 2013. HIV-1 entry inhibitors: recent development and clinical use. Curr Opin Virol **3:**51-57.

8. **Ray N, Doms RW.** 2006. HIV-1 coreceptors and their inhibitors. Curr Top Microbiol Immunol **303:**97-120.

9. **Picard L, Simmons G, Power CA, Meyer A, Weiss RA, Clapham PR.** 1997. Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. J Virol **71:**5003-5011.

10. **Samson M, LaRosa G, Libert F, Paindavoine P, Detheux M, Vassart G, Parmentier M.** 1997. The second extracellular loop of CCR5 is the major determinant of ligand specificity. J Biol Chem **272:**24934-24941.

11. **Siciliano SJ, Kuhmann SE, Weng Y, Madani N, Springer MS, Lineberger JE, Danzeisen R, Miller MD, Kavanaugh MP, DeMartino JA, Kabat D.** 1999. A critical site in the core of the CCR5 chemokine receptor required for binding and infectivity of human immunodeficiency virus type 1. J Biol Chem **274:**1905-1913.

12. **Pontow S, Ratner L.** 2001. Evidence for common structural determinants of human immunodeficiency virus type 1 coreceptor activity provided through functional analysis of CCR5/CXCR4 chimeric coreceptors. J Virol **75:**11503-11514.

13. **Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393:**648-659.

14. **Poignard P, Saphire EO, Parren PW, Burton DR.** 2001. gp120: Biologic aspects of structural features. Annu Rev Immunol **19:**253-274.

15. **Do Kwon Y, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gorman J, Joyce MG, Guttman M, Ma X, Narpala S, Soto C, Terry DS, Yang Y, Zhou T, Ahlsen G, Bailer RT, Chambers M, Chuang GY, Doria-Rose NA, Druz A, Hallen MA, Harned A, Kirys T, Louder MK, O'Dell S, Ofek G, Osawa K, Prabhakaran M, Sastry M, Stewart-Jones GB, Stuckey J, Thomas PV, Tittley T, Williams C, Zhang B, Zhao H, Zhou Z, Donald BR, Lee LK, Zolla-Pazner S, Baxa U, Schon A, Freire E, Shapiro L, Lee KK, Arthos J, Munro JB, Blanchard SC, Mothes W, Binley JM, et al.** 2015. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. Nat Struct Mol Biol **22:**522-531.

16. **Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC.** 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330:**1066-1071.

17. **Saita Y, Kodama E, Orita M, Kondo M, Miyazaki T, Sudo K, Kajiwara K, Matsuoka M, Shimizu Y.** 2006. Structural basis for the interaction of CCR5 with a small molecule, functionally selective CCR5 agonist. J Immunol **177:**3116-3122.

18. **Westervelt P, Trowbridge DB, Epstein LG, Blumberg BM, Li Y, Hahn BH, Shaw GM, Price RW, Ratner L.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J Virol **66:**2577-2582.

19. **Pastore C, Nedellec R, Ramos A, Pontow S, Ratner L, Mosier DE.** 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol **80:**750-758.

20. **Cashin K, Sterjovski J, Harvey KL, Ramsland PA, Churchill MJ, Gorry PR.** 2014. Covariance of charged amino acids at positions 322 and 440 of HIV-1 Env contributes to coreceptor specificity of subtype B viruses, and can be used to improve the performance of V3 sequence-based coreceptor usage prediction algorithms. PLoS One **9:**e109771.

21. **Dobrowsky TM, Rabi SA, Nedellec R, Daniels BR, Mullins JI, Mosier DE, Siliciano RF, Wirtz D.** 2013. Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins. Sci Rep **3:**3014.

22. **Low AJ, McGovern RA, Harrigan PR.** 2009. Trofile HIV co-receptor usage assay. Expert Opin Med Diagn **3:**181-191.

23. **Prosperi MC, Bracciale L, Fabbiani M, Di Giambenedetto S, Razzolini F, Meini G, Colafigli M, Marzocchetti A, Cauda R, Zazzi M, De Luca A.** 2010. Comparative determination of HIV-1 co-receptor tropism by Enhanced Sensitivity Trofile, gp120 V3-loop RNA and DNA genotyping. Retrovirology **7:**56.

24. **Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, Paxinos E, Arts EJ, Robertson DL, Mimms L, Quinones-Mateu ME.** 2012. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One **7:**e49602.

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

26. **Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I, Heera J, Valdez H, Harrigan PR.** 2011. Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients. J Infect Dis **203:**237-245.

27. **Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR.** 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One **4:**e5683.

28. **Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R.** 2007. Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol **25:**1407-1410.

29. **Jensen MA, Li FS, van 't Wout AB, Nickle DC, Shriner D, He HX, McLaughlin S, Shankarappa R, Margolick JB, Mullins JI.** 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol **77:**13376-13388.

30. **Sirois S, Sing T, Chou KC.** 2005. HIV-1 gp120 V3 loop for structure-based drug design. Curr Protein Pept Sci **6:**413-422.

31. **Thielen A, Lengauer T, Swenson LC, Dong WW, McGovern RA, Lewis M, James I, Heera J, Valdez H, Harrigan PR.** 2011. Mutations in gp41 are correlated with coreceptor tropism but do not improve prediction methods substantially. Antivir Ther **16:**319-328.

32. **Gulick RM, Su Z, Flexner C, Hughes MD, Skolnik PR, Wilkin TJ, Gross R, Krambrink A, Coakley E, Greaves WL, Zolopa A, Reichman R, Godfrey C, Hirsch M, Kuritzkes DR, Team ACTG.** 2007. Phase 2 study of the safety and efficacy of vicriviroc, a CCR5 inhibitor, in HIV-1-Infected, treatment-experienced patients: AIDS clinical trials group 5211. J Infect Dis **196:**304-312.

33. **Kirchherr JL, Lu X, Kasongo W, Chalwe V, Mwananyanda L, Musonda RM, Xia SM, Scearce RM, Liao HX, Montefiori DC, Haynes BF, Gao F.** 2007. High throughput functional analysis of HIV-1 env genes without cloning. J Virol Methods **143:**104-111.

34. **Connor RI, Chen BK, Choe S, Landau NR.** 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology **206:**935-944.

35. **He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR.** 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol **69:**6705-6711.

36. **Michael N, Kim JH.** 1999. HIV protocols, Totowa, N.J. Humana Press,.

37. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754-1760.

38. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357-359.

39. **Masso M, Vaisman, II.** 2010. Accurate and efficient gp120 V3 loop structure based models for the determination of HIV-1 co-receptor usage. BMC Bioinformatics **11:**494.

40. **Rodriguez C, Soulie C, Marcelin AG, Calvez V, Descamps D, Charpentier C, Flandre P, Recordon-Pinson P, Bellecave P, Pawlotsky JM, Masquelier B, Group AAS.** 2015. HIV-1 Coreceptor Usage Assessment by Ultra-Deep Pyrosequencing and Response to Maraviroc. PLoS One **10:**e0127816.

41. **Jensen MA, van 't Wout AB.** 2003. Predicting HIV-1 coreceptor usage with sequence analysis. AIDS Rev **5:**104-112.

42. **Korber B, Gnanakaran S.** 2009. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. Curr Opin HIV AIDS **4:**408-417.

43. **Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi A, Kwon YD, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD.** 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science **329:**811-817.

44. **Mishra SH, Shelley CM, Barrow DJ, Jr., Darby MK, Germann MW.** 2006. Solution structures and characterization of human immunodeficiency virus Rev responsive element IIB RNA targeting zinc finger proteins. Biopolymers **83:**352-364.

45. **Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, Stewart-Jones GB, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L,**

Munro JB, Blanchard SC, Mothes W, Connors M, Kwong PD. 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature **514:**455-461.

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

47. **da Silva J, Wyatt SK.** 2014. Fitness valleys constrain HIV-1's adaptation to its secondary chemokine coreceptor. J Evol Biol **27:**604-615.

48. **Moreno-Gamez S, Hill AL, Rosenbloom DI, Petrov DA, Nowak MA, Pennings PS.** 2015. Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. Proc Natl Acad Sci U S A **112:**E2874-2883.

49. **Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M, Nussenzweig MC.** 2015. HIV-1 integration landscape during latent and active infection. Cell **160:**420-432.

Chapter 4

Discussion and Future Directions

4.1 Summary

HIV/AIDS remains a significant public health problem worldwide. The cause of this problem can be attributed to the following three factors. Firstly, the lack of access to diagnostic testing and counselling in low- and middle-income countries with high disease burden, limited financial resources, and poor public health infrastructure leads to a high undiagnosed rate and further spread of the disease. WHO estimated that a large share of HIV-1 infection is undiagnosed in regions of Western Europe, Eastern Mediterranean, and Sub-Saharan Africa (WHO). Secondly, the lack of access to transmission prevention measures such as anti-retrovirals for pregnant HIV-1 positive women to block mother-to-child transmission, clean needles for intravenous drug users, and condoms to block sexual transmission contributes to the 2 million annual new infection cases. This is likely due to lack of financial resources, social stigma and punitive government policy (UNAIDS). Thirdly, the lack of consistent and reliable access to effective HAART drugs for treatment eligible individuals in low- and middle-income countries with high HIV/AIDS prevalence continues to bring morbidity and mortality to a socioeconomically disadvantaged population. The need remains to continue development of less-toxic, more effective, new classes of antiretroviral agents, and potentially curative agents in pharmaceutical pipelines to battle issues with the current life-long therapy, such as hepatic and renal toxicity, and development and circulation of drug-resistant strains.

One such new class of antiretroviral agent is entry inhibitors. These inhibitors work by blocking viral entry, the first step of the viral replication cycle. HIV entry into a host cell is a dynamic, multi-step process that involves viral attachment to the host cell extracellular matrix or receptor, viral envelope protein engagement with primary receptor CD4, recruitment and engagement of one of the two coreceptors CCR5 and CXCR4, followed by eventual formation of

the six-helix bundle and fusion pore, and translocation of viral nucleocapsid core into the host cell cytoplasm, as discussed in the first chapter of this dissertation. Every step in this process is being targeted for the development of novel entry inhibitors. Currently, two FDA-approved entry inhibitors include Enfuvertide (Fuzeon, Roche) that competitively blocks six-helix bundle formation and fusion, as well as Maraviroc (Selzentry, Pfizer) that is an allosteric inhibitor of CCR5 and blocks envelope-CCR5 interaction. A number of other small molecule inhibitors and monoclonal antibody based biologics are being developed through preclinical and clinical phases, among which are several coreceptor antagonists, notably CCR5 antagonists Vicriviroc (VCV) and TAK779, and a potent CXCR4 inhibitor AMD3100 which are not approved for HIV-1 indication due to its side effects. During the development process and clinical application of some of these coreceptor inhibitors, the outgrowth of resistant strains were detected in patients despite tropism testing prior to treatment commencement. Multiple mechanisms were found to be responsible for resistance development, including a tropism change of the virus from CCR5 to CXCR4, and a gain in the ability of the virus to engage the CCR5-antagonist complex for entry. The former mechanism of coreceptor shift represents the majority of CCR5-antagonist resistant cases, and could be further broken down into two pathways, including (1) selective expansion of a minor, pre-existing CXCR4-tropic viral variant that is not detected by tropism testing, under the drug pressure; and (2) *de novo* mutations of the virus to gain the ability to utilize CXCR4 as its entry coreceptor. A thorough understanding of the coreceptor shift process and the route it takes would help with the development of more suitable clinical criteria for CCR5 antagonist eligibility, and with the design of more effective treatment plans in order to minimize resistance development and provide benefit to the largest possible population of eligible patients.

This study utilized a panel of patient-derived HIV-1 envelope amplicon samples before and after the development of VCV resistance in order to advance our understanding of the coreceptor shift process *in vivo*, as well as the basic biology of envelope-coreceptor interaction in general. Using a combination of single clone and next-generation sequencing (NGS) approaches discussed in Chapter 2 and 3, we characterized a large number of clinical isolates of HIV-1 VCV-resistant envelope, and examined in-depth the quasispecies landscape in these patients before and after they became resistant to VCV treatment. We discovered drastic changes in envelope sequences over the course of treatment from the single clone analysis, and observed dynamic quasispecies population shift over the course of 2-8 weeks by NGS. With an array of phenotypic and genotypic assays, we analyzed a large batch of phenotypically validated envelope sequences, and with the Pacific Biosciences technology we were able to study mutations occurring along the full-length gp160 coding sequence. We detected three regions of interests that strongly associate with CXCR4-usage, including but not limited to the V3 loop, which is the principal determinant of coreceptor specificity supported by a large body of research. We identified important residues that were (1) previously reported to play a role in coreceptor choice within and surrounding to the V3 loop; (2) reported to be CD4 contact residues, and (3) located on the gp120-gp41 interaction surface within one gp160 heterodimer and between the neighboring heterodimers within an envelope trimer spike. In addition to confirming previous finding that residues in the V3 loop and CD4 binding site, which make direct contact with coreceptor and CD4, affect coreceptor specificity, we provided new evidence that the interaction between gp120 and gp41, as well as the interaction between the heterodimers in an envelope spike, could modulate coreceptor binding. This new line of evidence leads to new hypothesis that the envelope protein interaction with coreceptor could be influenced possibly through more

global mechanisms, such as the open or closed state of envelope quaternary structure, stability of a particular conformational state that allows for coreceptor binding, and the affinity and duration of coreceptor engagement that leads to fusion and viral entry.

Understanding of the structure-function relationship of HIV-1 envelope and the process of coreceptor shift is critically important, not only for the advancement in our basic scientific knowledge of protein-protein interaction, but also with profound clinical implications in development of diagnostic methods and future pharmaceutical agents. The findings from this *in vivo* study provides candidates for further *in vitro* causative study on envelope structure-function relationship, as well as potential ways to improve the current tropism predictive algorithms by incorporating residues with differentiating power for CXCR4-usage. It also provides new thoughts along the lines of modulating gp120-gp41 interaction via small molecules or monoclonal antibodies to restrict its quaternary structure as an approach to block entry. Last but not least, the knowledge of specific envelope-CXCR4 interaction sites in combination with known interaction sites between CXCR4 and CXCL12 could potentially lead to design of specific inhibitors that preferentially block envelope binding to CXCR4 but do not interfere with the physiological CXCR4 signaling that is essential for immune function. Further investigations are needed in the above mentioned areas to impact HIV/AIDS diagnosis, monitoring and treatment options.

4.2 Discussion

One of the major strengths of this study is that we were able to observe the viral quasispecies evolutionary process *in vivo*. The *in vivo* nature of study design allowed us to make conclusions that are based on physiological processes occurring in their natural environment. The patient samples were paired and contained before and after treatment samples for each study participant. Our assumption is that factors shaping the immune environment, including but not limited to cytotoxic T lymphocyte (CTL) and antibody responses, as well as the specific genetic makeup such as HLA, CD4, and coreceptor genotype, remained largely constant throughout the course of treatment; thus the major variable and driving force for coreceptor shift is the introduction of CCR5 antagonist into the system. The evolutionary process of coreceptor shift was primarily driven by VCV, with important constraints from the above-mentioned specific immune and genetic environment that is mostly lost in many *in vitro* evolutionary studies. Without the CTL and antibody responses and in the presence of non-physiological levels of CD4 and coreceptor expression, the mutations detected in an *in vitro* drug resistance development setting often bear no resemblance to the ones seen in patients. In fact, over the three passages of viral quasispecies in the process to generate functional libraries as described in Chapter 3, we detected a small group of V3 variants in Sample 1.2 that lost the GPG crown motif and deviated significantly from the input sequences. With proper quality controls based on the quasispecies pool in the amplicon and plasmid libraries, we were able to manually remove these artefactual variants that could potentially affect downstream analysis. This highlights the necessity of studying the evolutionary process in a controlled, physiological setting to avoid erroneous and irrelevant conclusions.

The flip-side of the story for an *in vivo* study is that it is difficult to determine causation in the structure-function relationship analysis. The conclusions were based on a strong association between certain specific residue changes and the utilization of CXCR4, yet it is unclear on the differentiation of key driver mutation versus auxiliary mutations. Further *in vitro* site-directed mutagenesis studies are required to understand the effect of one mutation or a combination of mutations on envelope-coreceptor interaction. Identifying key residues in multiple patients would also strongly indicate a causative role of these particular loci in coreceptor shift. With the availability of the remaining six patient-derived envelope amplicon samples, we certainly hope to continue examine the coreceptor shift process in more patients, to increase the power of the analysis and understand the findings in a broader context that are applicable to multiple patients.

One limitation of this study comes from the sample collection methods utilized. Since the patient-derived envelope amplicon samples were prepared from patient plasma, we are limited to exam the quasispecies population in circulation and do not have access to those variants residing in the peripheral lymphoid organs or the latently infected reservoirs. A recent study showed that the clonally expanded T cells in the circulation are more likely to harbor defective provirus containing deleterious mutations and large deletions than the quiescent non-dividing latently infected memory T cells, which usually harbor replication competent proviruses (1). This report is consistent with our observation that a large proportion of the single clones are non-functional, though molecular cloning artifact would also have a minor contribution towards the detection of non-functional single clones. On the other hand, previous reports have discussed the development of drug resistant variants in local tissues due to imperfect penetration and the mismatched half-lives of antiretroviral agents (2). Viral quasispecies in the local tissue may be

sheltered from full-strength HAART, and make slow, step-wise mutation in the context of the microenvironment of temporal and spatial monotherapy, only "spilling over" into the circulation upon achieving full drug resistance. This is consistent with our observation that the quasispecies population underwent a complete change of constituents over the course of treatment, and few drug-resistant intermediate variants were detected. This observation could be due to insufficient resolution and limits of detection to reveal rare, low-fitness variants in the circulation. Alternatively, it might be because we were looking in the wrong place, and should examine the quasispecies pool at the peripheral tissue instead. Future research should be conducted with the help of powerful NGS technologies to closely analyze viral quasispecies population in the latent reservoir and the peripheral tissues such as the central nervous system, in the hope to complete the missing pieces in the evolutionary pathway.

Overall, this study made a significant contribution to our knowledge of HIV-1 envelope structure-function relationship. The findings suggest a modular approach to understand how each region of the HIV-1 envelope protein, such as the V3 loop, CD4 binding site, and gp120-gp41 interaction site, might exert an impact on envelope-coreceptor interaction and coreceptor specificity, through direct or indirectly mechanisms. Despite of recent progress in understanding the HIV-1 envelope trimer structure at different states of activation, it has been a challenge to understand the multimeric, dynamic interaction process of HIV entry using structural methods such as crystallography (3, 4). In-depth association studies using multiple patients on NGS platforms, biochemical and biophysical analysis of single clones, as well as site-directed mutagenesis studies that focus on a fine-resolution understanding of effects of single residues, are needed to piece together the jigsaw puzzle of the big picture of HIV envelope-coreceptor interaction. The information gained through thorough structure-function analysis proposed above

would hopefully contribute to a comprehensive model of protein interaction energy landscape, with which the impact of mutations and pharmaceutical interventions to enhance or disrupt the said interaction could be accurately simulated, thus extending our knowledge of the basic biology of protein-protein interaction, and guiding the development of future therapeutic agents.

4.3 Future work

4.3.1 Phenotypic characterization of envelope-coreceptor interaction using biochemical, biophysical, and biological assays

In order to further our understanding of envelope sequence signature and its relationship with regard to tropism, more patient samples should be analyzed using a combination of the single clone and NGS approach. For the single clone analysis, additional work should be performed to fully characterize specific envelope clones that exhibit interesting coreceptor usage properties. Biochemical assays to analyze glycosylation levels of patient-derived clones might be useful in understanding how glycosylation of the envelope influences coreceptor binding and specificity superimposed on top of other observed mutations (5, 6). Biophysical approaches to accurately determine the binding affinities between envelope clones and CCR5 or CXCR4 can be performed using atomic force microscopy, as a previous study showed that HIV-1 envelope binds less strongly to CXCR4, and mutations in gp41 region that affect the formation of six-helix bundle might turn out to be a rate-limiting factor while the envelope is in weak association with CXCR4 (7). Biological assays including inhibitor IC50 analysis on the single clones or the quasispecies libraries might show a difference in how the VCV sensitive versus resistant envelopes respond to varying concentrations of coreceptor antagonists, thus probing the mechanism of resistance (8-10).

In addition, infection of chimeric coreceptor expressing U87-CD4 cell lines (NIBSC, UK) using patient-derived single clones and functional libraries would help dissect the interaction between envelope and specific regions of CCR5 and CXCR4. The three chimeric cell lines FC-1, FC-4b, and FC-6 were constructed so each one contains specific region of CXCR4 on a CCR5 backbone (Figure 4.4.1). Previous studies categorized R5-tropic lab adapted and patient derived

single clones on these cell line into narrow-R5 and broad-R5, and showed that broad-R5 variants are more likely to undergo coreceptor shift to obtain the ability to use CXCR4 for entry (11-14). With these cell lines, we hope to define a sequence signature on HIV-1 envelope that correlates with its ability to utilize a particular chimeric cell line, thus deducing the key interaction domains on CXCR4 that is required for entry.

The three chimeric cell lines were quality-controlled using PCR, sequencing and flow cytometry. PCR using primers binding to the CCR5 amino-terminus and CXCR4 carboxyl-terminus generated correct-sized products from the genomic DNA of FC-1, FC-4b, and FC-6, confirming the presence of chimeric coreceptor transgene in the genome. Sequencing of these PCR products confirmed the correct identity of these transgenes in the three cell lines, respectively. Flow cytometry of surface stained FC-1, FC-4b, and FC-6 cells using antibodies against CD4 (clone Q4120), CCR5 (clone 12G5) and CXCR4 (clone 2D7) showed a high level of cell surface expression of CD4, CCR5, and CXCR4, making these cell lines ideal for infectivity assays (data not shown).

Infection of chimeric cell lines was performed using single clones as well as functional libraries from Sample 1.2 passaged on X4. Interestingly, the single control clone from Sample 1.2 that was previously determined to be dual-tropic with a preference for CXCR4 was able to infect FC-4b at a comparable efficiency as the CXCR4-expressing cell line, and to a lesser extent FC-6, but was completely non-infectious on the FC-1 cell line (Figure 4.4.2). Similarly, the functional library from Sample 1.2 passaged on the CXC4-expressing cell line could infect FC-4b, and to a lesser extent, FC-6, but was unable to infect FC-1. This observation suggests that a combination of the CXCR4 N-terminus and CCR5 extracellular loops were counterproductive for dual-tropic viral entry. A single, dual-tropic envelope clone might utilize different

determinants to interact with CCR5 or CXCR4, and a "mixed-and-match" scenario for the two-site interaction model did not allow viral entry. Additionally, it showed the importance of CCR5 N-terminus for an envelope-R5 interaction, which is supported by previous studies (15, 16). Upon receiving the NGS dataset of FC-4b and FC-6 functional libraries of Sample 1.2 passaged on X4-expressing cells, analysis will be performed to understand the quasispecies that can utilize the two specific chimeric coreceptors, and hopefully reveal the sequence signature that allows them to interact with one chimeric cell line but not the other. More single clones from Subject 1 and 2 should also be tested to differentiate narrow versus broad R5 clones, as well as narrow versus broad dual/X4 clones.

4.3.2 Genotypic analysis of envelope-coreceptor interaction using bioinformatics tools

To better understand the quasispecies evolutionary process, more bioinformatics tools could be used to study the relationship between different envelope single clones or variants, and make correlation between evolution and function. In particular, dN/dS analysis could be performed using variants from before and after treatment, alone and combined, to reveal the similarity between each pair and position of every variant on a maximum likelihood phylogenetic tree. The input for this above-mentioned dN/dS study could be full-length gp160 envelope sequence from single clone sequencing or Pacific Biosciences Read of Insert, or sequences of specific envelope regions, such as gp120, gp41, or sub-regions such as each individual conserved regions C1 through C5, and variable regions V1 through V5, or even using a sliding window format to determine the evolutionary selection on every continuous stretch of the envelope gene. A dN/dS value greater than 1 could indicate significant positive selection

pressure on the specific input region, due to a relatively weak power of this analysis on closely related species.

Another bioinformatics analysis that should be done with the existing dataset is a linkage analysis between different single nucleotide polymorphism (SNP) loci along the full-length gp160 coding sequence. Thanks to the long read length by Pacific Biosciences sequencing platform, it is now possible to detect positions that co-vary in X4-using or R5-using variants. A Viral Epidemiology Signature Pattern Analysis (VESPA) tool on HIV database website can be used to examine residue frequencies in the two different groups of variants, thus identifying sequence signature patterns associated with coreceptor specificity across the length of the envelope.

4.4 Conclusions

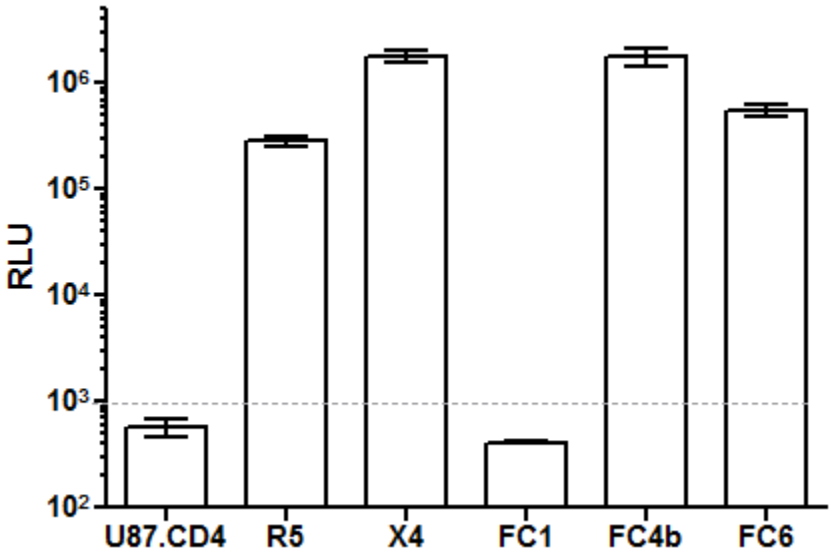
This study utilized a panel of patient-derived envelope amplicon samples to interrogate the evolutionary process of coreceptor shift *in vivo* under the specific selection pressure of a CCR5 antagonist therapy. Four samples from two subjects were examined in-depth to reveal drastic population-level shift of quasispecies from before to after CCR5 antagonist treatment. SNP analysis highlighted regions of interest on the envelope protein that strongly associate with coreceptor specificity, which includes residues near and on the V3 loop, at the CD4 binding site, and at the interaction surface between gp120 and gp41 within the same heterodimer and among neighboring heterodimers in an envelope trimeric spike. Future work should continue to perform phenotypic and genotypic analysis on more patient-derived samples using single clone and NGS approach on CCR5, CXCR4, and chimeric coreceptor expressing cell lines to (1) understand if the findings on two patients are broadly applicable to other patients; (2) explore structure-function relationship *in vitro* with biochemical, biophysical, and biological methods; and (3) analyze the evolutionary relationship between quasispecies using bioinformatics approaches. The enhanced understanding of HIV-1 envelope interaction with host cell coreceptor CXCR4 may have implications in basic biology of protein-protein interaction, as well as important clinical implications in the diagnosis, monitoring, and treatment options for HIV/AIDS.

4.5 Figures

Figure 4.5.1 Schematic diagram of CCR5, CXCR4, and the chimeric coreceptors FC-1, FC-4b, and FC-6.

The diagram is color coded with parental constructs CXCR4 (left) in black, and CCR5 (right) in pink. Chimeric coreceptor constructs FC-1, FC-4b and FC-6 are colored accordingly, with regions from CXCR4 colored black, and regions from CCR5 colored pink.

Figure 4.5.2 Luciferase based infectivity assay on U87.CD4 chimeric coreceptor cell lines using Sample 1.2 single control clone.



U87.CD4 is the negative control parental cell line and does not express significant levels of CCR5 or CXCR4. R5: U87.CD4.CCR5 cell line; X4: U87.CD4.CXCR4 cell line. A grey dotted line at RLU = 917 indicts the cut-off value below which is considered negative background. RLU: relative light unit.

4.6 References

1. **Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M, Nussenzweig MC.** 2015. HIV-1 integration landscape during latent and active infection. Cell **160:**420-432.

2. **Moreno-Gamez S, Hill AL, Rosenbloom DI, Petrov DA, Nowak MA, Pennings PS.** 2015. Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. Proc Natl Acad Sci U S A **112:**E2874-2883.

3. **Bartesaghi A, Merk A, Borgnia MJ, Milne JL, Subramaniam S.** 2013. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. Nat Struct Mol Biol **20:**1352-1357.

4. **Tran EE, Borgnia MJ, Kuybeda O, Schauder DM, Bartesaghi A, Frank GA, Sapiro G, Milne JL, Subramaniam S.** 2012. Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. PLoS Pathog **8:**e1002797.

5. **Ogert RA, Lee MK, Ross W, Buckler-White A, Martin MA, Cho MW.** 2001. N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism. J Virol **75:**5998-6006.

6. **Polzer S, Dittmar MT, Schmitz H, Schreiber M.** 2002. The N-linked Glycan g15 within the V3 Loop of the HIV-1 External Glycoprotein gp120 Affects Coreceptor Usage, Cellular Tropism, and Neutralization. Virology **304:**70-80.

7. **Dobrowsky TM, Rabi SA, Nedellec R, Daniels BR, Mullins JI, Mosier DE, Siliciano RF, Wirtz D.** 2013. Adhesion and fusion efficiencies of human immunodeficiency virus type 1 (HIV-1) surface proteins. Sci Rep **3:**3014.

8. **Jiang X, Feyertag F, Meehan C, McCormack G, Travers SA, Craig C, Westby M, Lewis M, Robertson DL.** 2015. Characterising the diverse mutational pathways associated with R5-tropic maraviroc resistance: HIV-1 that uses the drug-bound CCR5 coreceptor. J Virol doi:10.1128/JVI.01384-15.

9. **Westby M, Smith-Burchnell C, Mori J, Lewis M, Mosley M, Stockdale M, Dorr P, Ciaramella G, Perros M.** 2007. Reduced maximal inhibition in phenotypic susceptibility assays indicates that viral strains resistant to the CCR5 antagonist maraviroc utilize inhibitor-bound receptor for entry. J Virol **81:**2359-2371.

10. **Roche M, Salimi H, Duncan R, Wilkinson BL, Chikere K, Moore MS, Webb NE, Zappi H, Sterjovski J, Flynn JK, Ellett A, Gray LR, Lee B, Jubb B, Westby M, Ramsland PA, Lewin SR, Payne RJ, Churchill MJ, Gorry PR.** 2013. A common mechanism of clinical HIV-1 resistance to the CCR5 antagonist maraviroc despite divergent resistance levels and lack of common gp120 resistance mutations. Retrovirology **10:**43.

11. **Cavarelli M, Mainetti L, Pignataro AR, Bigoloni A, Tolazzi M, Galli A, Nozza S, Castagna A, Sampaolo M, Boeri E, Scarlatti G.** 2014. Complexity and dynamics of HIV-1 chemokine receptor usage in a multidrug-resistant adolescent. AIDS Res Hum Retroviruses **30:**1243-1250.

12. **Cavarelli M, Karlsson I, Zanchetta M, Antonsson L, Plebani A, Giaquinto C, Fenyo EM, De Rossi A, Scarlatti G.** 2008. HIV-1 with multiple CCR5/CXCR4 chimeric receptor use is predictive of immunological failure in infected children. PLoS One **3:**e3292.

13. **Cavarelli M, Karlsson I, Ripamonti C, Plebani A, Fenyo EM, Scarlatti G.** 2010. Flexible use of CCR5 in the absence of CXCR4 use explains the immune deficiency in HIV-1 infected children. AIDS **24:**2527-2533.

14. **Ripamonti C, Leitner T, Lauren A, Karlsson I, Pastore A, Cavarelli M, Antonsson L, Plebani A, Fenyo EM, Scarlatti G.** 2007. Biological and genetic evolution of HIV type 1 in two siblings with different patterns of disease progression. AIDS Res Hum Retroviruses **23:**1531-1540.

15. **Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SS, Stanfield RL, Robinson J, Sodroski J, Wilson IA, Wyatt R, Bewley CA, Kwong PD.** 2007. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. Science **317:**1930-1934.

16. **Seibert C, Cadene M, Sanfiz A, Chait BT, Sakmar TP.** 2002. Tyrosine sulfation of CCR5 N-terminal peptide by tyrosylprotein sulfotransferases 1 and 2 follows a discrete pattern and temporal sequence. Proc Natl Acad Sci U S A **99:**11031-11036.