4-22-2020 11:00 AM

Designing a Novel HIV-1 Candidate Vaccine

Rahul Pawa, *The University of Western Ontario*

Supervisor: Mann, Jamie F.S., *The University of Western Ontario*
Co-Supervisor: Arts, Eric J., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Microbiology and Immunology

Abstract

Currently no vaccine has been developed that can prevent the spread of HIV-1. During sexual transmission, a single viral variant called the Transmitted/Founder (T/F) purportedly with unique physical properties, establishes infection in 70-80% of individuals. Unlike previous studies that have tried to identify T/F viruses based on their structure glycan composition and amino acid sequence, we have analyzed the RNA sequences of HIV-1 to help identify T/F variants. Using a combination of both *in silico* data analysis and *in vitro* assays, we have identified that T/F viruses have higher numbers of immunostimulatory motifs than HIV virions that fail to infect. Moreover, we are the first to demonstrate that a higher composition of TNF-a inducing motifs within T/F RNA may enhance transmission success. Using the identified T/F viruses identified previously, we have designed novel vaccine immunogens by engineering the T/F Env glycoproteins to enhance their abilities to elicit neutralizing antibody responses.

Keywords

HIV-1, transmission, Transmitted/founder virus, pro-inflammatory, TNF-α, Envelope glycoprotein, CD4 binding site, broadly neutralizing antibodies, vaccine

Summary for Lay Audience

The human immunodeficiency virus-1 (HIV-1) is the cause of the acquired immunodeficiency syndrome (AIDS) where individuals succumb to alternative infections due to a weakened immune system. The primary route of transmission for this virus is sexual intercourse. During coital acts, the recipient is exposed to millions of viruses from the donor; however, in majority of the cases, only one virus enters the blood stream and causes the infection. This virus is known as the Transmitted/founder (T/F) virus. The virus is able to hide from the immune system by adding sugars on its surface protein, envelope (Env). Unfortunately, there is no vaccine that has been able to encourage the immune system to produce protective broadly neutralizing antibodies (bnAbs). The goal of this study was to create a novel antigen that can be used as a vaccine to produce bnAbs. To start, we created a computer program that is able to read through thousands of viral RNA sequences from transmission pairs and enumerate specific pro-inflammatory motifs. These motifs are sequences that are 4 nucleotides long. Lastly, we removed specific sugars that surround a key region on Env, called the CD4 binding-site (bs). By removing sugars, we create a "glycan hole" on the target protein that can be targeted by the immune system. The presence of a glycan hole on the target protein may encourage bnAbs against the CD4bs to bind. In the end, we demonstrated that the RNA of T/F viruses have more pro-inflammatory sequences and that these viruses are more likely to be present in the semen of the male donor. We proposed that more pro-inflammatory motifs in T/F viruses may explain the enhanced transmission success. Concurrently, we showed that by removing a specific glycan, bnAbs targeting the CD4bs can easily bind and prevent the virus from infecting a cell. Altogether, by removing a specific glycan on potential T/F viruses, we may be able to create a vaccine that can elicit bnAbs. In theory, by targeting the specific virus that causes the infection, we can prevent HIV-1 transmission and stop the AIDS pandemic. [343]

Co-Authorship Statement

The experiments and analysis presented in this MSc dissertation were predominantly carried out by Rahul Pawa under the supervision of Dr. Jamie F. S. Mann and Dr. Eric J. Arts. Details regarding contributions are listed below:

Chapter 2: Katja Klein and Rahul Pawa, Chanuka N. Wijewardhana, Ryan Troyer, Eric J. Arts and Jamie F. S. Mann. The Antigenicity of HIV-1 viral RNA may contribute to its transmission fitness. ***Nat Comm. Manuscript in preparation for submission***

Rahul Pawa wrote a first draft of the manuscript (a version which appears in Chapter 2). Dr. Jamie F. S. Mann, Dr. Katja Klein and Dr. Eric J. Arts helped in editing the manuscript. All genomic sequence analysis was performed by Rahul Pawa. Dr. Katja Klein, listed as co-author, performed the tissue explant experiment (paper under review).

Chapter 3: Rahul Pawa, Seth Kibel, Renata Ceccacci, Katja Klein, Meijuan Tian, Eric J. Arts and Jamie F. S. Mann. Engineering HIV Virus-Like-Particles for Improved Env Antigenicity to CD4 binding-site Directed Neutralizing Antibodies. ***Additional experiments underway and manuscript in preparation***

Rahul Pawa wrote a first draft of the manuscript (a version which appears in Chapter 3). Rahul Pawa performed the design of the mutagenic primers and the initial yeast-based recombinations. Additional support was provided by Seth Kibel, Renata Ceccacci and Meijuan Tian on performing yeast cloning. Meijuan Tian helped with performing the neutralization assays.

Contribution statement: Figures 1-4, 1-6 and 2-9(A) were adapted from image assets retrieved from the Mann Lab image repository. Katja Klein conducted the tissue explant assay in Figure 2-9(B).

Acknowledgments

I would like to thank Dr. Jamie F. S. Mann and Dr. Eric J. Arts for their indispensable advice and guidance in completing this work. Their constant reassurance, positive feedback and words of encouragement have been key to my personal growth as a scientist and the success of the overall work presented in this dissertation. I would also like to extend my gratitude to Dr. Katja Klein; thank you for always believing in me when I failed to do so. This accomplishment would definitely not be possible without the support that you have all provided me throughout my time in the lab. Thank you for pulling me up and encouraging me to keep the big picture in mind during my tough moments.

I would also like to thank my advisory committee, Dr. Stephan Barr and Dr. Rodney DeKoter for their continued intellectual support throughout my research. Thank you for the constant critical feedback throughout our committee meetings. I appreciate the opportunity to work closely with my committee members and have learned a lot from them –skills that will definitely benefit my future academic endeavors.

I am extremely grateful for the previous lab members in the Mann lab, Mr. Josh Pankrac and Chanuka Wijiwardhana for continuously providing me with reassurance and a helping hand when moments became very stressful. Thank you, Josh, for taking the time to learn and play squash with me and providing me with a platform to destress.

I would also like to extend my thanks to members of our collaborative lab space, Darshit Patel, Ermela Paparisto, Ankita Kambli, Adam Meadows, Spencer Yeung and Jason Knapp for their constant support and guidance through my graduate studies. In particular, thank you Darshit for continuously pushing me to do the best that I can and for always reassuring me that things will be okay.

Finally, I would like to thank my family (mom, dad and sister) and my girlfriend, Maitri Makwana, for their constant support and words of encouragement. It would not have been possible without you. From meal preps, to teaching me how to cook, to always reminding me of my potential, I am fortunate to have the support of beautiful hardworking people.

Table of Contents

List of Tables

List of Figures

List of Supplementary Figures

List of Abbreviations

Abs – antibodies
AIDS – Acquired Immunodeficiency Syndrome
AP-1 – activator protein
Asn – asparagine
BCR – B cell receptor
bnAbs – broadly neutralizing antibodies
CA – capsid protein
CDC – Center for Disease Control
CDR – complementary-determining region
CD4bs – CD4 binding site
CLR – C-type lectin
CRF – circulating recombinant forms
CTL – Cytotoxic T lymphocyte
DCs – Dendritic cells
DC-SIGN – dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin
Env – envelope glycoprotein
FTM – female to male
Gln – glutamine
HAART – highly active antiretroviral therapy
HET – heterosexual
HIV – Human Immunodeficiency Virus
HTLV –Human T cell Leukemia Virus
IDU – intravenous drug users
Ig – Immunoglobulin
IN – integrase protein
INSTIs – integrase inhibitors
LanL – Los Alamos National Laboratory HIV Database
LC – Langerhans's cells
LTR – long terminal repeat
MA – matrix protein
MPER – membrane proximal external region
MSM – men who have sex with men
MTF – male to female
NC – nucleocapsid
Nef – negative regulating factor
NFAT – nuclear factor of activated T cells
NF-κB –nuclear factor κB
NHP – non-human primates
NNRTIs – non-nucleoside reverse transcriptase inhibitor
NPC – nuclear pore complex
NRTIs – nucleoside reverse transcriptase inhibitors
ORFs – open reading frames
PAMPS – pattern recognition receptors
PBS – Phosphate-Buffered Saline
PIC – pre-integration complex
PNLGs – potential N-linked glycans

PR – protease
PrBS – primer binding site
Rev – RNA splicing regulator
RT – reverse transcriptase
RTC – reverse transcription complex
SIV – Simian Immunodeficiency Virus
SMH – somatic hypermutation
STIs – sexually transmitted diseases
TAR – transactivation response
Tat – transactivator protein
T/F – Transmitted/founder
Tfh – T follicular helper cells
TLRs – Toll-like receptors
TNF – Tumor necrosis factor
Vif – viral infectivity protein
Vpr – virus protein r
Vpu – viral protein unique
Vpx – Virus protein x

Chapter 1 Introduction

General Characteristics of the Human Immunodeficiency Virus (HIV)

History of HIV

1981 marked a year in the United States where physicians started noticing a rise in infections that would ultimately become the global epidemic known as Acquired Immunodeficiency Syndrome (AIDS). Young, gay men started entering hospital systems and succumbed to variety of opportunistic infections, which their immune systems should have fended off, and rare malignancies [1–3]. One of the distinctive telltale signs of infection was the patient's presentation with dark purple lesions on their arms and face. These lesions were due to a rare and aggressive form of cancer known as Kaposi's sarcoma[4]. AIDS was initially termed the "gay plague"; however, the Center for Disease Control (CDC) in 1982 termed the disease AIDS based on its disease pathogenesis[2]. The following year, new epidemiological evidence showed that the causative agent for AIDS was transferred by bodily fluids such as exposure to contaminated blood[2,5,6,7]. While the initial reports suggested male bisexual activity and blood transfusions as the primary risk factors for AIDS, follow-up studies in regions outside of the United States indicated heterosexual transmission as the predominant mode of spread[2,8–10]. The documentation of heterosexual transmission of AIDS in 1983 by the CDC slowly started to change the perception of the disease away from the heavily stigmatized populations of men who have sex with men (MSM) and intravenous drug users (IDU).

With the growing prevalence of the disease, many groups were inclined to isolate the causative agent. In 1983, Luc Montagnier and Francois Barre Sinousi, isolated a new retrovirus in a pre-AIDS patient with acute lymphadenopathy[2]. Based on the disease pathogenesis, it was believed that the causative agent was a strain of the Human T cell Leukemia Virus (HTLV). However, by electron microscopy, it appeared to attain a different structure[2,11–13]. A year later, Robert Gallo provided the causal link between the virus and AIDS and the same virus was also independently isolated by Jay Levy and colleagues[2,14,15]. It was not until 1986 that Harold Varmus, the chair of the International

Committee of the Taxonomy of Viruses, recommended naming the pathogen the Human Immunodeficiency Virus (HIV)[2,16]. While both viruses target the same subset of T cells, HTLV transforms these cells, resulting in adult T-cell leukemia, whereas HIV depletes CD4 T cells[2]. Luc Montagnier and Francois Barre Sinousi were subsequently awarded the Nobel prize for their initial discovery of the HIV-1 virus.

In the drive of understanding the origin of the virus, researchers identified a similar virus to HIV that was isolated from captive, laboratory primates –Simian Immunodeficiency Virus (SIV). Based on the biology of the two viruses, the zoonotic transmission of the virus is thought to have occurred through cutaneous or mucous membrane exposure to infected primates[3]. The most common method of exposure would be in the context of bushmeat hunting[3,17]. Peeters and her colleagues performed field work in Cameroon where they tested more than 1000 blood samples from bushmeat and showed that all species of Old-World nonhuman primates (NHP) carry SIV[18]. The common practice of bushmeat hunting and butchering of primates therefore presents the most plausible origin of HIV[17,18]. Based on phylogenetic and statistical analysis of infected blood tissues, it is believed that the probable location of the early epidemic emerged in colonial west central Africa –in particular, Leopoldville/Kinshasa[19]. The cause for the HIV/AIDS pandemic is most likely due to the urbanization and globalization of urban populations in west Africa. Large cities like Leopoldville developed to ensure efficient means of transportation which allowed infected individuals to carry the virus outside of the city. This eventually fueled what is now a global pandemic that affects virtually every nation in the world[3].

Diversity of HIV

HIV can be divided into two strains, HIV-1 and HIV-2. HIV-1 is believed to be derived from Chimpanzee and Gorilla species and transmitted via SIVcpz*Ptt* (*Pan troglodytes troglodytes*) and SIVgor, while HIV-2 was transmitted from several zoonotic transmissions from sooty mangabey monkeys. It is the HIV-1 strain that is responsible for causing the global AIDS pandemic[20]. HIV-2 is mainly found in west Africa and while its pathogenesis is similar to HIV-1, the progression to AIDS is slower and the virus is less transmissible[3]. HIV-1 is further divided into four main groups –M, N, O and P– with each representing a distinct zoonotic event. Groups M, N and O represent separate transmission events from

chimpanzees, while Group P represents the transmission event from gorillas[21]. Of the 4 HIV-1 groups, group M is the cause of the global pandemic and contains 10 genetically unique subtypes: A-D, F-H, K-L[19,21]. Subtype L is an extremely rare subtype of HIV-1 that was discovered in 2019 by researchers at Abbott[22]. The majority of the subtypes are isolated in Africa which is referred to as the epicenter of the global HIV epidemic. Subtype C predominates in southern Africa and India while subtype A predominates in eastern Africa[21]. Subtypes C and D are usually found circulating together. Subtype B predominates in western Europe, the Americas (north and south) and Australia. All these subtypes can also co-infect individuals and create recombinant forms where parts of each subtype create a genetically unique circulating strain. These viral strains are referred to as circulating recombinant forms (CRF) and China is considered a hotbed for these, with the most CFRs identified within its borders than any other country[23]. On a more global scale, subtype C is the dominant subtype causing about 46.6% of infections, compared to subtypes A (10.2%) and B (12.1%)[21].

HIV-1 structure and genes

The Human Immunodeficiency Virus (HIV) is a Lentivirus that comes from the family of *Retroviridae*; more specifically the subfamily Orthoretrovirinae[24]. During the infection cycle, the formation of the viral particle undergoes various stages to eventually create a mature HIV virion that is round in shape and about 100-120 nm in diameter[24,25]. Since the virus buds from the infected cell, the virus contains an outer lipid membrane that comes from the host. The virus packages two identical positive-sense, single-stranded RNA molecules within the core of the virus particle. The RNA is 9749 nucleotides long, contains a 5’ cap (Gppp), a 3’ poly(A) tail and the unspliced nature of the genome encodes for nine open reading frames (ORFs) to allow production of viral proteins necessary for viral replication, survival and pathogenesis[26,27].

The HIV genome is flanked by long terminal repeat (LTR) sequences on both ends each containing three regions: U3, R and U5. There are three functional regions within the U3 region, and one functional region within the R region that all play a role in regulating viral transcription post integration[28]. The presence of a transactivation response (TAR) element within the R region allows the accessory protein transactivator (Tat) to bind and promote

transcription[29]. The U3 region of the LTR codes for the basal/core promotor, core enhancer and modulatory regions that collectively bind to host transcription factors such as nuclear factor κB (NF-κB), nuclear factor of activated T cells (NFAT) and activator protein 1 (AP-1) to mediate transcription of viral mRNA[30]. Since the 3' LTR is a mirror image of the 5'LTR, the same regions and functional properties exist. Due to interference with the 5'LTR, the 3'LTR does not a play a dominating role in transcription initiation, however it does function in the polyadenylation of viral transcripts[31].

The 5' LTR codes for the promotor to initiate transcription of ORF encoding genes in the direction 5' to 3'. Three of these ORFs encode for polyproteins Gag, Pol and Env, which are subsequently proteolyzed into proteins common to all retrovirus. The remaining six ORFs are much shorter and encode for accessory proteins that play a role in regulating parts of the viral life cycle and pathogenesis[24,32]. The Gag polyprotein encodes for the matrix (MA, p17), capsid (CA, p24), nucleocapsid (NC, p7) and p6 which have a structural role[24,33]. The CA is a 24 kD protein that forms the conical capsid that protects the viral RNA and additional proteins necessary for replication[33]. The MA is a 17 kD protein that is myrislated and forms the inner membrane layer[24]. The NC is a 7 kD protein that plays a key role in forming the nucleoprotein/RNA complex necessary to shuttle the unspliced RNA complex to the site of viral egress[24,34]. The Pol polyprotein encodes for the protease (PR, p10), reverse transcriptase (RT, p51), RNase H and integrase (IN, p32)[24,32,35]. The PR is a 10 kD protein that performs the proteolytic cleavage of Gag (Pr55) and Gag-Pol (Pr160GagPol) precursor protein to release structural and viral protein[24]. The RT is a 51 kD protein that performs the transcription of HIV RNA to DNA to allow for integration into the host genome. The RNase H functions to degrade the RNA in the viral RNA/DNA replication complex[36]. The IN is a 32 kD protein that functions to successfully integrate the proviral DNA made by RT into the host genome[37]. The successful integration and location within the host of the genome will allow production of progeny virus. The final retrovirus polyprotein, Env, is synthesized as a gp160 precursor that is cleaved post-translationally to yield two associated glycoproteins, the gp120 surface subunit and gp41 transmembrane subunit[38,39]. This then arranges as a heterotrimer on the viral surface[40,41].

While gp120 binds to the CD4 receptor and either the CCR5 or CXCR4 co-receptor on T helper cells, gp41 mediates fusion of viral and cellular membranes[40,42].

Besides the core retroviral proteins, the HIV genome also encodes for the following accessory proteins. Transactivator protein (Tat) and the RNA splicing regulator (Rev) function to activate the transcription of viral genes and regulate the export of non-spliced and partially spliced viral mRNA, respectively. The additional regulatory proteins – negative regulating factor (Nef), viral infectivity protein (Vif), virus protein r (Vpr) and viral protein unique (Vpu)– influence HIV replication, virus production *in vivo*, infectivity and pathogenesis[24,32,43–45]. In particular, Nef and Vpu share the same function of CD4 downregulation on target cells, however, Nef can also downregulate HLA[46–48]. While the accessory proteins are not necessary for viral replication and infection, their presence enhances overall infectivity by protecting the virus from immune-mediated attack and clearance. In comparison to Vpu in HIV-1, HIV-2 encodes for virus protein x (Vpx) which functions to degrade SAMDH1, a host restriction factor involved in viral defense. Interestingly, the presence of this protein in HIV-2 is also responsible for its weakened pathogenesis[49,50]. Refer to **Figure 1-1** for a map of the HIV-1 genome.

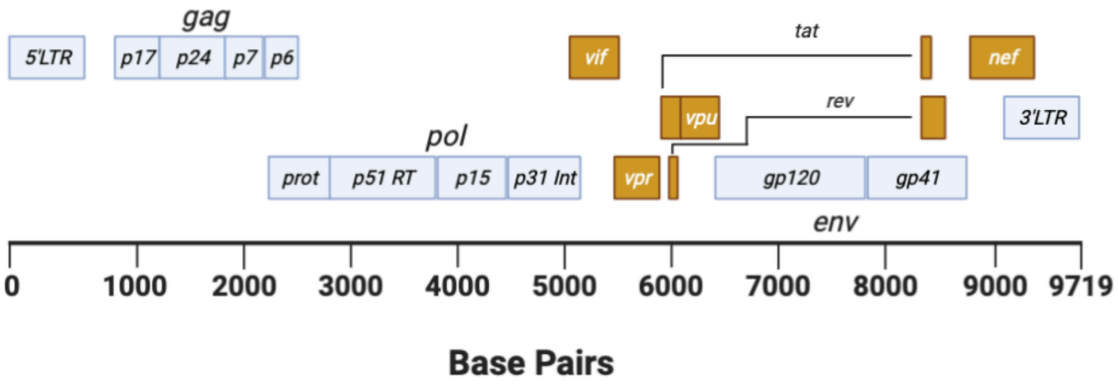


Figure 1-1: HIV-1 Full-length genome

Light blue genes are common retroviral proteins necessary for virion production. Golden yellow genes are accessory proteins expressed by HIV-1 that play a role in regulating parts of the viral life cycle and pathogenesis.

HIV-1 lifecycle

All cells that express CD4 such as macrophages, dendritic cells, T-helper cells and astrocytes, are susceptible to HIV-1 infection[51]. Attachment to target cells can be either non-specific, where Env interacts with negatively charged cell surface heparan sulfate proteoglycans or specific, where Env interacts with α4β7 integrin or pattern recognition receptors (PRRs) such as dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin (DC-SIGN)[52–55]. Regardless of what mediates the attachment, this action brings the Env protein in close proximity to the target receptor CD4 and coreceptor. Upon initial recognition of CD4 to gp120, a sequence of events is triggered that involves a series of conformational changes in gp120. While these events lead to higher affinity CD4 binding, they also result in the exposure of coreceptor binding sites on gp120 and the N-terminal and C-terminal heptad repeat regions of gp41[56]. These structural changes promote further engagement of gp120 to its specific chemokine receptor and induces further conformational changes in the trimer, triggering gp41 activation. With consequent refolding, the activated gp41 mediates viral and target cell membrane fusion and deposits the viral core within the cytoplasm of the target cell[51]. More recently, it has been shown that Env is a dynamic structure that spontaneously transitions between three distinct conformational states: a pre-triggered Env (State 1), a default intermediate (State 2) and a three-CD4-bound conformation (State 3)[57].

There are three models that provide the mechanism of capsid disassembly once the core has entered the cytoplasm. The first model –rapid core disassembly– suggests that the HIV capsid core dissociates almost immediately on viral entry. This is supported by biochemical studies that have failed to isolate CA and conical capsid core from the reverse transcription complex (RTC)[58,59]. However, recent observations have questioned this model. Observations such as the protective role of the capsid core from DNA/RNA sensors and infection of non-dividing cells have reevaluated the timing of the disassembly[60–62]. The second model –cytoplasmic uncoating– suggests that a significant amount of the core disassembles from the viral RTC within the first hour and the remainder traffic towards the nucleus. Evidence from immunofluorescent experiments have shown that cytoplasmic RTCs contain detectable amounts of CA at 4-hours post infection[63,64]. However, this model

is brought into question because the perceived purpose of the CA during infection is to protect the genome from host DNA/RNA sensors and restriction factors. The loss of CA integrity during reverse transcription and trafficking towards the nucleus would abrogate the protective function. It is possible that the protection later during infection is mediated by the RTC or the pre-integration complex (PIC) through the recruitment of other cellular factors in a CA-dependent manner[65,66]. However, this is yet to be validated through experimental work. The last model –cores at the pores– suggests that the CA remains intact until arrival to the nuclear pore complex (NPC) where nuclear import of the viral RTC into the nucleus is concomitant with the uncoating of the CA[67]. Evidence from electron micrographs show intact CA at the NPC and provides an explanation for protection from innate sensors like cGAS[60,68]. However, this model is approached with caution due to issues of timing in various studies.

Upon entry into the cytoplasm, reverse transcription is activated and begins to follow a highly ordered sequence of events to ensure production of DNA[36,69]. Briefly, the host $tRNA^{Lys3}$ binds to the primer binding site (PrBS) near the 5' end of the positive-strand of the (+) RNA and begins synthesis of the negative-strand (-) DNA to the 5' end of the RNA using the bound tRNA as a primer. Concomitantly, the RNase H hydrolyses the RNA that is bound to the newly synthesized DNA in the RNA:DNA hybrid product. As the RT reaches the end, the RT:DNA complex hybridizes with the R region at the 3' end of the (+) RNA and continues to elongate the (-) DNA. The RNase H continues to hydrolyze the RNA bound to the newly synthesized DNA except at two regions (cPPT and 3'PPT) as they are resistant to RNase H. Upon completion of the synthesis of (-) DNA, the PPT serves as the primer for the initiation of positive-sense (+) DNA synthesis. RNase H removes the tRNA and PPT and a second strand transfer between the newly formed PrBS on the (+) DNA strand and the (-) DNA strand completes the synthesis of the (+) DNA. This results in a linear DNA duplex with LTRs at both ends[36,69,70].

Integrase (IN) catalyzes the integration of the viral DNA into the host genome. The two key reactions catalyzed by IN are 3'-end processing and strand transfer events. The former reaction creates a reactive 3' hydroxyl end and the latter reaction uses the reactive hydroxyl end to nucleophilically attack a 5'-a-phosphate in the host DNA. To overcome the

unfavorable charge interaction between the viral DNA and the host DNA and encourage successful integration during this reaction, three evolutionarily conserved amino acid residues within IN (Asp64, Asp116 and Glu152) chelates two Mg+2 ions.[71] After the transfer, there remains junctions with unpaired regions of DNA which is thought to induce a host cellular DNA damage response known as "gap repair". This step is critical as improper repair can result in cellular death and therefore end the lifecycle of the virus. Proteins such as polymerase, nuclease and ligase are thought to extend to fill gaps, remove 5' dinucleotide flaps on the HIV DNA and ligate unbound segments, respectively. This fully integrated DNA is now referred to as a "proviral DNA"[72]. Refer to **Figure 1-2** for a schematic of the lifecycle.

Antiretroviral therapy

At the beginning of the AIDS epidemic, a positive HIV diagnosis was regarded as a certain death sentence due to the lack of available drugs to treat individuals. Fortunately, in the mid 1990's, antiretroviral drugs were developed that were effective at controlling viral replication. This evolved into multidrug combinations that were highly active antiretroviral therapies (HAART), and quickly became the standard treatment for HIV infection within economically affluent countries. HAART effectively prevents the decrease in CD4 T cells and reduces the likelihood of the virus developing resistance to the administered drugs.[73] HAART is administered as a cocktail of three drugs and normally includes two nucleoside reverse transcriptase inhibitors (NRTIs) and a non-nucleoside reverse transcriptase inhibitor (NNRTI).[74] Besides NRTIs and NNRTIs, there are additional classes of drugs included in antiretroviral therapy. These include protease inhibitors, entry or fusion inhibitors and integrase inhibitors (INSTIs). HAART functions by interfering with various stages in the viral replication cycle to actively and efficiently suppress viremia (**Figure 1-2**). Since HAART only targets the replicating virus, the quiescent population of cells that are latently infected are "hidden" from the actions of HAART and the clearing actions of the immune system.[75] Therefore, despite HAART being highly effective at maintaining the virus at undetectable levels, it is not and is unlikely to be a cure initself.[76]

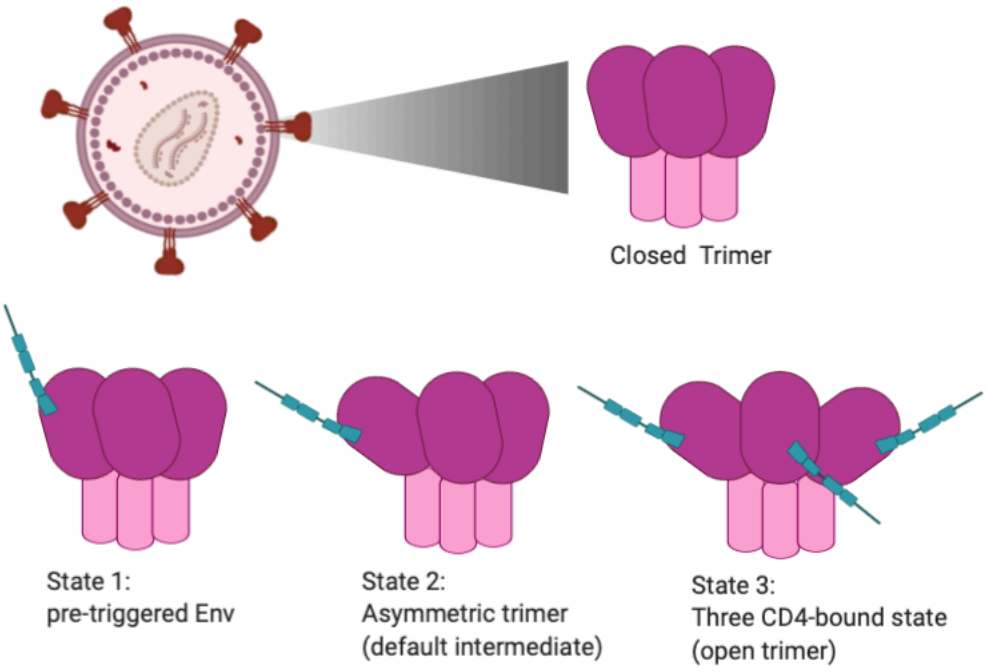


Figure 1-2: HIV-1 Env is a dynamic structure.

The Env structure has recently been shown to transition between three distinct conformational states: a pre-triggered Env (State 1), a default intermediate (State 2) and a three-CD4-bound conformation (State 3). Dark pink, gp120; light pink, gp4l'; green, CD4.

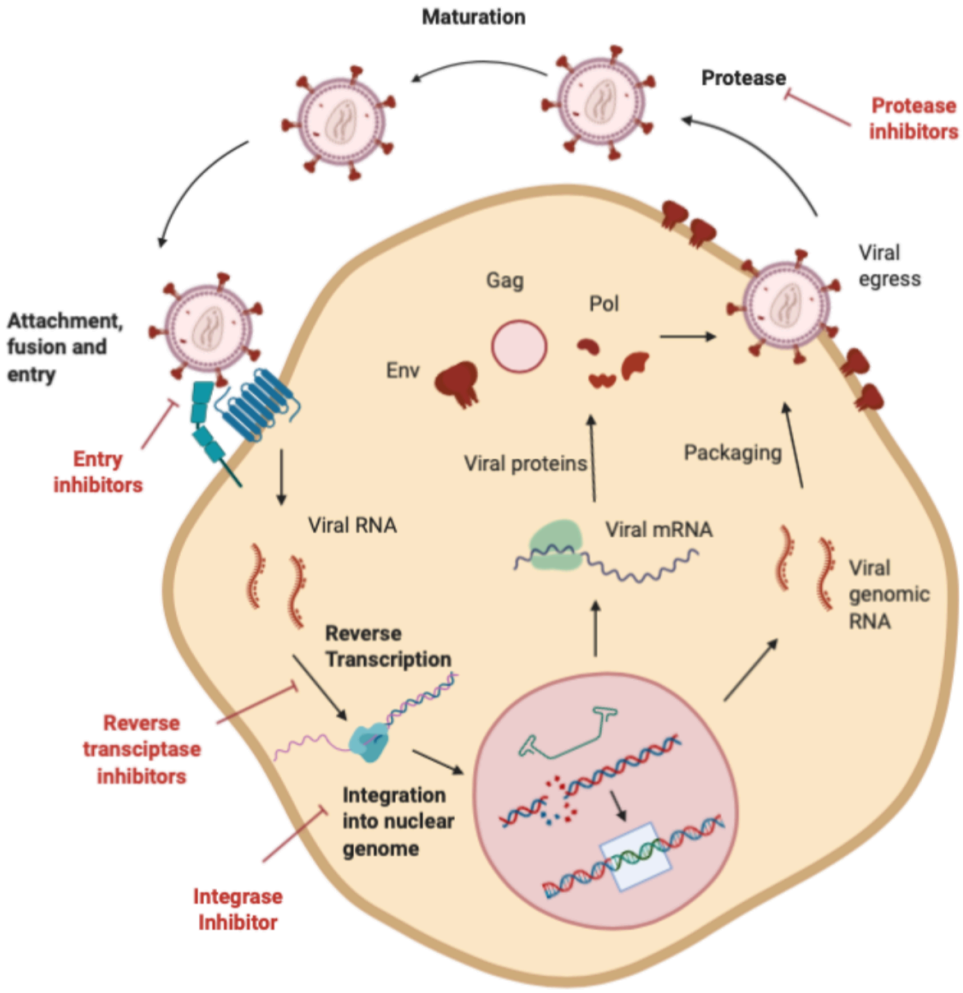


Figure 1-3: Replication Cycle and classes of antiretroviral drugs.

The virus infects cells expressing CD4 and either chemokine co-receptor CXCR4 or CCR5. Upon Env engagement, the virus begins to fuse with the cellular membrane mediating entry of viral particles into the cell. Reverse transcriptase activity converts the viral RNA into cDNA. Following import of the viral cDNA into the nucleus, integrase orchestrates integration of the viral cDNA into the host genome to create an integrated provirus. Proviral transcription mediated by host RNA polymerase II produces viral mRNA that produce the necessary viral proteins to create progeny virus. During the late stage of infection, Rev-dependent export shuttles full-length of viral RNA into the cytoplasm. Viral capsid formation concomitantly packages the full-length viral RNA and progeny virus begin to bud from the cell. Protease within the viral particles begins to process the packaged viral proteins and the viral particles mature, ready to infect neighby target cells.

HIV-1 depletion of CD4 T cells leads to an immunocompromised state

The hallmark of an HIV-1 infection is the depletion of CD4 T cells and eventual disease progression to AIDS. As a result, the individual succumbs to opportunistic infections that they should otherwise be able to fend off[77]. However, not all CD4 T cells serve as the target for HIV-1 entry and infection[78]. This has to do with the metabolic status of the CD4 T cells. Compared to the naïve counterparts, an activated CD4 T cell is more susceptible to infection[79]. A naïve CD4 T cell has not interacted with its target antigen in the context of MHC with co-stimulation and is therefore transcriptionally silent. On the other hand, an activated CD4 T cell is prepared with an arsenal of transcription factors and has undergone various histone modifications to ensure high levels of transcription of genes associated with inflammation[78]. As mentioned earlier, a myriad of host transcription factors, along with viral proteins, are required for successful completion of the HIV lifecycle (such as reverse transcription).

Originally, it was believed that the dramatic decline of CD4 T cells was due to cytopathic effects mediated by the virus[80], however subsequent studies demonstrated that the decline was attributed to caspase-1 dependent pyroptosis in non-permissive CD4 T cells (ie. naïve)[81,82]. The accumulation of incomplete RT DNA products within these cells is detected by DNA sensor IF116 resulting in the assembly of an inflammasome[82]. While this accounts for 95% of the decline in CD4 T cells, the remaining 5% comes from permissive CD4 T cells (ie. activated) in which productively infected cells undergo caspase-3-mediated apoptosis[82].

In addition to disruption of T cell symbiosis, HIV-1 infection delays plasma antibody response in B cells of GALT. Histopathological changes in secondary lymphoid structures such as Peyer's patches, display decreased numbers of B cell germinal centers and increased follicular lysis– leading to a shift from a naïve phenotype to a memory phenotype in the terminal ileum and serum compartment[83]. This B cell dysfunction results in a decrease in HIV-induced antibody response and thus an impaired humoral response directed towards the virus.

HIV-1 transmission

Nearly 70% of HIV-1 transmission events occur through sexual intercourse where there is contact with contaminated blood or semen[84]. The other 30% of the infections are caused by alternative, less frequent, routes such as anal intercourse, vertical transmission or through intravenous inoculation. During sexual transmission, the most important factor that increases risk of infection is the viral load[85]. The stages of infection are divided into early/acute, acute and chronic infection, where the nomenclature is based on the length of infection. At the early/acute to acute stage of infection, the plasma has a high viral load which increases chances of infection[86]. An additional factor that drives HIV-1 transmission includes the concomitant presence of other sexually transmitted disease (notably genital ulcers, herpes simplex-2 infection and bacterial vaginosis)[20,84,87]. Behavioral factors such as multiple sexual partners, stigma, discrimination and punitive laws against high-risk groups (MSM and IV drug users) can also act as drivers to HIV infections[88].

Crossing the mucosal barrier

During each inoculum, the recipient is exposed to millions of viral variants from the donor. However, in 70-80% of transmission events, only a single virus is found to be able to successfully establish a systemic infection[89] **(Figure 1-4)**. In generating a systemic infection, the viral swarm that the recipient is initially exposed to encounters a dramatic genetic bottleneck. A myriad of biological and viral factors can influence the transmission rate such as virus strain, inoculum, epithelial trauma during intercourse, concomitant anogenital infection, circumcision, and the use of contraceptives[90–92]. As such, the mechanism of HIV-1 entry will be affected by the physiological and anatomical state of the vagina and ectocervix in women, the glans penis and inner foreskin in men as well as a number of other factors[92]. There are three potential routes for HIV-1 translocation across the mucosal-epithelium barrier: (1) virus penetration across epithelium; (2) virus transcytosis by epithelium cells; or (3) trans-epithelial emigration of infected Langerhans cells (LC)[92–94]. The first two routes are unlikely as epithelium cells are not susceptible to infection[90]. On the contrary, an intact genital mucosal barrier will provide protection by ensuring minimal crossing of virus[92]. Anything that results in the weakening or loss of this protective layer –such as co-infections or damage due to sexual intercourse– may enhance

transmission[90,91]. In particular for vaginal transmission, it has been identified that vaginal epithelial LC and dendritic cells (DC) are the major viral targets for initial infection[92]. While the main target for HIV Env is the CD4 receptor and chemokine receptors CCR5 and CXCR4, there are also non-specific interactions mediated through C-type lectin (CLR), adhesion molecules, complement receptors, Fc receptors and heparin sulfate proteoglycans[90,95–97]. While it is clear that HIV-1 interacts with DC through these aforementioned receptors, its interaction with LCs is primarily mediated through langerin, a LC-specific CLR that has been shown to bind HIV gp120 with a high affinity, CD4 and CCR5[92,95,98]. DC play a primary role of recognizing, interacting and presenting antigens to CD4 T cells, however the anatomical location within the mucosal epithelium and the ability of LC to migrate to regional lymph nodes present a much more likely mode of transmission for HIV to infect its target cell[99]. While macrophages, DC and T cells express CD4, CXCR4 and/or CCR5, LC play a more significant role in HIV transmission because of the following characteristics: (1) location within the anatomical region of the vagina; (2) constitutive expression of CD4; (3) susceptibility for R5 tropic viruses; (4) ability to migrate to lymph node; and (5) ability to transmit HIV to its target cell type[92].

Through the use of non-human primate models (NHP), a general timeline of the infection has become well-understood. From the donor's genital secretion, the recipient mucosal barrier is swarmed with a genetically diverse quasispecies of virus. Within hours, the virus is able to cross the mucosal epithelium and enter the lamina propria where it establishes founder populations. These founder populations begin to rapidly expand and within days to a week, the virus is able to spread into the local lymph nodes and become a self-propagating infection[100]. The virus then becomes systemic as it spreads to secondary lymphoid tissues, brain, liver, lungs and gut. Within these different organs, HIV is responsible for a myriad of complications such as severe HIV-associated dementia when present in the central nervous system, or chronic inflammation within the GALT[83]. The impact of HIV-1 on GALT is significant because, as the largest immunological organ in the body, it houses 60% of the total body lymphocytes (i.e. a hotspot for HIV replication)[83,101]. Compared to the serum compartment and other lymphoid organs, the mucosal CD4 T cells display activated effector and central memory ($CD45R0^{+}$) phenotypes and express $CCR5^{hi}$ on the surface. These lymphocytes are present within distinct sub-

compartments of the GALT such as the Peyer's patches in the small intestine, lymphoid follicles in the colon and the lamina propria. Due to the hostile environment created by constant exposure to a plethora of antigens (mainly driven by the gut microbiota) the lymphocytes within the GALT are in a constant state of "physiological inflammation[83]." Due to their high degree of susceptibility to infection by HIV, these cells undergo an immense percentage of overall death mediated by direct viral infection, activation-induced cell death and host-derived cytotoxic responses[83]. Interestingly, despite the high degree of death of CD4 T cell in the GALT, this loss is not sufficient to lead to AIDS. Studies using non-pathogenic simian immunodeficiency virus in NHP have shown that despite the loss of CD4 T cells, there is no official diagnosis of AIDS[83,102]. This may be due to a physiological response that provokes proliferation of CD4 lymphocytes in the serum compartment, in an attempt to sustain overall lymphocyte population[83]. However, this response simply delays the inevitable prognosis of AIDS. The necessary pro-inflammatory milieu created by Th17 cells is balanced with anti-inflammatory functions by regulatory T cells which together orchestrate the CD4 T cells' responses in the gut. The imbalance caused by HIV-mediated depletion of lymphocytes in the gut disrupts the mucosal barrier, which encourages microbial translocation leading to chronic immune activation. The chronic inflammation facilitates downstream activation of cells, further feeding into productive infection of HIV virions in target cells.

From a genetic swarm to a single viral variant

HIV is spread primarily by sexual intercourse, through which the virus is capable of passing the mucosal barrier to infect target CD4 T cells[103]. While the recipient is exposed to millions of viral variants from the donor, in 70-80% of infections, a single virus is able to successfully establish a systemic infection[89]. In generating a systemic infection, the viral swarm that the recipient is initially exposed to encounters a dramatic genetic bottleneck. Work published by Klein *et al.* demonstrates that early HIV-1 isolates within the blood displayed a more homogenous genotype than clones in the female genital tract. Their analysis of early mucosal infection in women revealed that despite the high diversity of HIV-1 in the vaginal tract, a possible mucosal sieve effect may select for a single variant to establish a homogenous systemic infection[104]. Similarly, Williams-Wietzikoski *et al.*

analyzed viral sequences in semen and plasma –in the case of male-to-female transmission– and reported seeing significant genetic differences between viral RNA isolated from seminal plasma and blood within certain males in their transmission pairs[105]. This in turn suggests the occurrence of unique viral variants present in compartments relevant to transmission via sexual intercourse. Evidence for the success of these Transmitted/founder (T/F) variants is heavily influenced by viral fitness inferred from sequence conservation, viral subtype, gender, viral load, presence of other sexually transmitted infectious agents and the integrity of the immune system[106–112].

Phenotypic Properties of Transmitted/founder (T/F) viruses

In particular, the success of a T/F is highly attributed to phenotypic differences within the envelope (Env) glycoprotein. While the specific phenotypic differences vary between subtypes, it is well accepted that there are unique characteristics in the Env of T/F virus compared to Env found on virus at later time points of infection. Work by Derdeyn *et al.* has shown that T/F viruses of subtype C share two unique properties –shorter V1-V4 regions and a significantly lower number of potentially N-linked glycan sites (PNLGs)– that may have allowed the variants to transmit more easily, and successfully establish infection[113]. Similarly, a study in Kenya looked at early infection sequences in subtype A and noticed that they had significantly shorter V1-V2 loop sequences and fewer PNLGs[114]. On the contrary, subtype B T/F virus have been shown to have higher Env expression and incorporation of Env on virions[114]. While each subtype may have its own unique feature, it is well appreciated that T/F viruses have features that may enhance infection. With majority of the transmission events occurring due to sexual intercourse, a major goal for an HIV-1 vaccine is to prevent acquisition at mucosal surfaces. Ideally, targeting the viral variant that causes the systemic infection may be a more effective vaccine formulation. As such, understanding T/F viruses will provide essential insights onto HIV-1 transmission and novel vaccine designs.

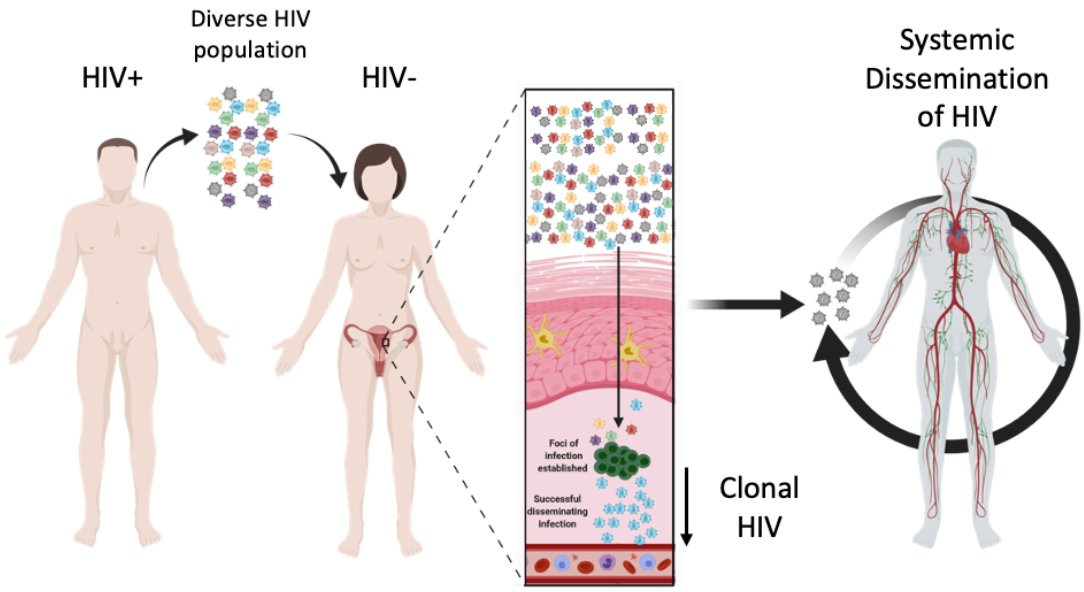


Figure 1-4: HIV-1 Transmission experiences a genetic bottleneck.

During sexual transmission, the recipient is exposed to millions of genetic viral variants. In 70-80% of infections, one viral variant is able to cross the physical mucosal barrier, overcome immune mediated clearance and establish the systemic infection. This viral variant is called the Transmitted/founder (T/F) virus. Overtime, the T/F virus begins to diverge from the ancestral strain and create a diverse pool of HIV variants.

Immune response to HIV-1

During infection, the immune system initiates a very strong response against HIV-1. Studies of infection in non-human primates (NHP) have provided insight on critical early events during transmission[115]. Within hours of exposure, the virus is capable of crossing the mucosal epithelium and establishing a population of infected cells, called a founder population[115–117]. This population begins to rapidly expand and evolve into a larger foci of infection that eventually becomes systemic[115–117]. In more than 75% of infected individuals, a single viral variant, referred to as the T/F virus, is enough to establish a systemic infection[118,119].

Early in infection, the immune system and the virus enter an evolutionary arms race whereby the immune system attempts to counter the high mutational burden mediated by HIV RT as the virus begins to proliferate into genetically unique variants. Peak viremia is achieved 10-15 days post-infection and declines to viral set point after 30 days. The dramatic inflection point in the HIV population count is mediated by a rapid amplification in CD8 T cells. However, HIV is able to constantly evade the evolving cytotoxic T lymphocyte (CTL) pressure and thus circumvent immune-mediated clearance.

Prior to peak viremia, there is a burst in low-affinity B cell plasmablasts —short-lived extrafollicular plasma cells with minimal immunoglobulin (Ig)-variable region diversification[120]. These B cell plasmablasts can account for over 50% of all circulating B cells, but only ~1.5% of these are thought to be HIV-specific[120]. Proliferation of B cells requires the binding of an antigen to its B-cell receptor (BCR) or via cognate interactions with T cells that present peptides on the major histocompatibility complex[121]. The antigen presented by CD4 T follicular helper cells (Tfh) is significant because it positively selects for B cells with the highest affinity for that antigen and provides them with signals for survival and proliferation[120,122–125]. Co-stimulatory factors can promote B cell proliferation, Ig class switching and somatic hypermutation, and can guide differentiation[126]. The activated B cells either become short-lived plasma cells that produce low-affinity IgG or enter a lymphoid follicle where they form a germinal center. Those in the germinal center alternate between the dark and light zones, leading to the production of both high affinity memory B cells and long-lived plasma cells[122]. As the virus acquires mutations, certain B

cells will also mutate and proliferate to keep up with the genetic diversity of the virus. The degree of selection and clonal expansion of B cells depends on the relative affinities of competing clones for the most prevalent antigen[122,127]. As the virus acquires immune escape mutations in Env, the affinity of cognate antibodies will weaken.

The early antibody response produced during an HIV infection has shown to be polyreactive and are directed to the gp41 glycoprotein before gp120[100]. As such, the immunodominant nature of gp41 elicits production of antibodies that fail to neutralize the infecting virus. Interestingly, an additional hypothesis to the predominant gp41 antibody-driven response may be due to the cross-reactive priming by the microbiome in the gut. More specifically, anti-gp41 IgA antibodies are present about 13.5 days post-infection followed by anti-gag IgA antibodies 25 days post-infection[100].

In turn, this clinical timeline was adapted to Fiebig stages whereby specific laboratory-defined stages of HIV infection were used to further elucidate the time course of infection. These stages are defined as followed: Fiebig stage I is when HIV RNA is first detected; Fiebig stage II is when HIV p24 is detected along with HIV RNA; Fiebig stage III is when HIV Ab is detected via ELISA but negative for p24 on Western blot; Fiebig stage IV is positive for Ab but indeterminate on Western blot; Fiebig stage V is positive for Western blot (but missing the p31 band); and finally Fiebig stage IV is positive for Western blot and includes the p31 band.

Production of broadly neutralizing antibodies (bnAbs)

Over the years, the field began to acknowledge the contribution of an arms race between the evolving virus and the adapting immune system, in the development of broadly neutralizing antibodies (bnAbs). While all individuals generate autologous neutralizing antibodies, after 2-4 years, only 20-50% of those individuals will generate bnAbs[127–130,131]. This small subset of antibodies can neutralize various primary isolates because they have the ability to bind to epitopes on Env that are highly conserved[132] –sites that are functionally important for the virus and therefore struggles to change/dispense with. The HIV-1 trimer complex has five such sites that are vulnerable to bnAb recognition: the CD4 binding site (CD4bs) of gp120, PG9 and PG16 recognition site in gp120, V3 loop, the

membrane proximal external region (MPER) in gp41 and the gp120/gp41 interface region[133]. bnAbs possess unique features that allow them to neutralize various viral strains; they are highly mutated, especially in the framework region[134], and they tend to have abnormally long and charged immunoglobulin heavy complementary-determining regions (CDR; region that binds to the antigen)[135]. The mutations introduced in bnAbs are a result of somatic hypermutation (SMH)[136]. Interestingly, there is a positive correlation between the number of mutations and the potency of bnAbs[137,138]. The high level of SMH and the late emergence of these bnAbs suggests years of chronic antigenic exposure, selection and mutations to develop such potent antibody lineages. Moreover, the extended CDR enables the paratope of the antibody to penetrate deep into conformationally hidden and/or heavily glycosylated sites for contact with its epitope[139,140–142]. As a consequence of these features, some bnAbs are polyreactive[135]. This may explain the relatively low abundance of bnAbs in HIV-1-infected individuals, as viral mimicry to self-antigens can delete or tolerize autoreactive B cell clones, thus impeding the development of bnAbs[135]. Constant evolution of bnAb lineages can force strong divergence from their germline ancestors. As a result, these ancestors usually fail to interact with concealed and conserved epitopes of their progeny target after years of evolution[143,144]. Nonetheless, how these bnAbs are generated or why they are only generated in less than 50% of HIV-infected patients are questions that remain unanswered. A prevailing theory surrounding the generation of bnAbs emphasizes the interplay between the dark and light zones in the germinal center and SMH[145]. It is thought that the frequency of bnAb-target sites on Env may be high in a select few patients and thus select for such high affinity B cells[146]. As these B cell clones are selected, they undergo extensive SMH to allow the antibody to evolve simultaneously with the virus[147,148].

The innate immune response and HIV-1

The selective pressures that shape the HIV-1 bottlenecks and lead to successful transmission of T/F virus can occur at different stages in the transmission cycle – such as at the site of transmission within the donor, during the transmission process between the donor and recipient or within the recipient upon exposure to the viral swarm from the donor[149]. Since the stochastic and selective pressures will be different for each donor and

recipient, it is therefore unlikely that a single phenotype or genetic sequence will inherently select for a successful T/F virus. Many biological factors beyond the control of the virus, such as pro-inflammatory factors, have been implicated in enhancing transmission of HIV-1. While proteins can activate various pro-inflammatory signal cascades, viral RNA has been shown to mediate similar effects. RNA has been implicated to play an essential role in the evolution and regulation of DNA/protein-based life. This includes, but is not limited to: protein coding, sequences with affinity to target proteins that can regulate the proteins' functions, and secondary structures that play a role for virus function like ribosomal entry structure and internal ribosomal entry site[150]. In response, our innate immune system has evolved to identify viral RNA sequences and initiate a cascade of pro-inflammatory and antiviral responses. Toll-like-receptors (TLRs) and other innate pattern recognition receptors like RIG-I play a crucial role to identify viral RNA (such as through unmethylated CpG motifs on the RNA) and induce production of antiviral responses like IFN-α to combat the virus[151,152]. More specifically, TLR7 and TLR8 recognize ssRNA and initiate a signal cascade that eventually produces IFN-α and pro-inflammatory cytokines like TNF-α, IL-6, and IL-12[153,154]. Creating a pro-inflammatory environment at the site of infection is thought to be beneficial as it can activate macrophages, induce apoptosis and recruits immune cells, all of which play a role in fighting and clearing the infection[155].

A combination of pro-inflammatory cytokines, chemokines and type I and II IFNs play an essential role in host defense against HIV. Recognition of HIV-1 by TLR7 requires attachment and endocytosis of HIV-1 but does not require the downstream infection stages[156]. As the virus is engulfed, it is shuttled into an early endosome where viral RNA induces IFN-α stimulation via TLR 7 signaling in pDCs[157]. The significance of viral attachment comes from studies that have compared nonfusogenic Env with WT Env and demonstrated that production of IFN-α in pDCs precedes fusion of the virus[158]. Without the correct attachment on the surface and subsequent fusion, pDCs elicit less IFN-α. This also indicates the necessity of virus entry and not replication in the activation of TLR7[158,159]. While the stimulation of TLR8 would initiate a protective response, ironically, stimulation by HIV-1 fosters replication within DCs[160]. If the virus is successfully able to integrate into the host genome, subsequent stimulation of TLR8 rather initiates

transcription of HIV by recruiting RNA polymerase[160]. In concomitant with signals from DC-SIGN, HIV-1 is able to use the immune sensors as a fuel to drive production of viral proteins. However, the response initiated by TLR8 within monocytes and macrophages is rather protective than beneficial for HIV-1. Upon recognition of HIV-1 in monocytes and macrophages, TLR8 promotes the secretion of IFN-α and pro-inflammatory and regulatory cytokines such as CXCL10 –a key protein involved in the recruitment of additional immune cells to the site of infection[161]. Furthermore, the ssRNA of HIV contains many uridine-rich ligands that potently signal TLR7/8 resulting in high titers of IL-18 and IL-1β production, further adding to the inflammatory niche[154,161]. Monocytes and macrophages within an HIV-1-infected niche also activate the NLRP3 inflammasome which leads to activation of casapase-1 and gasdermin D-mediated pyroptotic cell death[162]. Although it is an inflammatory form of programmed cell death, it promotes the rapid clearance of HIV-1 infection by removing intracellular replication niches and enhancing host response by recruiting additional immune cells and contribute to inflammation within the tissue to clear the infection.

However, as these innate sensors are activated by HIV-1, it has slowly become evident that majority of these are either disabled or are changed to bias production of proteins to enhance infection. HIV-1 is able to support its infection during the acute stages through the expression of Vpu –an accessory protein expressed by HIV. Doehle B *et al.* were able show a molecular interaction between Vpu and IRF3 in infected cells and demonstrate the shuttling of IRF3 by Vpu to the endolysosome for proteolytic degradation[163]. As a consequence, HIV-1 is able to repress production of IFN-β and evade anti-viral mechanism hoisted by the immune system. Additionally, the drastic rise in IFN-α mediated by TLR7 in turn induces expression of hundreds of ISGs in surrounding cells that contributes to HIV-1 restriction[158]. Paradoxically, type I IFNs secreted by pDCs in the cervical mucosa aid in HIV-1 replication. Activated pDCs beneath the cervicovaginal epithelial cells recruit CCR5+ CD4 T cells through the secretion of type I IFNS and chemokines CCL3 and CCL4[158]. Interestingly, high levels of IFN are found during the early/acute stage of infection and overtime viremic patients have low levels of IFN in response to HIV-1[158]. While the high IFN response initially may limit the spread of HIV-1 systemically, the

response within the cervicovaginal environment recruits more target cells for HIV-1. As the immune system orchestrates an inflammatory response to clear the infection, HIV-1 in turn utilizes that to enhance replication and establish a systemic infection.

Tumor necrosis factor alpha

One of the pro-inflammatory molecules that is induced upon TLR7/8 activation is TNF-α, a trimeric protein. TNF is considered a pleiotropic cytokine as many cells in the body can induce its production; however, production of this protein is dominated by activated macrophages and pro-inflammatory T cells at the site of infection[164]. First discovered by Carswell *et al*. in 1975, subsequent studies a decade later showed that TNF is a member of a large superfamily known as TNF/TNFR superfamily[164]. The primary produces of TNF-α are cells of the monocytic lineage –macrophages, astroglia, Langerhans cells, Kupffer cells and alveolar macrophages[164,165]. There are two receptors that TNF-α binds to mediate its effects –TNFR1 (p55 or p60) and TNFR2 (p75 or p80)[165]. Both of these receptors contain cysteine-rich repeats in the extracellular domain that interact with the trimer ligand. Upon binding, the receptor:ligand complex is internalized via clathrin-coated pits to be degraded in secondary lysosomes[165]. Unlike TNFR1, which is constitutively expressed in most cells, TNFR2 is found mainly on immune cells[164,165]. This cell-specific expression pattern results in differences in function. Activation of TNFR1 leads to cytotoxicity, cell growth, NF-κB activation and upregulation of adhesion molecules and cytokines. On the other hand, activation of TNFR2 leads to proliferation of lymphoid cells and in some immune cells, cytotoxicity and NF-κB activation. Both of these receptors are necessary to mount a proper immune response as independent knock-out studies of TNF1 and TNF2 in mice have shown failure to clear bacteria and poor TNF-induced necrosis in T cells, respectively[166,167].

TNF-α plays a significant role in immunostimulation, resistance to infection and resistance to tumors. In experimental challenges with leishmaniasis, insufficient TNF-α was correlated with disease progression indicating the necessary role of TNF-α in modulating the immune system[168]. The upregulation of adhesion molecules mediated by TNF-α

contributes to resistance to infection by recruiting neutrophils and platelets, enhancement of macrophage and NK cell killing and overall activation of the immune system.

Remarkably, HIV-1 may have evolved to make use of this pro-inflammatory niche in aiding the establishment of infection. Work by Marein *et al.* showed that TNF-α increases HIV-1 replication in Langerhans cells (LCs) whereas Pam3CSK4, a TLR1/TRL2 ligand, enhanced HIV-1 capture within this subset of cells[169]. Furthermore, TNF-α has been implicated in enhancing HIV-1 replication by creating an inflammatory niche that attracts DCs, macrophages and T cells[169–173]. Ironically, while this pro-inflammatory niche arms the immune system to fight the infection, it creates the ideal opportunity for HIV-1 to infect activated CD4 T cells and establish a foci of infection.

Envelope Glycosylation: an immune evasion tactic

HIV-1 Env spikes are shielded from potentially neutralizing antibodies by three main mechanisms: glycan shielding[174], conformational masking[140,175] and sequence variation[176–178]. The Env glycoprotein is heavily glycosylated, with N-linked glycans accounting for as much as 50% of the molecular weight of the protein[179,180]. On average, gp120 contains ~25 putative N-linked glycosylation sites (PNLGs): ~4 located in the inner domain, ~7-8 in the V1/V2 and V3 loops and the rest in the outer domain of gp120[181–183]. These N-linked glycans restrict antibody access to 97% of the Env surface[184]. This poses a challenge for the generation of a vaccine against HIV-1, as vulnerable immunogenic sites on Env remain hidden.

The presence of oligosaccharide constituents of gp120 were identified and described soon after the discovery of HIV-1. Montefiori *et al.* discovered the contributing factors of oligosaccharides to virus infectivity and immune evasion[185]. Studies thereafter began to examine the effect of glycan removal on viral infectivity and Env structural effects, using single and combinatorial site mutants[183,186,187]. As anticipated, viral infectivity was impaired in mutants with the loss of multiple glycosylations, as it significantly impacted the native 3D conformation of the trimer. Furthermore, Env glycosylations affect the generation of an effective immune response against HIV-1. Development of potent bnAbs requires constant interaction with the antigen and high levels of SMH. However, the glycan

shield forms a non-immunogenic cloak, protecting the underlying protein surface[188,189] **(Figure 1-5)**. Since these glycans are synthesized in the endoplasmic reticulum of the host cell, they are generally less immunogenic and begin to present the HIV-1 Env trimer as an immunologically silent protein[190]. Although the glycosylation pattern on Env has been shown to be dynamic, consensus sites on gp140 have been mapped. Glycosylation profiling of these sites can guide vaccine developers in determining the quality of glycosylation on their immunogen. Differences in Env immunogenicity have been observed when only surface glycosylation is modified. Thus, immunogens based on mature Env will not likely be able to activate the target germline-B cells[191,192]. Env-based immunogens that first select for bnAb precursors, and then continuously select for the subsequent iterations in the lineage, may help to elicit bnAbs upon vaccination[193–196].

The CD4 binding site (bs) and glycosylation

Env sequence variation plays an important role in the modification of glycosylation patterns, by adding or removing Asparagine (Asn) amino acids. Mutations that add Asn to a specific region in Env allow for the presence of PNLGs whereas deletions of already present Asn can remove glycans. Although these changes enable the virus to create immune-escape variants, they must not be deleterious to its survival. The CD4 binding site (bs) is an example of a region that cannot be heavily modified, as it will directly impact the ability of the virus to engage with its receptor. As a result, the virus mutates regions around the CD4bs to circumvent bnAbs which target this site for virus neutralization.

Zhou *et al.* and Liang *et al.* have identified PNLGs surrounding the CD4bs that upon removal, significantly increase VRC01 bnAb binding. These sites include: N197, N276, N301, N363, N386 and N486[197,198]. Upon removal of the putative N-linked glycosylation's, the binding affinity and neutralization breadth of VRC01 dramatically increases. This suggests that glycans surrounding the CD4bs serve to increase steric hindrance and limit the ability of CD4bs-directed bnAbs to successfully bind or remain bound. While most sequence variations involve nonsynonymous mutations of Asn to Alanine (Ala), a specific mutation at site 197 requires a substitution with glutamine (Gln). Liang *et al.* were able to demonstrate that removal of the putative N-linked glycan-binding site 197, had little effect on the overall structure of the Env trimer. More importantly, they

revealed a steric clash to the CD4bs in the presence of the glycan which destabilizes certain Env-antibody complexes –such as Env-VRC01. The presence of Gln in place of Asn serves two purposes: (1) exposes the CD4bs and (2) increases flexibility of the V1/V3 region, leading to increased V3 exposure. This makes the Env trimer susceptible to both CD4bs-directed bnAbs and V3-specific nAbs[198]. Collectively, future Env immunogens lacking certain PNLGs can be used to immune-focus necessary structural components to increase the chance of culminating a bnAb response.

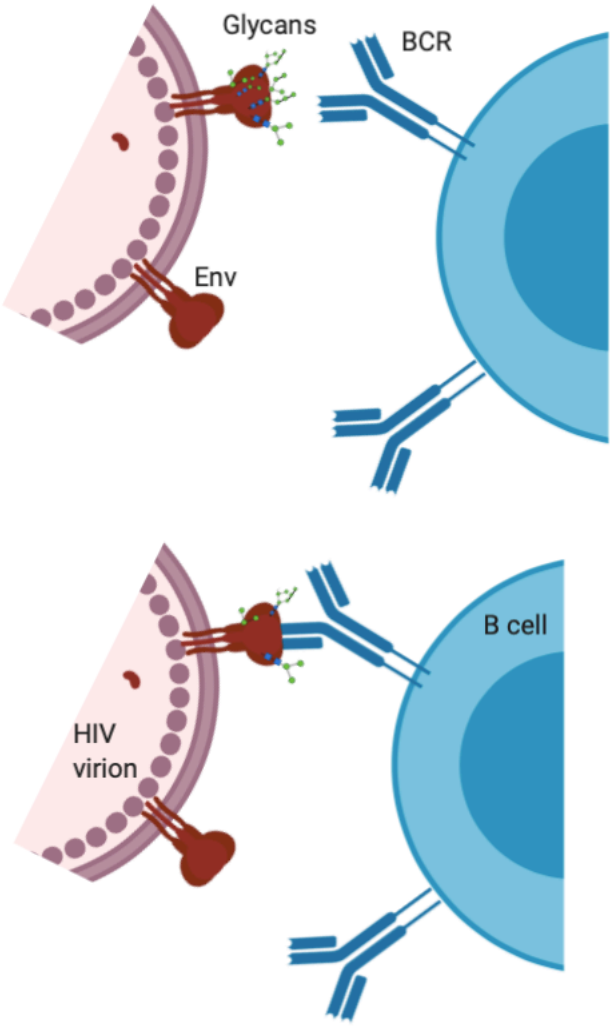


Figure 1-5: N-linked glycans on HIV-1 Env spikes mask vulnerable immunogenic epitopes from B cell receptors.

The top panel demonstrates a B cell receptor (BCR) on a B cell experiencing steric hindrance from the N-linked glycans and as such, fails to bind to its target epitope on the Env spike. As a result, this impairs proper B cell education and thus poor evolution with the virus. The bottom panel demonstrates a lack of glycans surrounding the target epitope and as a result, successful binding of the BCR. This in turn encourages B cell evolution with the virus and results in the development of bnAbs.

Developing a novel HIV-1 vaccine antigen

The goal of this MSc project is to design an efficacious vaccine that can elicit a bnAbs response against HIV-1 (**Figure 1-6**). As successful HIV-1 transmission is normally caused by a single viral clone from within the quasispecies, it stands to reason that the ability to target immune responses towards that particular T/F virus may lead to a more effective vaccine than targeting all the clones of a highly divergent viral swarm. To achieve this, we required the ability to identify and select viruses that could have T/F potential. The identified T/F sequences would then be used to select antigens from which to make a vaccine. As a way to screen viral sequences for potential T/F viruses that could be used in vaccine development, we developed an *in-silico* coding script that can simultaneously identify and enumerate various 4-mer motifs within thousands of viral input sequences. We hypothesized that the presence of the various 4-mer sequence motifs within the HIV genome, previously identified to induce TNF-α secretion in primary human cells, may contribute to or enhance HIV-1 transmission. Concurrently, we are performing preliminary proof-of-concept studies on eliciting antibody responses against the CD4bs, a site of known vulnerability to bnAbs. Here we have been removing glycans peripheral to the CD4bs in order to reduce any steric hindrance towards antibody binding. The aim being to identify which glycans maximally expose the CD4bs to antibodies. The overall aim would be to then combine antigens selected from the *in-silico* analysis with glycan deletion studies to develop a novel HIV-1 vaccine that elicits neutralizing antibodies against T/F viruses.

Therefore, the specific aims for my MSc project are as follows:

Specific Aim 1:

1. Identify the distribution of TNF-α, IFN-α and CpG pro-inflammatory motifs within the genome of HIV-1
2. Analyze viral RNA sequences from the serum compartment of transmission pair cohorts to understand the potential difference in pro-inflammatory motifs within T/F viruses

3. Analyze viral RNA sequences in different anatomical compartments (semen, cervix, serum) to understand potential differences in pro-inflammatory motifs within different compartments

Specific Aim 2:

4. Characterize the effect of removing specific glycans surrounding the CD4bs by assessing the binding ability of bnAbs directed towards this epitope.
5. Introduce the most effective glycan deletion to candidate T/F based on any unique RNA features.
6. Perform a sequential vaccination protocol comparing WT to glycan mutant VLP formulations to evaluate CD4-bs-directed Ab lineage selection and overall Ab efficacy.

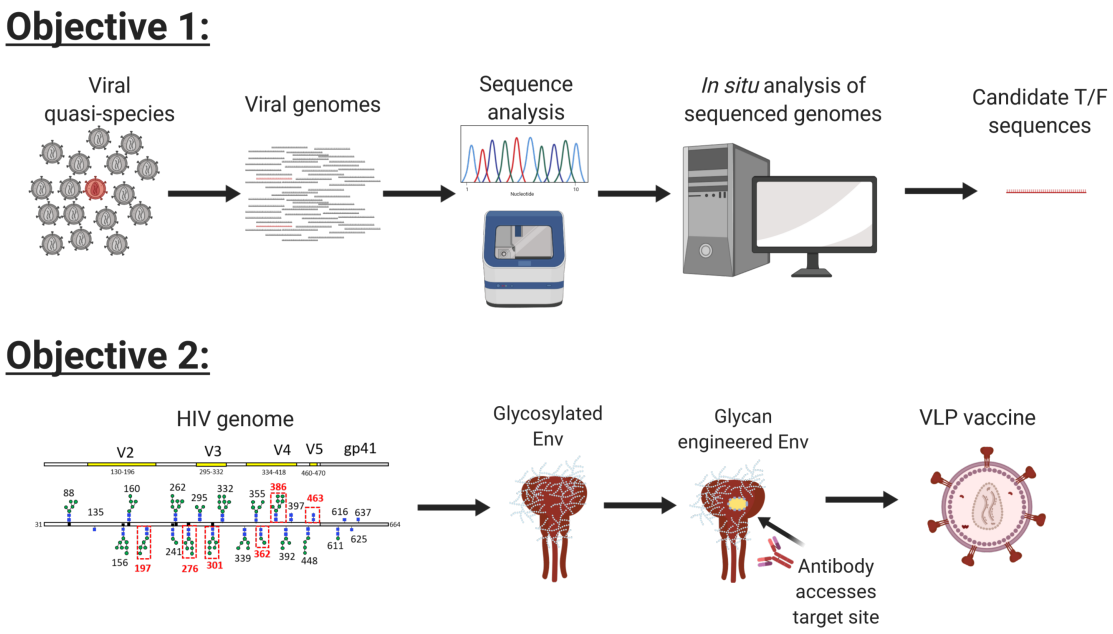


Figure 1-6: Schematic explaining the two objectives to the MSc Project.

The goal of Objective one is to analyze various RNA sequences in transmission pair cohorts to understand potential differences in pro-inflammatory motifs within T/F viruses. To do this, an *in-situ* analysis of thousands of sequences was performed to understand the pro-inflammatory composition within the sequences. The goal of Objective two is to remove glycans peripheral to the CD4bs in order to reduce any steric hindrance towards antibody binding. The overall aim would be to then combine antigens selected from the *in-situ* analysis with glycan deletion studies to develop a novel HIV vaccine that elicits neutralizing antibodies against T/F viruses.

References

1. Pneumocystis Pneumonia --- Los Angeles. Available at: https://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm. (Accessed: 9th January 2020)
2. Greene, W. C. A history of AIDS: Looking back to see ahead. *Eur. J. Immunol.* **37**, (2007).
3. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841–a006841 (2011).
4. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR. Morb. Mortal. Wkly. Rep.* **30**, 305–308 (1981).
5. Jaffe, H. W. *et al.* National case-control study of Kaposi's sarcoma and Pneumocystis carinii pneumonia in homosexual men: I. Epidemiologic results. *Ann. Intern. Med.* **99**, 145–151 (1983).
6. De Cock, K. M., Jaffe, H. W. & Curran, J. W. The evolving epidemiology of HIV/AIDS. *AIDS* **26**, 1205–1213 (2012).
7. Possible transfusion-associated acquired immune deficiency syndrome (AIDS) - California. *MMWR. Morb. Mortal. Wkly. Rep.* **31**, 652–654 (1982).
8. Update: acquired immunodeficiency syndrome (AIDS)--United States. *MMWR. Morb. Mortal. Wkly. Rep.* **33**, 661–664 (1984).
9. Selik, R. M., Haverkos, H. W. & Curran, J. W. Acquired immune deficiency syndrome (AIDS) trends in the United States, 1978-1982. *Am. J. Med.* **76**, 493–500 (1984).
10. Harris, C. *et al.* Immunodeficiency in Female Sexual Partners of Men with the Acquired Immunodeficiency Syndrome. *N. Engl. J. Med.* **308**, 1181–1184 (1983).
11. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (80-.).* **220**, 868–871 (1983).
12. Schüpbach, J. *et al.* Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science (80-.).* **224**, 503–505 (1984).
13. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science (80-.).* **224**, 497–500 (1984).
14. Gallo, R. C. *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (80-.).* **224**, 500–503 (1984).
15. Levy, J. A. *et al.* Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science (80-.).* **225**, 840–842 (1984).
16. What to call the AIDS virus? *Nature* **321**, 10 (1986).
17. WATANABE, M. E. Origins of HIV: The Interrelationship between Nonhuman Primates and the Virus. *Bioscience* **54**, 810 (2004).
18. Peeters, M. *et al. Risk to Human Health from a Plethora of Simian Immunodeficiency Viruses in Primate Bushmeat*. *Emerging Infectious Diseases* • **8**, (2002).
19. Worobey, M. *et al.* Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664 (2008).

20. Maartens, G., Celum, C. & Lewin, S. R. HIV infection: Epidemiology, pathogenesis, treatment, and prevention. in *The Lancet* **384**, 258–271 (Lancet Publishing Group, 2014).
21. Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**, 679–689 (2011).
22. Yamaguchi, J. *et al.* Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L. *JAIDS J. Acquir. Immune Defic. Syndr.* 1 (2019). doi:10.1097/QAI.0000000000002246
23. Lau, K. A. & Wong, J. J. L. Current trends of HIV recombination worldwide. doi:10.4081/idr.2013.s1.e4
24. Clinical Information Human Immunodeficiency Virus (HIV). (2016). doi:10.1159/000445852
25. Gelderblom, H. R. Assembly and morphology of HIV: Potential effect of structure on viral function. *AIDS* **5**, 617–637 (1991).
26. Processing of Retroviral RNA - Retroviruses - NCBI Bookshelf. Available at: https://www.ncbi.nlm.nih.gov/books/NBK19413/. (Accessed: 14th January 2020)
27. Wilusz, J. *Putting an "End" to HIV mRNAs: capping and polyadenylation as potential therapeutic targets*. (2013).
28. Mbondjiwonje, C. *et al.* Distinctive variation in the U3R region of the 5' Long Terminal Repeat from diverse HIV-1 strains. *PLoS One* **13**, (2018).
29. Mbonye, U. & Karn, J. Transcriptional control of HIV latency: Cellular signaling pathways, epigenetics, happenstance and the hope for a cure. *Virology* **454–455**, 328–339 (2014).
30. Donahue, D. A. & Wainberg, M. A. Cellular and molecular mechanisms involved in the establishment of HIV-1 latency. *Retrovirology* **10**, (2013).
31. Krebs, F. C., Hogan, T. H., Quiterio, S., Gartner, S. & Wigdahl, B. *Lentiviral LTR-directed Expression, Sequence Variation, and Disease Pathogenesis*.
32. Frankel, A. D. & Young, J. A. T. *HIV-1: Fifteen Proteins and an RNA*. *Annu. Rev. Biochem* **67**, (1998).
33. Freed, E. O. HIV-1 Gag proteins: Diverse functions in the virus life cycle. *Virology* **251**, 1–15 (1998).
34. Bernacchi, S. *et al.* HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **317**, 385–399 (2002).
35. Baldwin, D. N. & Linial, M. L. The roles of Pol and Env in the assembly pathway of human foamy virus. *J. Virol.* **72**, 3658–65 (1998).
36. Hu, W.-S. & Hughes, S. H. HIV-1 Reverse Transcription. doi:10.1101/cshperspect.a006882
37. Craigie, R. & Bushman, F. D. HIV DNA Integration. doi:10.1101/cshperspect.a006890
38. West, A. P. *et al.* Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* **156**, 633–48 (2014).
39. Munro, J. B. & Mothes, W. Structure and Dynamics of the Native HIV-1 Env Trimer. *J. Virol.* **89**, 5752–5 (2015).
40. Ma, X. *et al.* HIV-1 Env trimer opens through an asymmetric intermediate in which individual protomers adopt distinct conformations. *Elife* **7**, (2018).

41. Garg, H. & Blumenthal, R. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. *Cell. Mol. Life Sci.* **65**, 3134–44 (2008).
42. Merk, A. & Subramaniam, S. HIV-1 envelope glycoprotein structure. *Curr. Opin. Struct. Biol.* **23**, 268–76 (2013).
43. Seelamgari, A. *et al.* Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontiers in bioscience : a journal and virtual library* **9**, 2388–2413 (2004).
44. Accessory Proteins and Assembly - Retroviruses - NCBI Bookshelf. Available at: https://www.ncbi.nlm.nih.gov/books/NBK19407/. (Accessed: 14th January 2020)
45. Sauter, D. *et al.* Human Tetherin Exerts Strong Selection Pressure on the HIV-1 Group N Vpu Protein. *PLoS Pathog.* **8**, (2012).
46. Mann, J. K. *et al.* Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. *Retrovirology* **10**, (2013).
47. Foster, J. L., Denial, S. J., Temple, B. R. S. & Garcia, J. V. Mechanisms of HIV-1 Nef function and intracellular signaling. *Journal of Neuroimmune Pharmacology* **6**, 230–246 (2011).
48. Haller, C. *et al.* HIV-1 Nef and Vpu Are Functionally Redundant Broad-Spectrum Modulators of Cell Surface Receptors, Including Tetraspanins. *J. Virol.* **88**, 14241–14257 (2014).
49. Mahdi, M., Szojka, Z., Mótyán, J. A. & Tőzsér, J. Inhibitory Effects of HIV-2 Vpx on Replication of HIV-1. *J. Virol.* **92**, (2018).
50. Vicenzi, E. & Poli, G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens* **81**, 61–71 (2013).
51. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. doi:10.1101/cshperspect.a006866
52. Saphire, A. C., Bobardt, M. D., Zhang, Z., David, G. & Gallay, P. A. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J. Virol.* **75**, 9187–200 (2001).
53. Arthos, J. *et al.* HIV-1 envelope protein binds to and signals through integrin α4β7, the gut mucosal homing receptor for peripheral T cells. *Nat. Immunol.* **9**, 301–309 (2008).
54. Cicala, C. *et al.* The integrin α4β7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20877–20882 (2009).
55. Geijtenbeek, T. B. H. *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587–597 (2000).
56. Garg, H. & Blumenthal, R. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. *Cellular and Molecular Life Sciences* **65**, 3134–3144 (2008).
57. Ma, X. *et al.* HIV-1 env trimer opens through an asymmetric intermediate in which individual protomers adopt distinct conformations. *Elife* **7**, (2018).
58. Fassati, A. & Goff, S. P. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.* **75**, 3626–35 (2001).
59. Miller, M. D., Farnet, C. M. & Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71**, 5382–90 (1997).
60. Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and

other retroviruses. *Science (80-.).* **341**, 903–906 (2013).
61. Rasaiyaah, J. *et al.* HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* **503**, 402–405 (2013).
62. Yamashita, M., Perez, O., Hope, T. J. & Emerman, M. Evidence for Direct Involvement of the Capsid Protein in HIV Infection of Nondividing Cells. *PLoS Pathog.* **3**, e156 (2007).
63. Hulme, A. E., Perez, O. & Hope, T. J. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9975–9980 (2011).
64. McDonald, D. *et al.* Visualization of the intracellular behavior of HIV in living cells. *J. Cell Biol.* **159**, 441–452 (2002).
65. Campbell, E. M. & Hope, T. J. HIV-1 capsid: The multifaceted key player in HIV-1 infection. *Nature Reviews Microbiology* **13**, 471–483 (2015).
66. Farnet, C. M. & Bushman, F. D. HIV-1 cDNA integration: Requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **88**, 483–492 (1997).
67. Hilditch, L. & Towers, G. J. A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Current Opinion in Virology* **4**, 32–36 (2014).
68. Arhel, N. J. *et al.* HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *EMBO J.* **26**, 3025–3037 (2007).
69. Esposito, F., Corona, A. & Tramontano, E. HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Mol. Biol. Int.* **2012**, 1–23 (2012).
70. Jonckheere, H., Anné, J. & De Clercq, E. The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.* **20**, 129–54 (2000).
71. Konsavage, W. M., Sudol, M. & Katzman, M. Effects of varying the spacing within the D,D-35-E motif in the catalytic region of retroviral integrase. *Virology* **379**, 223–233 (2008).
72. Delelis, O., Carayon, K., Saïb, A., Deprez, E. & Mouscadet, J.-F. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* **5**, 114 (2008).
73. Autran, B. *et al.* Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science* **277**, 112–6 (1997).
74. Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2**, a007161 (2012).
75. Kulkosky, J. & Bray, S. HAART-persistent HIV-1 latent reservoirs: their origin, mechanisms of stability and potential strategies for eradication. *Curr. HIV Res.* **4**, 199–208 (2006).
76. Huyghe, J., Magdalena, S. & Vandekerckhove, L. Fight fire with fire: Gene therapy strategies to cure HIV. *Expert Rev. Anti. Infect. Ther.* **15**, 747–758 (2017).
77. Weiss, R. A. How does HIV cause AIDS? *Science (80-.).* **260**, 1273–1279 (1993).
78. Siliciano, R. F. & Greene, W. C. HIV latency. *Cold Spring Harb. Perspect. Med.* **1**, (2011).
79. Cavrois, M. *et al.* Mass Cytometric Analysis of HIV Entry, Replication, and Remodeling in Tissue CD4+ T Cells. *Cell Rep.* **20**, 984–998 (2017).

80. Alimonti, J. B., Ball, T. B. & Fowke, K. R. Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *Journal of General Virology* **84**, 1649–1661 (2003).
81. Finkel, T. H. *et al.* Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nat. Med.* **1**, 129–134 (1995).
82. Doitsh, G. *et al.* Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* **505**, 509–14 (2014).
83. Karris, M. A. Y. & Smith, D. M. Tissue-specific HIV-1 infection: Why it matters. *Future Virology* **6**, 869–882 (2011).
84. Shaw, G. M. & Hunter, E. HIV transmission. *Cold Spring Harb. Perspect. Med.* **2**, (2012).
85. Quinn, T. C. *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N. Engl. J. Med.* **342**, 921–929 (2000).
86. Miller, W. C., Rosenberg, N. E., Rutstein, S. E. & Powers, K. A. The Role of Acute and Early HIV Infection in the Sexual Transmission of HIV. doi:10.1097/COH.0b013e32833a0d3a
87. Røttingen, J. A., Cameron, W. D. & Garnett, G. P. A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: How much really is known? *Sex. Transm. Dis.* **28**, 579–597 (2001).
88. Unaids. *UNAIDS report on the global AIDS epidemic 2013 GLOBAL REPORT*. (2013).
89. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7557 (2008).
90. Shattock, R. J. & Moore, J. P. Inhibiting sexual transmission of HIV-1 infection. *Nat. Rev. Microbiol.* **1**, 25–34 (2003).
91. Cohen, O. J. & Fauci, A. S. *Host Factors that Affect Sexual Transmission of HIV*.
92. Kawamura, T., Kurtz, S. E., Blauvelt, A. & Shimada, S. The role of Langerhans cells in the sexual transmission of HIV. *Journal of Dermatological Science* **40**, 147–155 (2005).
93. Bomsel, M. Transcytosis of infectious human immunodeficiency virus across a tight human epithelial cell line barrier. *Nat. Med.* **3**, 42–47 (1997).
94. Kawamura, T. *et al.* R5 HIV productively infects langerhans cells, and infection levels are regulated by compound CCR5 polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8401–8406 (2003).
95. Turville, S., Wilkinson, J., Cameron, P., Dable, J. & Cunningham, A. L. The role of dendritic cell C-type lectin receptors in HIV pathogenesis. *J. Leukoc. Biol.* **74**, 710–718 (2003).
96. Kavanagh, D. G. & Bhardwaj, N. A division of labor: DC subsets and HIV receptor diversity. *Nature Immunology* **3**, 891–893 (2002).
97. Wilflingseder, D., Banki, Z., Dierich, M. P. & Stoiber, H. Mechanisms promoting dendritic cell-mediated transmission of HIV. *Molecular Immunology* **42**, 229–237 (2005).
98. Turville, S. G. *et al.* Diversity of receptors binding HIV on dendritic cell subsets. *Nat. Immunol.* **3**, 975–983 (2002).

99. Ballweber, L. *et al.* Vaginal langerhans cells nonproductively transporting HIV-1 mediate infection of T cells. *J. Virol.* **85**, 13443–7 (2011).
100. Gao, Y., McKay, P. & Mann, J. Advances in HIV-1 Vaccine Development. *Viruses* **10**, 167 (2018).
101. Mehandru, S. & Dandekar, S. Role of the gastrointestinal tract in establishing infection in primates and humans. *Current Opinion in HIV and AIDS* **3**, 22–27 (2008).
102. Milush, J. M. *et al.* Virally Induced CD4 + T Cell Depletion Is Not Sufficient to Induce AIDS in a Natural Host . *J. Immunol.* **179**, 3047–3056 (2007).
103. Patel, P. *et al.* Estimating per-act HIV transmission risk: A systematic review. *AIDS* **28**, 1509–1519 (2014).
104. Klein, K. *et al.* Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS Pathog.* **14**, e1006754 (2018).
105. Williams-Wietzikoski, C. A. *et al.* Comparisons of Human Immunodeficiency Virus Type 1 Envelope Variants in Blood and Genital Fluids near the Time of Male-to-Female Transmission. *J. Virol.* **93**, (2019).
106. Haaland, R. E. *et al.* Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1. *PLoS Pathog.* **5**, e1000274 (2009).
107. Carlson, J. M. *et al.* Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science (80-.).* **345**, (2014).
108. Wawer, M. J. *et al.* Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *J. Infect. Dis.* **191**, 1403–1409 (2005).
109. Quinn, T. C. *et al.* Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. *N. Engl. J. Med.* **342**, 921–929 (2000).
110. Chakraborty, H. *et al.* Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: A probabilistic empiric model. *AIDS* **15**, 621–627 (2001).
111. Freeman, E. E. *et al.* Herpes simplex virus 2 infection increases HIV acquisition in men and women: Systematic review and meta-analysis of longitudinal studies. *AIDS* **20**, 73–83 (2006).
112. Stekler, J. *et al.* HIV Dynamics in Seminal Plasma during Primary HIV Infection. *AIDS Res. Hum. Retroviruses* **24**, 1269–1274 (2008).
113. Derdeyn, C. A. *et al.* Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**, 2019–22 (2004).
114. Chohan, B. *et al.* Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J. Virol.* **79**, 6528–31 (2005).
115. Hu, J., Gardner, M. B. & Miller, C. J. Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells. *J. Virol.* **74**, 6087–95 (2000).
116. Miller, C. J. & Hu, J. T cell-tropic simian immunodeficiency virus (SIV) and simian-human immunodeficiency viruses are readily transmitted by vaginal inoculation of rhesus macaques, and Langerhans' cells of the female genital tract are infected with SIV. *J. Infect. Dis.* **179 Suppl 3**, S413-7 (1999).

117. Spira, A. I. *et al.* Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques. *J. Exp. Med.* **183**, 215–25 (1996).
118. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7 (2008).
119. Klein, K. *et al.* Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS Pathog.* **14**, e1006754 (2018).
120. Buckner, C. M. *et al.* Characterization of plasmablasts in the blood of HIV-infected viremic individuals: evidence for nonspecific immune activation. *J. Virol.* **87**, 5800–11 (2013).
121. Noelle, R. J. & Snow, E. C. Cognate interactions between helper T cells and B cells. *Immunol. Today* **11**, 361–8 (1990).
122. Rodriguez-Garcia, M. *et al.* Dendritic cells from the human female reproductive tract rapidly capture and respond to HIV. *Mucosal Immunol.* **10**, 531–544 (2017).
123. Whitney, J. B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–7 (2014).
124. Goonetilleke, N. *et al.* The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* **206**, 1253–72 (2009).
125. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15**, 137–48 (2015).
126. Crotty, S. A brief history of T cell help to B cells. *Nat. Rev. Immunol.* **15**, 185–9 (2015).
127. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–76 (2013).
128. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–19 (2011).
129. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–70 (2011).
130. Doria-Rose, N. A. *et al.* Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables. *J. Virol.* **84**, 1631–6 (2010).
131. Euler, Z. *et al.* Longitudinal analysis of early HIV-1-specific neutralizing activity in an elite neutralizer and in five patients who developed cross-reactive neutralizing activity. *J. Virol.* **86**, 2045–55 (2012).
132. Kwong, P. D. & Mascola, J. R. Human Antibodies that Neutralize HIV-1: Identification, Structures, and B Cell Ontogenies. *Immunity* **37**, 412–425 (2012).
133. Shcherbakov, D. N., Bakulina, A. Y., Karpenko, L. I. & Ilyichev, A. A. Broadly Neutralizing Antibodies against HIV-1 As a Novel Aspect of the Immune Response. *Acta Naturae* **7**, 11–21
134. Klein, F. *et al.* Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization. *Cell* **153**, 126–138 (2013).
135. Mouquet, H. *et al.* Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* **467**, 591–5 (2010).

136. Pavri, R. & Nussenzweig, M. C. AID Targeting in Antibody Diversity. in *Advances in immunology* **110**, 1–26 (2011).
137. Breden, F. *et al.* Comparison of Antibody Repertoires Produced by HIV-1 Infection, Other Chronic and Acute Infections, and Systemic Autoimmune Disease. *PLoS One* **6**, e16857 (2011).
138. Lingwood, D. *et al.* Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* **489**, 566–570 (2012).
139. Mascola, J. R. & Haynes, B. F. HIV-1 neutralizing antibodies: understanding nature's pathways. *Immunol. Rev.* **254**, 225–244 (2013).
140. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–82 (2002).
141. Briney, B. S., Willis, J. R. & Crowe, J. E. Human peripheral blood antibodies with long HCDR3s are established primarily at original recombination using a limited subset of germline genes. *PLoS One* **7**, e36750 (2012).
142. Klein, F. *et al.* Antibodies in HIV-1 vaccine development and therapy. *Science* **341**, 1199–204 (2013).
143. Hoot, S. *et al.* Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4bs bNAbs. *PLoS Pathog.* **9**, e1003106 (2013).
144. Xiao, X. *et al.* Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: Implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun.* **390**, 404–409 (2009).
145. De Boer, R. J. & Perelson, A. S. How germinal centers evolve broadly neutralizing antibodies: the breadth of the follicular helper T cell response. *J. Virol.* **91**, (2017).
146. Doria-Rose, N. A. *et al.* Frequency and phenotype of human immunodeficiency virus envelope-specific B cells from patients with broadly cross-neutralizing antibodies. *J. Virol.* **83**, 188–99 (2009).
147. Meffre, E. *et al.* Immunoglobulin heavy chain expression shapes the B cell receptor repertoire in human B cell development. *J. Clin. Invest.* **108**, 879–86 (2001).
148. Meffre, E. & Wardemann, H. B-cell tolerance checkpoints in health and autoimmunity. *Curr. Opin. Immunol.* **20**, 632–8 (2008).
149. Joseph, S. B., Swanstrom, R., Kashuba, A. D. M. & Cohen, M. S. Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. *Nature Reviews Microbiology* **13**, 414–425 (2015).
150. Poltronieri, P., Sun, B. & Mallardo, M. RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. *Curr. Genomics* **16**, 327–335 (2015).
151. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition by the innate immune system. *Int. Rev. Immunol.* **30**, 16–34 (2011).
152. Yoo, J. S., Kato, H. & Fujita, T. Sensing viral invasion by RIG-I like receptors. *Current Opinion in Microbiology* **20**, 131–138 (2014).
153. Lester, S. N. & Li, K. Toll-like receptors in antiviral innate immunity. *Journal of Molecular Biology* **426**, 1246–1264 (2014).
154. Heil, F. *et al.* Species-Specific Recognition of Single-Stranded RNA via Till-like Receptor 7 and 8. *Science (80-.).* **303**, 1526–1529 (2004).
155. Chen, L. *et al.* Inflammatory responses and inflammation-associated diseases in

organs. *Oncotarget* **9**, 7204–7218 (2018).
156. Beignon, A. S. *et al.* Endocytosis of HIV-1 activates plasmacytoid dendritic cells via Toll-like receptor-viral RNA interactions. *J. Clin. Invest.* **115**, 3265–3275 (2005).
157. O'Brien, M., Manches, O. & Bhardwaj, N. Plasmacytoid Dendritic Cells in HIV Infection. in 71–107 (2012). doi:10.1007/978-1-4614-4433-6_3
158. Iwasaki, A. Innate Immune Recognition of HIV-1. *Immunity* **37**, 389–398 (2012).
159. Lepelley, A. *et al.* Innate sensing of HIV-infected cells. *PLoS Pathog.* **7**, (2011).
160. Gringhuis, S. I. *et al.* HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. *Nat. Immunol.* **11**, 419–426 (2010).
161. Donninelli, G., Gessani, S. & Del Cornò, M. Interplay between HIV-1 and Toll-like receptors in human myeloid cells: friend or foe in HIV-1 pathogenesis? *J. Leukoc. Biol.* **99**, 97–105 (2016).
162. Guo, H., Gao, J., Taxman, D. J., Ting, J. P. Y. & Su, L. HIV-1 infection induces interleukin-1β production via TLR8 protein-dependent and NLRP3 inflammasome mechanisms in human monocytes. *J. Biol. Chem.* **289**, 21716–21726 (2014).
163. Doehle, B. P. *et al.* Vpu Mediates Depletion of Interferon Regulatory Factor 3 during HIV Infection by a Lysosome-Dependent Mechanism. *J. Virol.* **86**, 8367–8374 (2012).
164. Parameswaran, N. & Patial, S. Tumor necrosis factor-a signaling in macrophages. *Critical Reviews in Eukaryotic Gene Expression* **20**, 87–103 (2010).
165. Idriss, H. T. & Naismith, J. H. TNF? and the TNF receptor superfamily: Structure-function relationship(s). *Microsc. Res. Tech.* **50**, 184–195 (2000).
166. Kalb, A., Bluethmann, H., Moore, M. W. & Lesslauer, W. Tumor necrosis factor receptors (Tnfr) in mouse fibroblasts deficient in Tnfr1 or Tnfr2 are signaling competent and activate the mitogen-activated protein kinase pathway with differential kinetics. *J. Biol. Chem.* **271**, 28097–28104 (1996).
167. Erickson, S. L. *et al.* Decreased sensitivity to tumour-necrosis factor but normal T-cell development in TNF receptor-2-deficient mice. *Nature* **372**, 560–563 (1994).
168. Goldfeld, A. E. & Tsai, E. Y. TNF-α and genetic susceptibility to parasitic disease. *Exp. Parasitol.* **84**, 300–303 (1996).
169. De Jong, M. A. W. P. *et al.* TNF-α and TLR agonists increase susceptibility to HIV-1 transmission by human Langerhans cells ex vivo. *J. Clin. Invest.* **118**, 3440–3452 (2008).
170. Matsuyama, T. *et al.* Enhancement of HIV Replication and Giant Cell Formation by Tumor Necrosis Factor. *AIDS Res. Hum. Retroviruses* **5**, 139–146 (1989).
171. Osborn, L., Kunkel, S. & Nabel, G. J. Tumor necrosis factor α and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor κB. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2336–2340 (1989).
172. Kimber, I. & Cumberbatch, M. Stimulation of Langerhans cell migration by tumor necrosis factor α (TNF-α). *J. Invest. Dermatol.* **99**, 48–50 (1992).
173. Vaidya, S. A. *et al.* Tumor necrosis factor α is associated with viral control and early disease progression in patients with HIV type 1 infection. *J. Infect. Dis.* **210**, 1042–1046 (2014).
174. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).

175. Munro, J. B. *et al.* Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science (80-.).* **346**, 759–763 (2014).
176. Gaschen, B. *et al.* Diversity Considerations in HIV-1 Vaccine Selection. *Science (80-.).* **296**, 2354–2360 (2002).
177. Korber, B. *et al.* Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **58**, 19–42 (2001).
178. Starcich, B. R. *et al.* Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**, 637–48 (1986).
179. Bitto, D. *et al.* Determination of N-linked Glycosylation in Viral Glycoproteins by Negative Ion Mass Spectrometry and Ion Mobility. *Methods Mol. Biol.* **1331**, 93–121 (2015).
180. Doores, K. J. *et al.* Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc. Natl. Acad. Sci.* **107**, 13800–13805 (2010).
181. Zhang, M. *et al.* Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* **14**, 1229–1246 (2004).
182. Pritchard, L. K. *et al.* Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6**, 7479 (2015).
183. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10**, 14 (2013).
184. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
185. Montefiori, D. C., Robinson, W. E. & Mitchell, W. M. Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9248–52 (1988).
186. Lee, W. R. *et al.* Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2213–7 (1992).
187. Ohgimoto, S. *et al.* Location-specific, unequal contribution of the N glycans in simian immunodeficiency virus gp120 to viral infectivity and removal of multiple glycans without disturbing infectivity. *J. Virol.* **72**, 8365–70 (1998).
188. Pantophlet, R. & Burton, D. R. GP120: Target for Neutralizing HIV-1 Antibodies. *Annu. Rev. Immunol.* **24**, 739–769 (2006).
189. Burton, D. R. & Hangartner, L. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu. Rev. Immunol.* **34**, 635–59 (2016).
190. Klasse, P. J., Sanders, R. W., Cerutti, A. & Moore, J. P. How can HIV-type-1-Env immunogenicity be improved to facilitate antibody-based vaccine development? *AIDS Res. Hum. Retroviruses* **28**, 1–15 (2012).
191. Dosenovic, P. *et al.* Immunization for HIV-1 Broadly Neutralizing Antibodies in Human Ig Knockin Mice. *Cell* **161**, 1505–15 (2015).
192. Jardine, J. G. *et al.* Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science (80-.).* **349**, 156–161 (2015).
193. Jardine, J. *et al.* Rational HIV Immunogen Design to Target Specific Germline B

Cell Receptors. *Science (80-.).* **340**, 711–716 (2013).
194. McGuire, A. T. *et al.* Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. *J. Exp. Med.* **210**, 655–63 (2013).
195. Kwong, P. D., Mascola, J. R. & Nabel, G. J. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat. Rev. Immunol.* **13**, 693–701 (2013).
196. Kulp, D. W. & Schief, W. R. Advances in structure-based vaccine design. *Curr. Opin. Virol.* **3**, 322–31 (2013).
197. Zhou, T. *et al.* Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep.* **19**, 719–732 (2017).
198. Liang, Y. *et al.* Changes in Structure and Antigenicity of HIV-1 Env Trimers Resulting from Removal of a Conserved CD4 Binding Site-Proximal Glycan. *J. Virol.* **90**, 9224–36 (2016).

Chapter 2 « Objective one preface »

Understanding and being able to identify what constitutes a T/F virus would undoubtedly advance the field in terms of identifying novel vaccine candidates and also, in understanding the basic biology of HIV infection. While the phenotypic differences attributed to T/F viruses may allow for enhanced transmission, here we focused on viral RNA antigenicity as a determinant in sexual transmission. Work by Forsbach *et al*., identified 4-mer sequence motifs that presented differential IFN-α and TNF-α stimulation on primary human cells. They identified AU-rich motifs induced a high amount of TNF-α while GU-rich motifs induced a high amount of IFN-α[1]. As such, we hypothesized that the presence of unique 4-mer sequence motifs that induce TNF-α within viral RNA may synergistically work with already established phenotypic differences in T/F viruses to enhance transmission of HIV-1. We believe that viral RNA that contains more immunostimulatory motifs may play a role in transmission.

To do this, we developed a bioinformatic tool that scans multiple genetic sequences to enumerate selected immunostimulatory motifs within viral RNA. The quantity of these immunostimulatory sequences were then compared between donors and recipients, between blood and the mucosal compartment within individuals, between blood and the semen compartment within individuals and overtime in patients. We observed that viral sequences that contained more TNF-α inducing motifs were isolated from the recipient in transmission pair cohorts. Furthermore, we showed that viruses in semen had a higher number of proinflammatory TNF-α inducing motifs than those isolated from the blood of the same individual. Finally, sequences identified in the blood contained more TNF-α inducing motifs than those found in the cervix within infected individuals. This ongoing analysis will help screen potential T/F viral sequences that could be used to select for novel vaccine antigens.

1. Forsbach, A. *et al.* Identification of RNA Sequence Motifs Stimulating Sequence-Specific TLR8-Dependent Immune Responses. *J. Immunol.* **180**, 3729–3738 (2008).

« The antigenicity of packaged viral ribonucleic acid (RNA) may contribute to mucosal HIV-1 transmission fitness »

Authors: Katja Klein[1] and Rahul Pawa[1], Chanuka N. Wijewardhana[1], Richard Gibson[1], Ryan Troyer[1], Robin J. Shattock[2], Eric J. Arts[1,3,4] and Jamie F.S. Mann[1,5*]

Affiliations:

[1]Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, Canada N6A 5C1.

[2]Department of infectious Diseases, Division of Medicine, Imperial College London, Norfolk Place, London, W2 1PG, UK.

[3]Joint Clinical Research Center, Kampala, UG

[4]Department of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

[5] School of Veterinary Sciences, University of Bristol, Bristol, United Kingdom.

*Correspondence to: J.F.S Mann, University of Bristol.

Manuscript submitted

Abstract

In the majority of sexual transmissions, a single genetic viral variant, termed the Transmitted/founder (T/F), successfully overcomes a series of selective genetic bottlenecks, to establish a productive infection within the host. We provide novel evidence that the viral genome may play an important role in the successful transmission of T/F viruses. We screened the HIV-1 genomes of previously published datasets for the presence of RNA sequence motifs known to be immunostimulatory and triggers of innate pathogen recognition receptors. We found that the numbers of immunostimulatory sequence motifs decreased significantly as individuals progress to AIDS, that the number of immunostimulatory RNA motifs in viruses sequenced from semen compared to the serum compartment were increased, and that viruses with higher numbers of immunostimulatory RNA motifs can exit the cervix more efficiently than viruses containing less immunostimulatory motifs. Overall, we present new evidence that helps to explain the efficiency and success of T/F viruses.

Introduction

Human Immunodeficiency Virus-1 (HIV-1) is a major cause of morbidity and mortality around the world. HIV-1 is an RNA virus that displays extraordinary rates of genetic change owing to various combinations of selective forces. Factors that affect the evolutionary trajectory of HIV-1 at the population and intra-patient level include, an error prone reverse transcriptase that lacks proof reading function as well as innate and adaptive immune pressures. HIV-1 is spread primarily by sexual intercourse, through which the virus is capable of passing the mucosal barrier to infect target CD4 T cells, primarily believed to be of the T_H17 phenotype[1]. While the recipient is exposed to millions of viral variants from the donor (termed the viral swarm), in 70-80% of infections, a single virus (called the transmitted/founder; T/F) is able to successfully establish a systemic infection within the recipient[2]. In generating a systemic infection, the viral swarm that the recipient is initially exposed to, encounters a dramatic genetic bottleneck that selects for the successful T/F. Work published by Klein *et al.* demonstrates that early HIV-1 isolates within the blood displayed a more homogenous genotype than clones in the female genital tract. Their analysis of early mucosal infection in women revealed that despite the high diversity of HIV-1 in the cervix, a possible mucosal sieve effect may select for a single variant to establish a homogenous systemic infection[3]. Similarly, Williams-Wietzikoski *et al.* analyzed viral sequences in semen and plasma –in the case of male-to-female transmission– and reported seeing significant genetic differences between viral RNA isolated from seminal plasma and blood within certain males in their transmission pairs[4]. This in turn suggests the occurrence of unique viral variants present in compartments relevant to transmission via sexual intercourse. Evidence for the success of these T/F viruses may be heavily influenced by viral fitness, inferred from sequence conservation, viral subtype, gender, viral load, presence of other sexually transmitted infectious agents and the integrity of the immune system[5–11].

What makes a T/F virus successful at establishing an infection within the recipient is currently unknown. The majority of research into why T/F viruses succeed tends to focus on the phenotypic differences within the envelope (Env) glycoprotein. While the specific phenotypic differences vary between subtypes, it is believed that there are unique

characteristics in the Env of T/F virus compared to Env found on virus at chronic stage of infection. Work by Derdeyn *et al.* has shown that T/F viruses of subtype C share two unique properties –shorter V1-V4 regions and a significantly lower number of potentially N-linked glycan sites (PNLGs)– that may have allowed the variants to transmit more easily, and successfully establish infection[12]. Similarly, a study in Kenya looked at early infection sequences in subtype A and noticed that they had significantly shorter V1-V2 loop sequences and fewer PNLGs[13]. On the contrary, subtype B T/F virus have been shown to have higher Env expression and incorporation of Env on virions[13]. While each subtype may have its own unique feature, it is understood that T/F viruses may have features that enhance their chances of infection.

The selective pressures that shape the various genetic bottlenecks and lead to successful transmission of T/F viruses occur at different stages of the transmission cycle – within the donor, at the site of transmission within the donor, during the transmission process between the donor and recipient, or within the recipient upon exposure to the viral swarm from the donor[14]. While the selective pressures will be different for each donor and recipient, it is unlikely that a single phenotypic or genetic sequence will inherently select for a successful T/F virus. Many biological factors beyond the control of the virus, such as pro-inflammatory factors, altered microbiome, and preexisting sexually transmitted diseases (STDs), have been implicated in enhancing transmission of HIV-1. While proteins can activate various pro-inflammatory signal cascades, viral RNA has been shown to mediate similar effects. RNA has been implicated to play an essential role in the evolution and regulation of DNA/protein-based life. These include, but are not limited to, protein coding, and secondary structures that play a role for virus function like ribosomal entry structure, internal ribosomal entry site and many more[15]. In response, our innate immune system has evolved to identify foreign viral RNA sequences and initiate a cascade of pro-inflammatory and antiviral responses to protect us from infection. Toll-like-receptors and other innate pattern recognition receptors, like RIG-I and MAVS, play a crucial role in sensing and reacting against identified viral genetic sequences, ultimately leading to production of antiviral responses spearheaded by type-1 interferons such as IFN-α[16,17]. More specifically, TLR7 and TLR8 recognize ssRNA and initiate a signal cascade that eventually produces

IFN-α and pro-inflammatory cytokines like TNF-α, IL-6, and IL-12[18,19]. Creating a pro-inflammatory environment is beneficial for the host as it can activate macrophages, can induce apoptosis of surrounding and infected cells and recruit immune cells to sites of infection, all of which play a role in fighting and clearing the infection[20].

Remarkably, HIV-1 may have evolved to make use of this pro-inflammatory niche by using it to foster its infection. Work by Marein *et al.* showed that TNF-α increases HIV-1 replication in Langerhans cells (LCs) whereas Pam3CSK4, a TLR1/TRL2 ligand, enhanced HIV-1 capture within this subset of cells[21]. Furthermore, TNF-α has been implicated in enhancing HIV-1 replication by creating an inflammatory niche that attracts DCs, macrophages and T cells[21–25]. Ironically, while this pro-inflammatory niche arms the immune system to fight the infection, it also creates the ideal opportunity for HIV-1 to establish infection. This could be achieved through activating the lymph node migratory potential of antigen presenting cells (APCs) for in trans-infection of CD4 T cells, or even direct infection of activated mucosal CD4 T cells for infectious foci development.

Understanding and being able to identify what constitutes a T/F virus would undoubtedly advance the field in terms of understanding the basic biology of HIV infection and in identifying novel vaccine candidates. While certain phenotypic differences attributed to T/F viruses may allow for enhanced transmission, here we focused on viral RNA antigenicity as a determinant in sexual transmission. Work by Forsbach *et al*. identified various immunostimulatory sequence motifs that could induce IFN-α and TNF-α production by primary human cells. Moreover, they identified that AU-rich motifs induced a high amount of TNF-α while GU-rich motifs induced a high amount of IFN-α[26]. As such, we hypothesized that an abundance of immunostimulatory sequence motifs within viral RNA may enhance the ability of HIV-1 T/F viruses to establish a productive infection compared to viruses that failed to infect which may have reduced numbers of proinflammatory motifs. To address this hypothesis, we developed a bioinformatic counting script that analyzes thousands of genetic sequences at once to enumerate selected immunostimulatory motifs within viral RNA. The quantity of these immunostimulatory RNA motifs were then compared within the sequenced HIV quasispecies (1) between donors and recipients; (2) between the blood and the mucosal compartment within

recipients; (3) between the blood and the semen compartment within HIV donors; and (4) how the numbers of motifs change over time within infected individuals. By analyzing thousands of sequenced viruses from various different cohorts and tissue compartments, we identified an intriguing biological phenomenon associated with successful viral transmission. Namely, HIV-1 viruses with a higher number of pro-inflammatory TNF-α motifs appeared to be more successful in the transmission process and these viruses appear to be compartmentalized within the semen of donors.

Results

The env gene from Clades A, B, C, and D, contains the highest prevalence of pro-inflammatory TNF-α inducing motifs, compared to the rest of the genome.

Growing evidence suggests that RNA immunostimulatory motifs play an important role in viral infections. We thought to characterize the number and distribution of pro-inflammatory motifs previously identified as TNF-α, IFN-α and non-methylated CpG in the HIV-1 genome. This would enable us to determine if these motifs are uniformly distributed throughout the viral genome or if they are expressed at higher frequencies in certain areas of the genome. First, we analyzed the prevalence of immunostimulatory motifs in the whole genome of the common HIV-1 group M reference strains A, B, C and D obtained from the Los Alamos National Laboratory HIV database (https://www.hiv.lanl.gov), excluding any recombinant strains. Analysis of the whole genome showed that the prevalence of TNF-α, IFN-α and CpG motifs is fairly similar between different clades (**Fig 2-1a**). To evaluate any differences between specific regions within the genome, we analyzed *gag*, *pol*, *env*, *vif-vpr*-5'*tat-vpu*, and *nef* separately. Interestingly, in all four clades, the *env* gene displayed the highest prevalence of TNF-α inducing motifs (**Fig 2-1**). In contrast, the accessory protein genes *vif-vpr*-5'*tat-vpu* showed the highest prevalence of IFN-α inducing motifs, with the exception of Clade B, where *env* had a higher prevalence of IFN-α inducing motifs (**Fig 2-1**). The highest prevalence of CpG motifs was observed in the *nef* region of the genome in Clades B, C and D, while *env* displayed a higher prevalence in Clade A (**Fig 2-1**). Based on our observations

we selected *env* as our region of interest for further analysis on the impact of these immunostimulatory motifs on viral transmission.

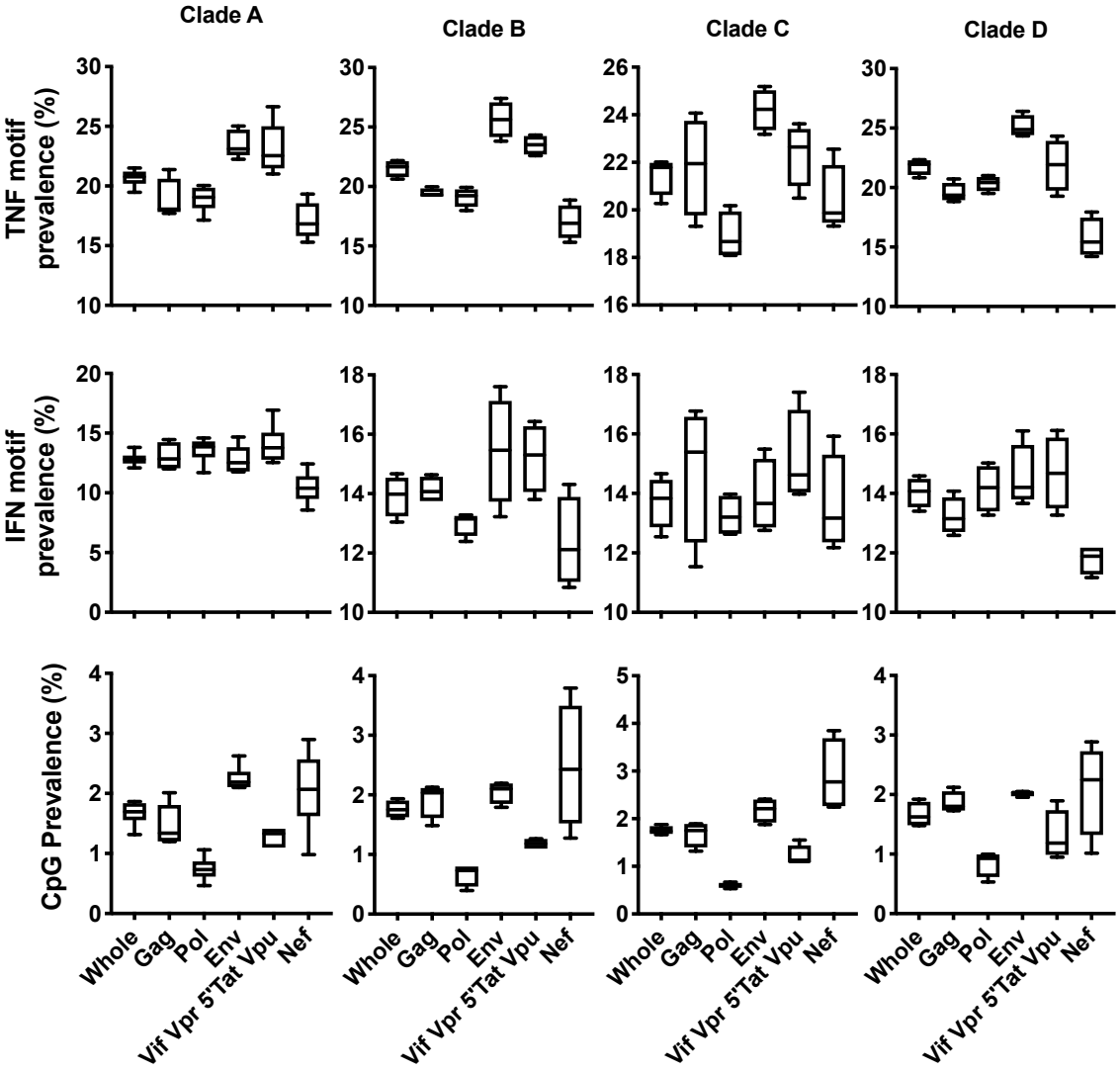


Figure 2-1: For clades A, B, C and D, env displays the highest prevalence of TNF-α motifs.

Whole genome sequences for the HIV-1 reference strains of each clade (A, B, C and D) were downloaded from the Los Alamos National Laboratory HIV database (https://www.hiv.lanl.gov). The genome was divided into 5 parts: *gag* (790-2086 bp), *pol* (2085-4230 bp), *env* (6225-8795 bp), *vif-vpr*-5'*tat-vpu* (5041-6310 bp), and *nef* (8797-9417 bp) based on HXB2 positioning. Sequences were aligned to HXB2 and trimmed where necessary to ensure that each sequence with a particular gene being analyzed was the same length. To normalize for differences in lengths between the various genes, the

results are shown as percentage (%) prevalence of TNF-α, IFN-α and CpG motifs within the gene. Number of reference sequences analyzed for each clade: A (*n*=6); B (*n*=4); C (*n*=4); and D (*n*=4).

The pro-inflammatory capability of HIV-1 *env* decreases overtime

Upon identifying the frequency of these immunostimulatory motifs in HIV-1 *env*, we were unsure of the dynamic nature of these motifs. For instance, does the immunostimulatory motif number remain stable or does it change over time? If the motif number does change over time, does it follow a random pattern associated with the stochastic nature of genetic drift, or does it represent an evolutionary pattern of adaptation towards fitness? If there are indeed changes in the motif number, genetic drift would result in equally increased and decreased motif numbers across the whole viral population, while evolutionary adaptation towards fitness would favor mutations in one direction over the other. To test this hypothesis, noting that the virus is under a range of selective pressures, we downloaded and analyzed 1850 HIV-1 clade B sequences from the Los Alamos database. Specifically, we analyzed the *env* sequences for Fiebig stages 5 (30 – 100 days post infection) and 6 (>100 days post infection) to assess if there was any change from Fiebig 1 at a global level of clade B infection. From this analysis we observed that the number of TNF-α inducing motifs significantly decreased ($p<0.0001$) from Fiebig stage 1 to stages 5 & 6, indicating that at a global level, the antigenicity of the virus begins to decrease over time in patients (**Fig 2-2**). This analysis suggests that the viruses early after transmission, contain higher number of proinflammatory TNF-α inducing motifs. Using the longitudinal sequence data from Shankarappa *et al.* (1999)[27], we did attempt to evaluate and confirm the same phenomenon seen in the Los Alamos data set, i.e. the pro-inflammatory motifs of viral isolates decrease over time over the course of infection. Unfortunately, due to the small nature of the cohort (n=8), we did not observe any clear bias in reduction of TNF-α motifs over time (**Supplementary Figure 2-1**). This data suggests that the reduction in TNF-α immunostimulatory motifs is a non-random phenomenon and more likely to be an adaptive evolutionary shift. It would be interesting to speculate that over time, the virus within the host adapts to the immune pressure by reducing the immunostimulatory potential of its RNA.

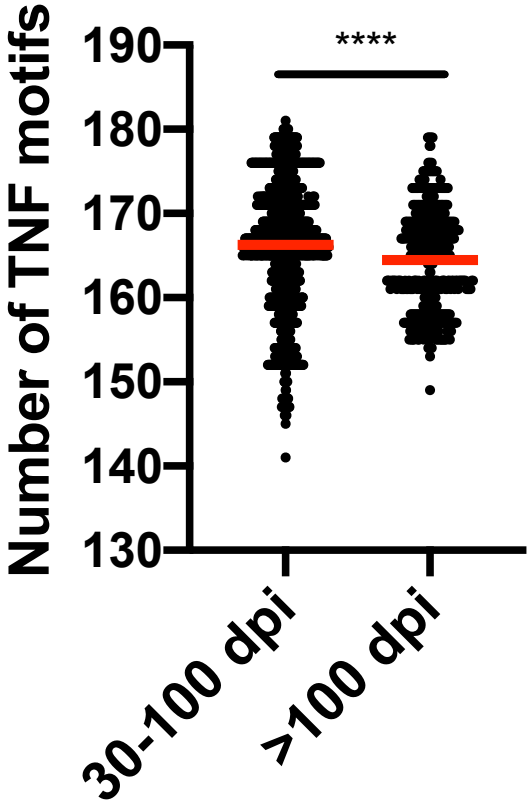


Figure 2-2: The number of TNF-α inducing motifs decreases overtime in HIV-1 infected individuals.

HIV-1 *env* sequences (HXB2 6225-8792 position) isolated from individuals at Fiebig stages 5 (30-100 dpi) and 6 (>100 dpi) of viral infection, were downloaded from the Los Alamos sequence database. Sequences were aligned to HXB2 *env* and trimmed where necessary to ensure the same sequence lengths were being analyzed (2570 bp). Red line represents the geometric mean number of TNF-α inducing motifs in the HIV sequences. Each point represents a unique viral sequence (30-100 dpi: n = 1364; >100 dpi n = 486). Mann-Whitney statistical analysis was performed with ****$p \leq 0.0001$.

Virus containing more TNF-α inducing motifs are more likely to be isolated within the serum compartment of the recipient compared to the donor.

The previous data indicated that while TNF-α motifs clustered within the *env* gene, we noticed an overall decrease overtime in the presence of TNF-α inducing RNA motifs. This means that for the vast majority of viruses, there are more TNF-α inducing motifs early in infection and as infection progresses there is a tendency to lose the motifs. This high pro-inflammatory viral RNA, early in infection, we argued may play a role in successful transmission of T/F viruses. Considering that 89.1% of HIV-1 infection events occur during the donor's acute stage infection where viral loads are high, this may have implications for the inflammatory nature of the viral RNA[28]. To assess the impact of pro-inflammatory motifs on viral transmission, we sought to analyze sequences from primary HIV viral isolates obtained from donor and recipient blood. We first analyzed the *env* gene sequences from transmission pairs obtained from a study by Iyer *et al.* (2017)[29]. The transmission pairs were stratified into MSM (Pairs 1, 2), MTF (Pair 3) and FTM (Pairs 4, 5, 6 and 7). Within the MSM and MTF cohorts (*n*=3), we observed that 100% (3/3) of viral *env* sequences obtained from the recipients showed higher TNF-α inducing motifs compared to the sequences derived from their respective donor (**Fig 2-3**). The IFN-α inducing motifs were representative of what we observed for the TNF-α inducing motifs, whereby 100% (3/3) of the sequences displayed a higher number of IFN-α motif in the recipients. Of note, a similar trend was observed with CpG inducing motifs whereby 67% (2/3) of the sequences in the recipient had higher CpG motifs with one pair demonstrating no change (Pair 2). We next analyzed the sequences from the FTM cohort (*n*=4) and unlike the previous cohort, we observed that 50% (2/4) of the recipients had viral *env* sequences with a higher number of TNF-α inducing motifs compared to their respective donor (**Fig 2-3**). Unlike the previous cohort, we observed 50% (2/4) of the recipients displayed an increase in the number of IFN-α motifs and CpG motifs in their viral isolates (**Fig 2-3**). In the one MTF transmission pair, we observed an increase in TNF-α, IFN-α and CpG motifs from sequences in the recipient compared to the donor. Although we can see slight

increases in the number of IFN-α and CpG motifs from viral isolates of the recipients in these cohorts, the trend is not as strong as that observed with TNF-α (**Fig 2-3**).

Having observed differences between donor and recipient within this small cohort, with the latter having virus with higher numbers of TNF-α inducing motifs, we sought to expand our analysis. Using *env* sequences from a transmission cohort published by Oberle *et al.* (2016) we analyzed an additional nine transmission pairs[30]. Additionally, the transmission pairs were stratified into MSM (Pairs 1, 3, 5, 7, 8 and 9) and MTF (Pair 2, 4 and 6) transmissions. Within the MSM cohort (*n*=6), we observed that 67% (4/6) of viral *env* sequences obtained from the recipients showed higher TNF-α inducing motifs compared to sequences derived from their respective donor (**Fig 2-4**). The IFN-α inducing motifs were representative of what we observed for the TNF-α inducing motifs, whereby 83% (5/6) displayed a higher number of IFN-α motif in the recipient (**Fig 2-4**). We next analyzed the viral sequences from the MTF transmission cohort (*n*=3) and unlike the previous transmission pairs, we observed that only 33% (1/3) of the recipients had viral *env* sequences with a higher number of TNF-α inducing motifs compared to their respective donor (**Fig 2-4**). Similarly, we observed that contrary to the previous cohort, 100% of the recipients displayed a decrease in the number of IFN-α motifs and 67% of the recipients showed an increase in CpG motifs in their viral isolates (**Fig 2-4**). The contrary analysis observed in the MTF cohort may be due to the limited sample size (*n*=3), whereas the MSM cohort had twice the number of transmission pairs (*n*=6). Regardless, when the MSM and MTF datasets from Oberle *et al.* were pooled together with the dataset from Iyer e*t al.*, 75% (12/16) of the recipients had viral *env* sequences with a higher number of TNF-α inducing motifs compared to their respective donors. This suggests that the virus that successfully establishes systemic infection in the recipient contains more pro-inflammatory TNF-α inducing motifs than the virus found in the donor's blood.

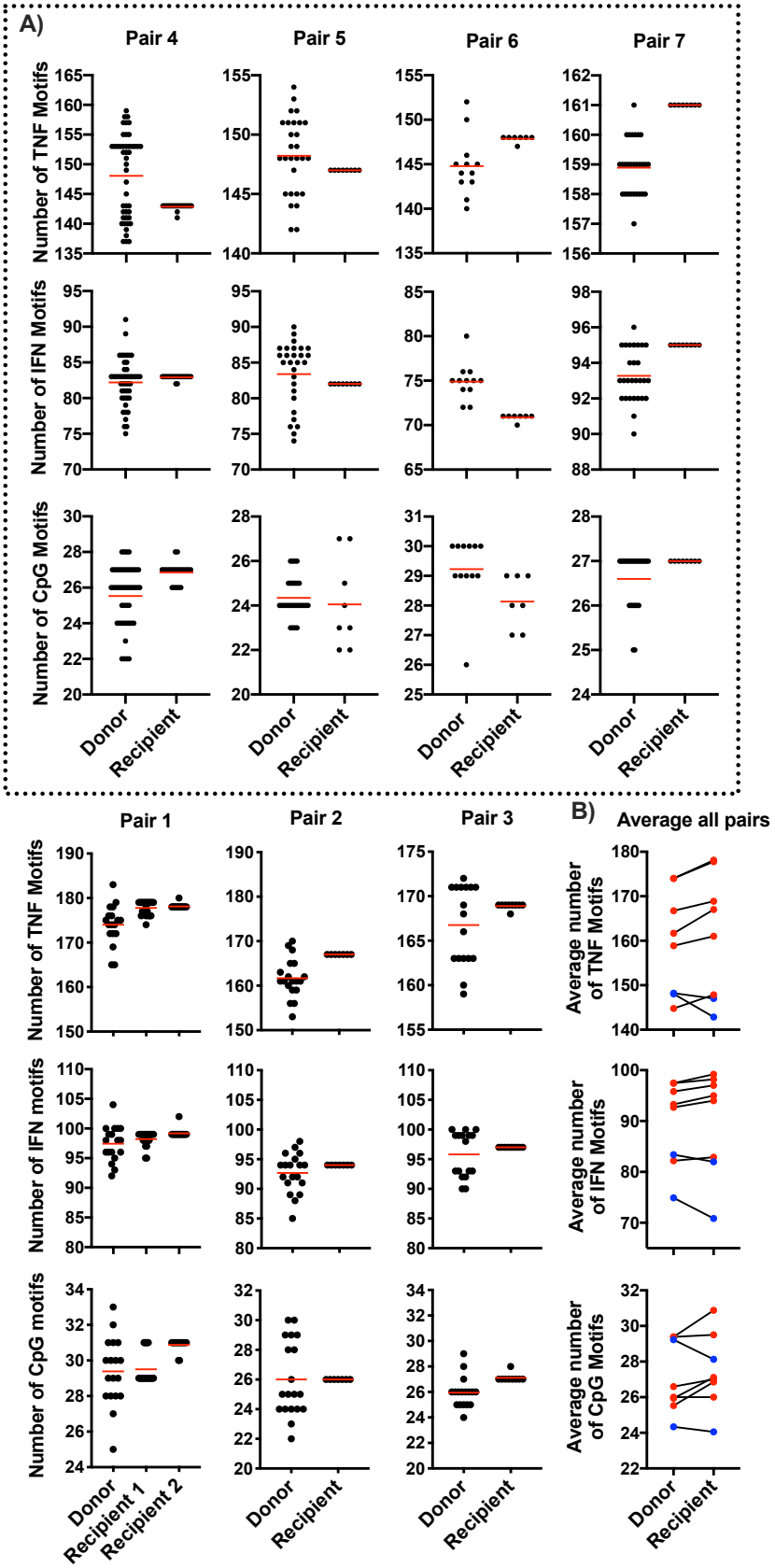

A)
Pair 4
Pair 5
Pair 6
Pair 7
Number of TNF Motifs
Number of IFN Motifs
Number of CpG Motifs
Donor
Recipient
Pair 1
Pair 2
Pair 3
Recipient 1
Recipient 2
B)
Average all pairs
Average number of TNF Motifs
Average number of IFN Motifs
Average number of CpG Motifs

Figure 2-3: Recipients in transmission pairs have viral sequences with higher numbers of immunostimulatory RNA motifs compared to their respective donors.

Donor and recipient *env* sequences were analyzed from the Iyer *et al*. 2017, dataset. Sequences were aligned to HXB2 to ensure we were analyzing the same region of *env* (6225-8795 bp) and were trimmed where necessary to ensure the same length (2570 bp) was being analyzed. For all donor recipient pairs (*n*= 7), graphs display the number of TNF-α, IFN-α and CpG motifs. (A) Individual transmission pairs where Pairs 1 to 2 represent men who have sex with men (MSM), Pair 3 represents male to female (MTF); and Pairs 4 to 7 represent female to male (FTM) transmissions (dotted square). Pairs 1 to 3 are subtype B infections while Pairs 4 to 7 are subtype C infections. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each point represents a unique viral sequence. (B) Average of all transmission pairs. Red dots from matched donors and recipients signifies increases in the number of immunostimulatory motifs while blue dots between donors and recipients signifies a decrease in the number of immunostimulatory motifs.

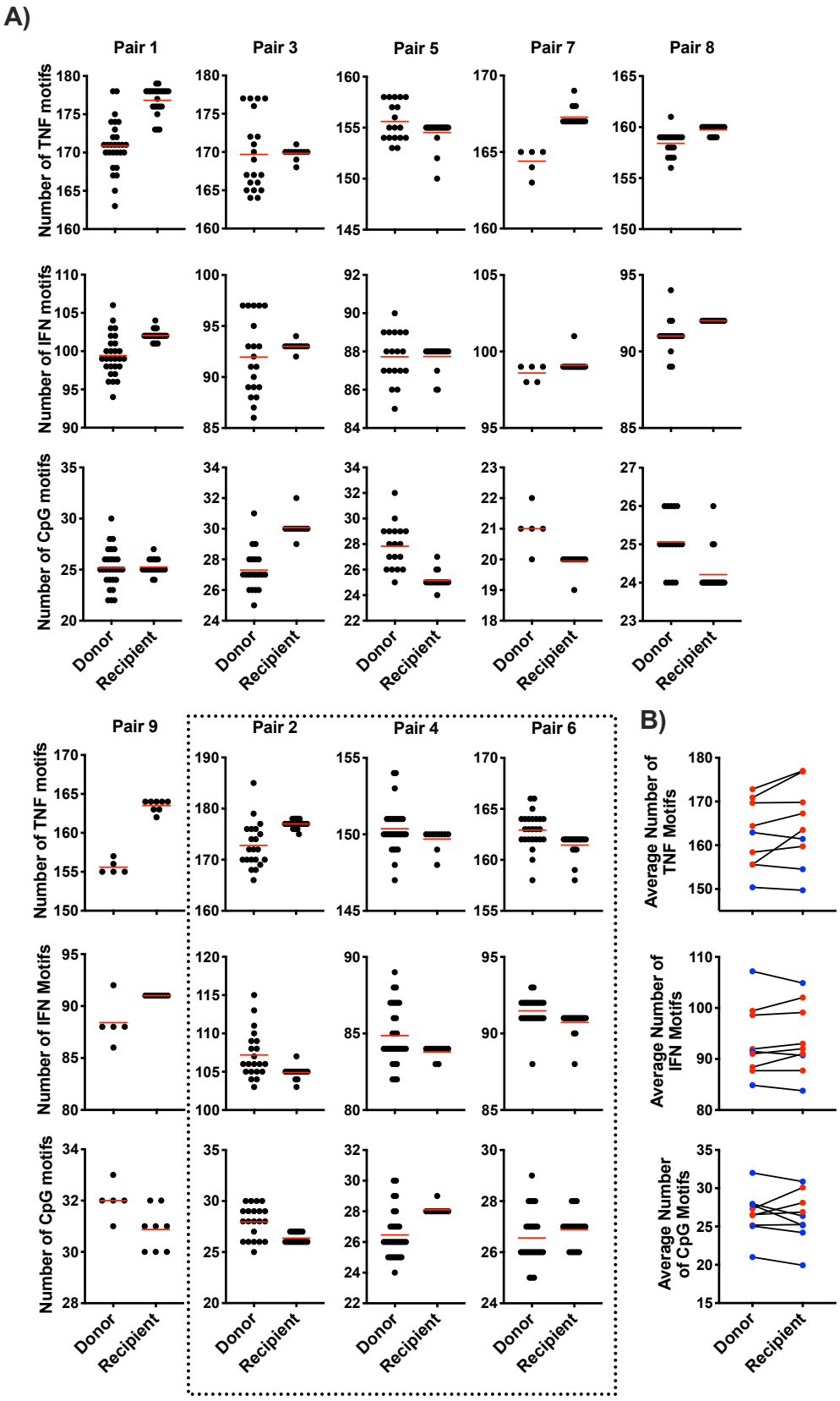

A)
Pair 1
Pair 3
Pair 5
Pair 7
Pair 8
Pair 9
Pair 2
Pair 4
Pair 6
Number of TNF motifs
Number of IFN motifs
Number of CpG motifs
Donor
Recipient
B)
Average Number of TNF Motifs
Average Number of IFN Motifs
Average Number of CpG Motifs

Figure 2-4: The recipients within transmission pairs have viral sequences containing a higher number of immunostimulatory motifs compared to their respective donors.

Donor and recipient sequences were analyzed from the Oberle *et al.,* 2016 dataset. **(A)** The dataset was stratified to male-to-male (MSM) and female-to-male (HET) transmissions (dotted box). The *env* sequences were aligned to HXB2 to ensure analysis of the *env* region (6225-8795 bp) and were trimmed where necessary to ensure the same length (2570 bp) was being analyzed. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each point represents a unique viral sequence. **(B)** Average of all MSM and MTF transmissions. For all donor recipient pairs, graphs display number of TNF-α, IFN-α and CpG motifs. Red dots from matched donors and recipients signifies increases in the number of immunostimulatory motifs while blue dots between matched donors and recipients signifies a decrease in the number of immunostimulatory motifs.

Virus with greater numbers of TNF-α inducing motifs in their genomes are more likely to be found in the donor's semen compared to their blood.

Based on the previous findings, we have presented a trend where TNF-α inducing RNA motifs are enriched in viral sequences isolated from the recipient. Thereby suggesting that viruses with higher numbers of these TNF-α inducing motifs are more likely to be transmitted. However, we were unsure where these high pro-inflammatory viral RNA isolates were originating from considering the virus isolated from the donor's blood was less inflammatory. The previous datasets present viral RNA sequences attained from the serum compartment; however, during sexual transmission, the key anatomical sites at play are the genital organs. While both the serum and semen compartments are subject to immune pressure, the testes have been described as an immune privileged site[31,32]. This is important because sperm, which is protected within testicles from the body's immune system, has been implicated in playing a role in transmission[33]. Owing to that, we hypothesized that differences in the immunological environment may compartmentalize viruses and shield them from immune pressure. Fortuitously, within the Iyer *et al.* data set, transmission pair 1 had both serum and semen derived viral sequences available, allowing us to compare immunostimulatory motifs from two different compartments within an individual. Interestingly, these viral isolates displayed a similar trend whereby TNF-α inducing and IFN-α inducing motifs were higher in virus isolated in the semen compared to the blood (**Fig 2-5).** On the contrary, the CpG motif displayed a slight decrease between the blood and semen (**Fig 2-5**). This proof-of-concept helped solidify our hypothesis that the virus in semen may be compartmentalized and more readily transmitted compared to virus in the serum compartment.

To further investigate this phenomenon, we sought to analyze a larger cohort of individuals. To that end, we analyzed the datasets from Anderson *et al.* (2010) for which viral *env* sequences were attained from both the blood and semen of 14 individuals[34]. We observed that 71% (10/14) of individuals in this cohort had viral *env* sequences with an increased number of TNF-α inducing motifs in the semen compared to the blood (**Fig 2-6a, b**).

Similarly, albeit less impressive, 57% (8/14) of the viral *env* sequences from the semen showed a higher number of IFN-α and CpG inducing motifs compared to the *env* of viral isolates from the blood (**Fig 2-6a, b**). Altogether, this suggest that the virus within the semen may have viral RNA that is on average, more immunostimulatory compared to that found in the blood.

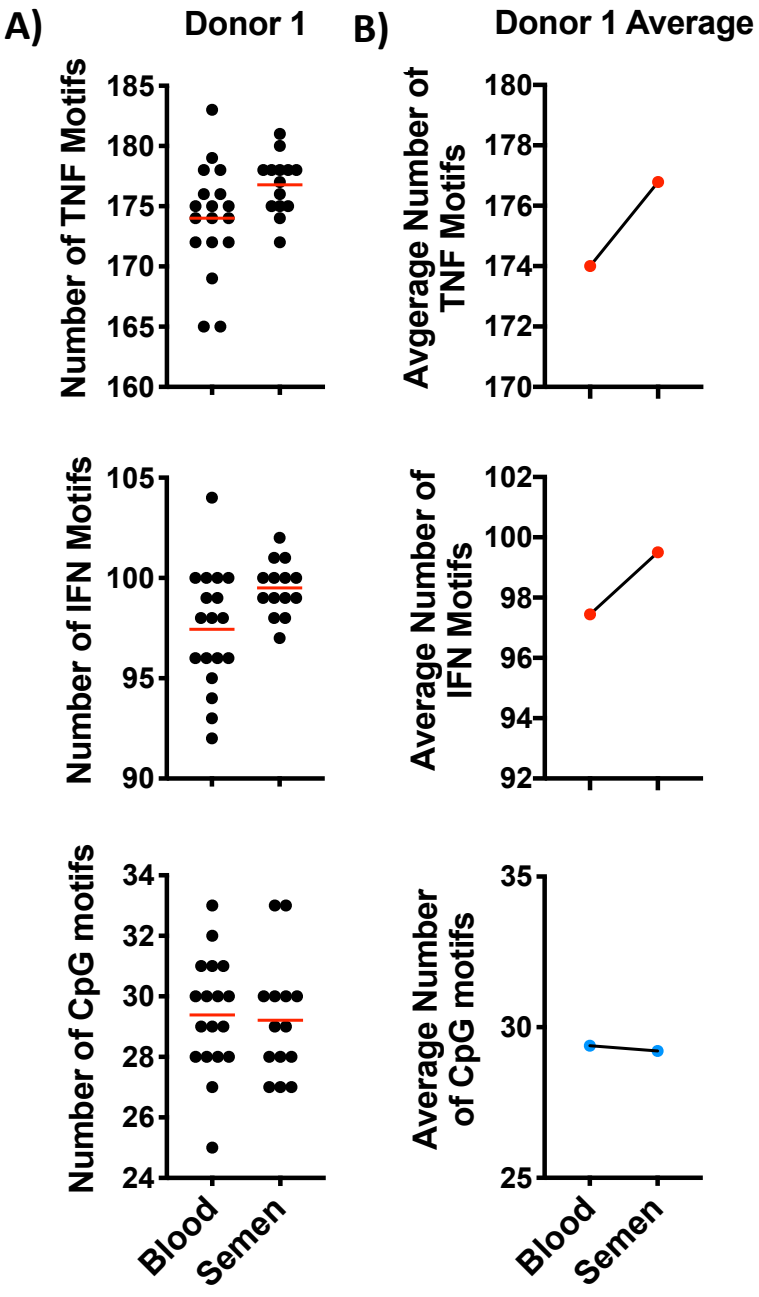


Figure 2-5: Donor 1 contained viral RNA with more pro-inflammatory motifs within the semen compared to the blood.

Blood vs Semen viral sequences were attained from Iyer *et al.* (2017) for Donor 1 from transmission pair 1 **(A)** The HIV *env* sequences were aligned to HXB2 and trimmed where necessary to ensure all regions of *env* (6225-8795 bp) used in the analysis were the same length (2570 bp). For all donor recipient pairs, graphs display number of TNF-α, IFN-α

and CpG motifs. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each dot represents a unique viral sequence. (B) Red dots from matched semen and serum samples signifies increases in the average number of immunostimulatory motifs while blue dots between matched semen and serum signifies a decrease in the average number of immunostimulatory motifs within the donor.

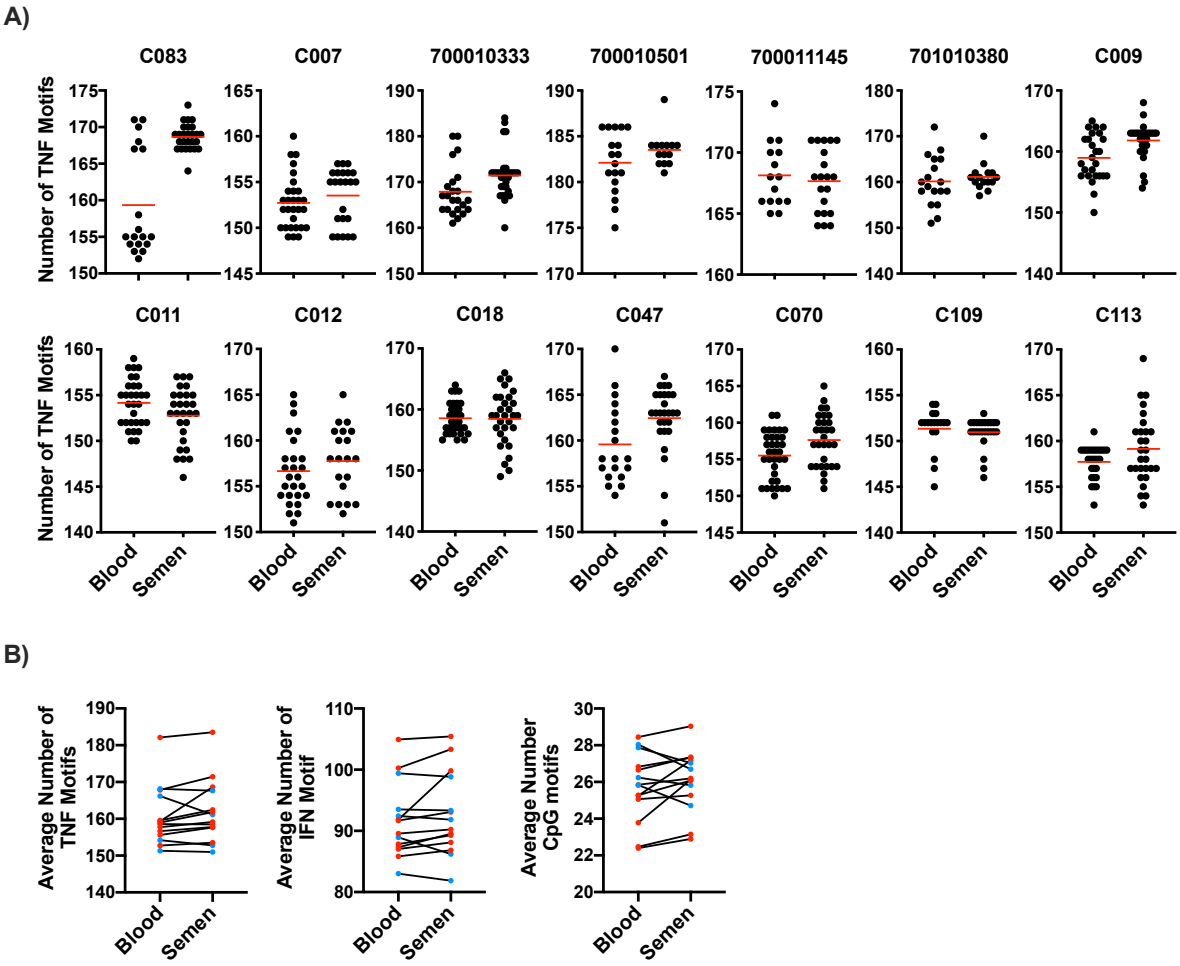


Figure 2-6: The semen contains viral isolates with a higher number of TNF-α motifs compared to the serum compartment of the same donor.

Blood vs Semen viral sequences were attained from Anderson *et al.* (2010). **(A)** 14 volunteers had *env* sequences of viral isolates collected from the serum and semen compartments. The graphs represent the number of TNF-α motifs present within the full *env* region (6225-8796 bp; 2570 bp in length) of the viral isolates based on alignment with HXB2. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. (**B**) Represents the average number of TNF-α, IFN-α and CpG motifs respectively for the serological and semen viral isolates. Red dots from matched semen and serum samples signifies increases in the number of immunostimulatory motifs while blue

dots between matched semen and serum signifies a decrease in the number of immunostimulatory motifs.

The virus with more TNF-α inducing motifs is more likely to cross the mucosal barrier and establish the systemic infection.

Previous work in the literature has demonstrated the compartmentalization of the virus between semen and blood within individuals[34]. In accordance to this compartmentalization, we observe that the viral RNA within the semen may be more immunostimulatory. Additionally, during transmission, the genetic swarm that the recipient is exposed to undergoes a genetic sieve whereby one viral variant, in 70-80% of infections, causes the systemic infection. As such, we hypothesized that the virus that would be transmitted would also contain viral RNA with a higher number of immunostimulatory motifs. Using the *env* sequences from Klein *et al*. (2018), we evaluated the difference in these immunostimulatory TNF-α inducing motifs in *env* sequences isolated from the cervix and blood within the recipient[3]. If our hypothesis is correct, we should see a higher number of immunostimulatory motifs in viruses that were transmitted (i.e. in the blood) than viruses in the cervix. Interestingly, we found that 69% (11/16) of individuals showed *env* sequences with a higher number of TNF-α inducing motifs in the blood compared to the cervix, with 13% (2/16) of individuals having an equivalent number of TNF-α motifs between the two compartments (**Fig 2-7**). There was no apparent trend observed with the number of IFN-α inducing or CpG inducing motifs (**Supplementary 2-2**). When we stratify the cohorts based on subtype infection, we observe that there are some differences. We found that in subtype A infections, 57% (4/7) of individuals showed *env* sequences with a higher number of TNF-α inducing motifs in the blood compared to the cervix, while subtype C and D showed 80% (4/5) and 75% (3/4) respectively. This may allude to some subtype differences, however, more samples for each subtype infection are needed to further corroborate this observation. Taken together, this suggests that the virus containing higher numbers of TNF-α inducing RNA motifs is more likely to enter the mucosa of the recipient and establish a disseminating infection.

Interestingly, within the Iyer *et al.* data set, we had access to blood and cervical viral sequences from two donors (transmission Pairs 4 and 5) of an FTM transmission, allowing us to compare immunostimulatory motifs from two different compartments within female donors. Here we observed that both female donors showed *env* sequences with a higher

number of TNF-α. IFN-α and CpG inducing motifs in the cervix compared to the blood (**Fig 2-8**). It appears this may contradict trends observed in the cohort from Klein *et al.*, however it is key to understand the differences in the observations. The initial analysis is observing the blood and cervix during a typical MTF transmission where the virus experiences the transmission bottleneck within the female recipient. On the contrary, the second analysis is observing the blood and cervix during an atypical FTM transmission where we observe potential compartmentalization of the T/F virus within the female genital tract –site of transmission.

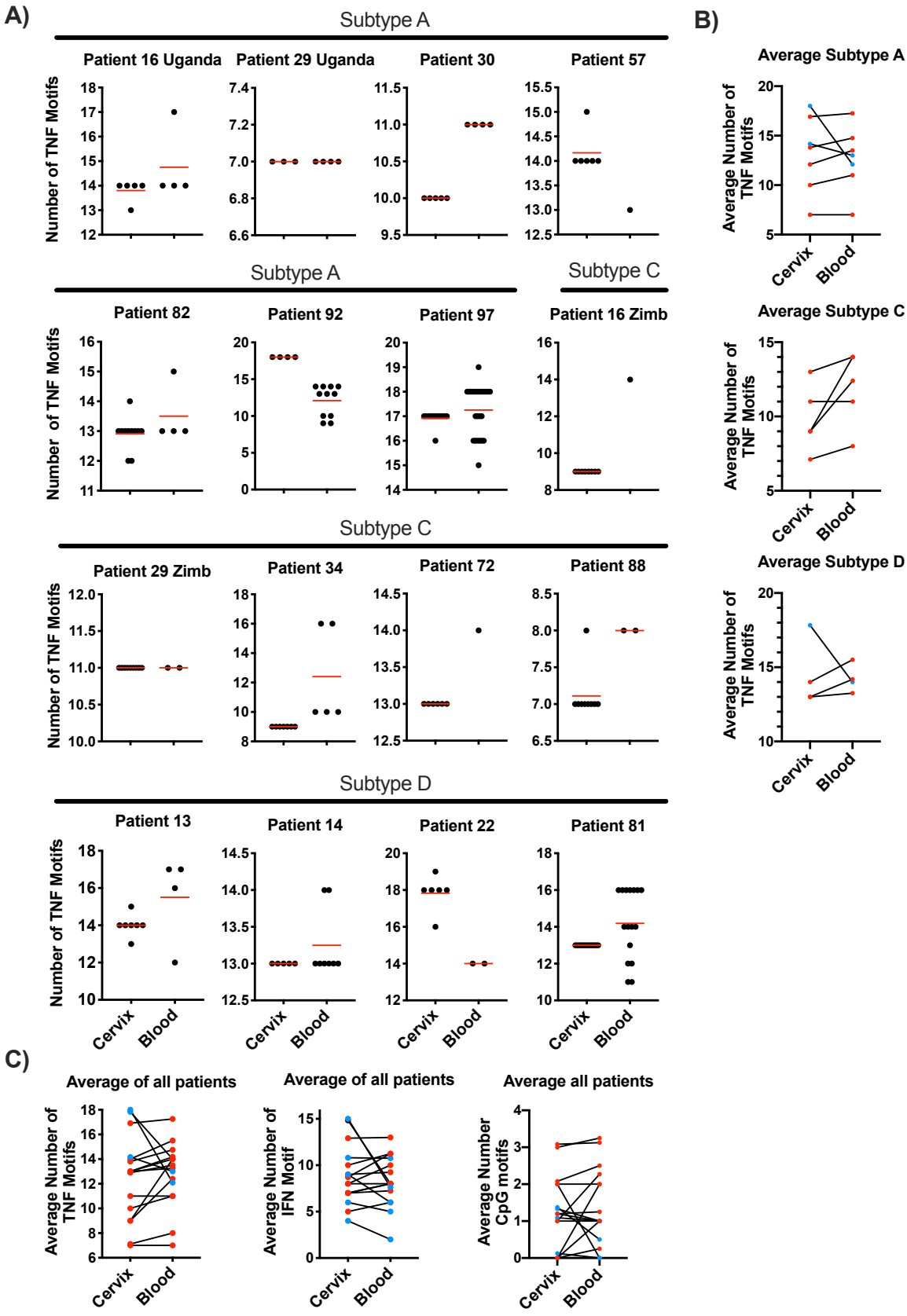

A)
Subtype A
Patient 16 Uganda
Patient 29 Uganda
Patient 30
Patient 57
Number of TNF Motifs
Subtype A
Subtype C
Patient 82
Patient 92
Patient 97
Patient 16 Zimb
Number of TNF Motifs
Subtype C
Patient 29 Zimb
Patient 34
Patient 72
Patient 88
Number of TNF Motifs
Subtype D
Patient 13
Patient 14
Patient 22
Patient 81
Number of TNF Motifs
Cervix
Blood
B)
Average Subtype A
Average Number of TNF Motifs
Average Subtype C
Average Number of TNF Motifs
Average Subtype D
Average Number of TNF Motifs
C)
Average of all patients
Average Number of TNF Motifs
Average of all patients
Average Number of IFN Motif
Average all patients
Average Number CpG motifs

Figure 2-7: The serum compartment contains viral isolates with a higher number of TNF-α motifs compared to the cervix.

Sequences were analyzed from the Klein *et al.* (2018) dataset. This cohort contains *env* sequences from the C2 to V3 region of *env* (405 bp in length). Sequences were aligned to HXB2 and trimmed to ensure the same sequence lengths were being analyzed. (**A**) Patients were stratified according to their subtype infection and then analyzed for the number of TNF-α motifs. Patients were infected with either subtype A, C or D. There were no AD recombinants. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each dot represents a unique viral sequence. (**B**) Graphs display the average number of TNF-α, IFN-α and CpG motifs according to subtype infection. Red dots from matched cervix and blood samples signifies increases in the number of immunostimulatory motifs while blue dots between matched cervix and blood samples signifies a decrease in the number of immunostimulatory motifs. (**C**) Graphs display the average number of TNF-α, IFN-α and CpG motifs of all patients. Again, red dots from matched cervix and blood samples signifies increases in the number of immunostimulatory motifs while blue dots between matched cervix and blood samples signifies a decrease in the number of immunostimulatory motifs

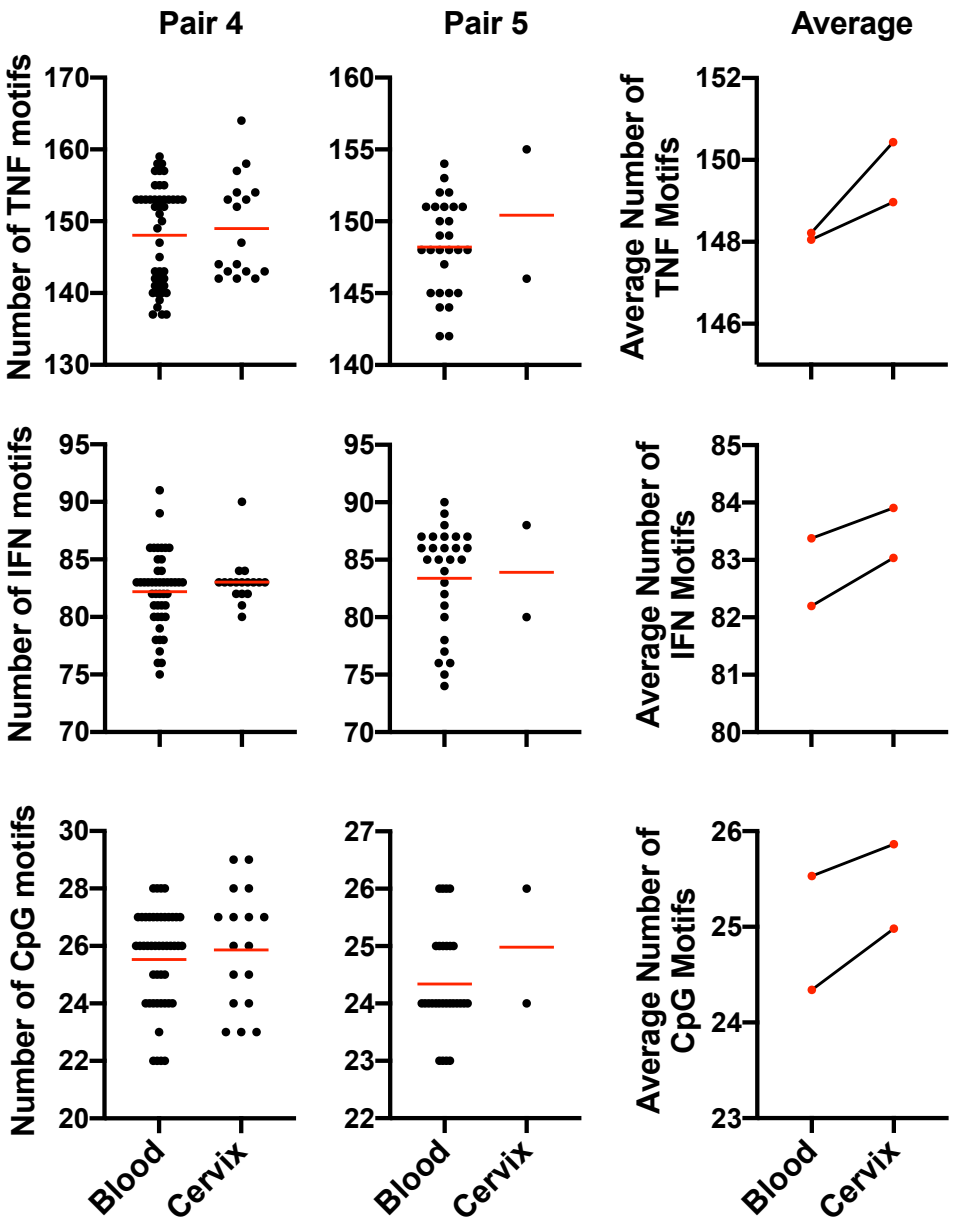


Figure 2-8: The cervix of female donors contains viral isolates with a higher number of TNF-α motifs compared to the blood.

Blood vs Cervix viral sequences were analyzed from the Iyer *et al.* (2017) dataset for transmission pairs 4 and 5. HIV *env* sequences (6225-8795 bp) were aligned to HXB2 and trimmed to ensure the same length of *env* was being analyzed (2570 bp). **(A)** Red line

represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each dot represents a unique viral sequence. **(B)** The average number of immunostimulatory motifs for both donors. Red dots from matched cervix and blood samples signifies increases in the number of immunostimulatory motifs while blue dots between matched cervix and blood samples signifies a decrease in the number of immunostimulatory motifs.

Transmitted/founder (T/F) viruses with more TNF-α motifs are more likely to be found within migratory cells and therefore establishes the systemic infection.

Based on analysis performed on the cohorts looking at transmission pairs and different anatomical compartments (serological vs semen; cervix vs serological), we hypothesized that successful transmission would occur with viruses containing RNA with a high number of TNF-α inducing motifs. To model mucosal HIV-1 transmission we performed multi-virus infection assays in *ex vivo* human explant tissues (**Fig 2-9a**). These assays were performed with chimeric viruses harboring *env* genes of viruses isolated from acute/early or chronic stage of infection. It is important to note that the acute and chronic viruses were produced on an NL4-3 backbone and differed solely on the inserted *env* sequence. This allowed for the genotypic differences and their ability to enhance or reduce transmission in the explant model to be attributed to *env*. Pools of four to five acute and three chronic Env chimeric viruses were competed in each competition assay to assess if any antigenic differences between *env* sequences isolated from different stages of HIV-1 infection influenced the viral transmission fitness. Analysis by Illumina MySeq deep gene sequencing revealed that the viral *env* sequences derived from acute/early viruses (B4, B7 and B9) were the predominant strains found in the migratory cell-PM1 T cell co-cultures in the vast majority of the assays (**Fig 2-9b**). Interestingly, analysis of the immunostimulatory pattern highlighted that the *env* sequences of B4, B7 and B9 displayed the highest number of TNF-α inducing motifs (**Fig 2-9c**). In contrast, viruses that displayed the lowest number of TNF-α inducing motifs B3 and B14 were detected only in very small quantities in migratory cell T cell co-cultures (**Fig 2-9b, c**). In general, we observed that the average number of TNF-α stimulating motifs in viruses that exited mucosal tissues and replicated in the migratory cell-PM1 T cell co-cultures (red circles) was higher than in those that failed to transmit (green circles) (**Fig 2-9d**). Collectively, this data indicates that viruses that contain RNA with a higher number of TNF-α inducing motifs have a greater chance of entering and then migrating through mucosal tissues via migratory cells. This we hypothesized, could be how HIV-1 migrates to the draining secondary lymphoid tissues for in-trans infection of susceptible CD4 T cells.

To further investigate the difference in antigenicity observed in acute vs chronic *env* RNA, we generated *env*-based RNA for the three most successful viruses (B4, B7, B9) and compared them to the three chronic *env* RNAs using a THP-1 dual reporter cell line. The dual reporter THP-1 cell line allows the concomitant study of NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) pathway and the IRF (interferon regulatory factor) pathway. Determination of IRF and NF-κB activation in the THP-1 cultures demonstrated that acute virus derived *env* RNA stimulated the THP-1 dual reporter cells significantly more than the chronic virus derived *env* RNA (**Fig 2-9e**). These viruses (B4, B7, B9) therefore had the most antigenic *env* RNA and out competed all the other viruses, further supporting the evidence from all the other cohort data presented in this manuscript. Together this may shine light on the role the RNA of the virus may play in aiding transmission.

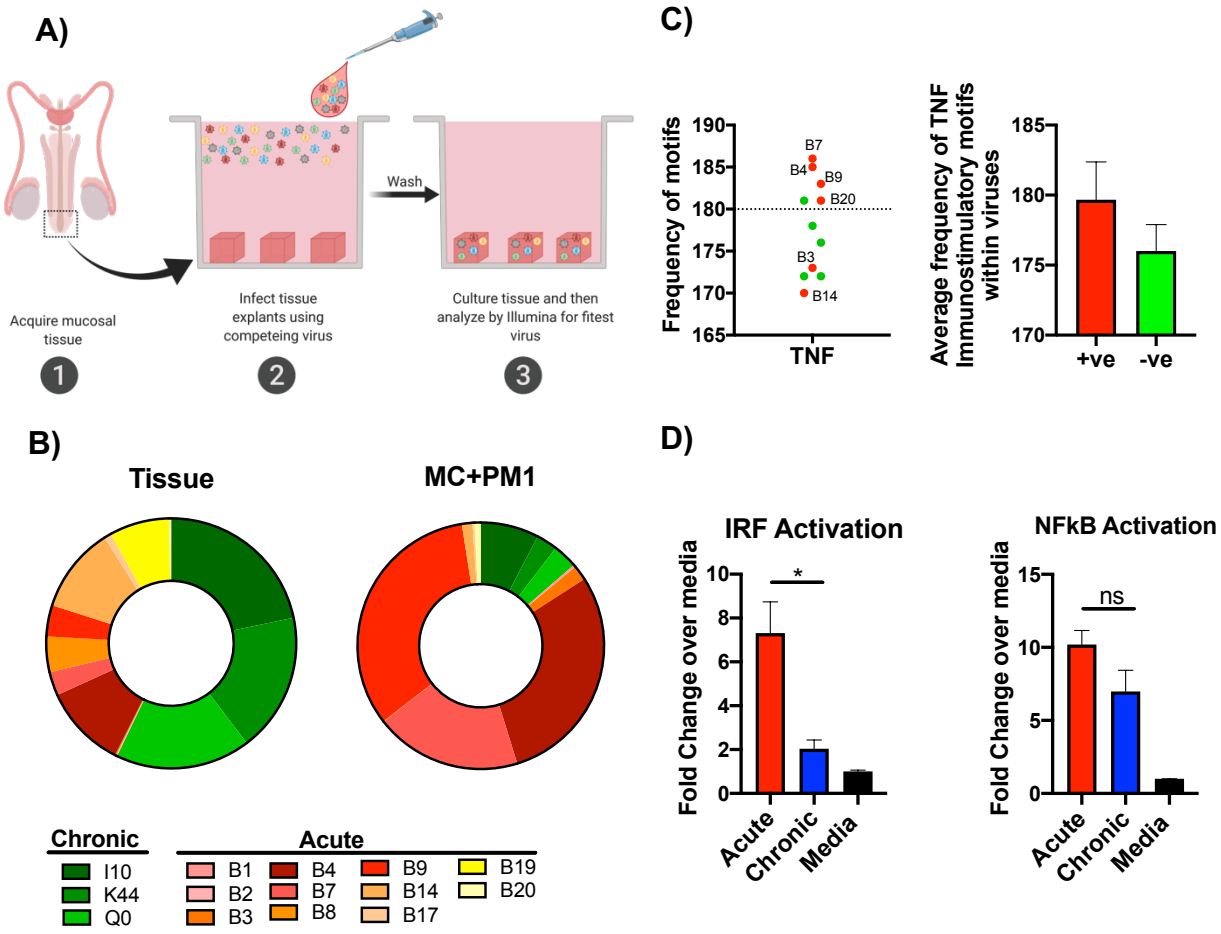


Figure 2-9: High number of TNF-α motifs enhances transmission in an explant model.

The *env* from viruses that have been identified previously as acute/early (*n*=20) and chronic (*n*=3) infectious viruses were PCR amplified and recombined into a pREC NL4-3 backbone vector. This ensured that the only genetic difference between all viruses was from the *env* gene and not any other viral protein. (**A**) Five acute/early and 3 chronic viruses were pooled together in a competition assay using cervical and penile tissue. Note, chronic viruses were reused with different combination of acute/early viruses. RNA isolated from the tissue and migratory cells were subject to Illumina sequencing and *gag* qRT-PCR for quantification of the number of copies in each compartment. (**B**) Composition of acute/early and chronic viruses detected in the Tissue and MC+PM1 (migratory cells). (**C**) Number of TNF-α motifs in the *env* of the acute/early viruses. (**D**) Average number of TNF-α motifs in *env* in all acute/early vs chronic viruses. (**E**) Acute/early and chronic *env*

were PCR amplified with a T7 overhang. A T7 *in vitro* transcription kit was used to make RNA that was complexed with a DOTAP transfection reagent to stimulate THP-1 dual report cells (n=6). Cell culture supernatant was collected 24h post stimulation for Interferon Regulatory Factor (IRF) and NF-κB activation. Luminosity was measured for IRF using QUANTI-Luc to measure the activity of Lucia luciferase in culture supernatant. NF-κB activation was measured using a QUANTI-Blue colorimetric enzyme assay that determines alkaline phosphate (AP) activity in cell culture supernatant. A Cytation 5 instrument was used to detect reactions. Mann-Whitney statistical analysis was performed *$p \leq 0.05$.

Table 2-1: Patient information from Oberle *et al.*

Patient	Stage of Infection/Treatment Given
Donors from pairs 7 and 8	Put on antiretroviral treatment immediately after infection; HIV transmitted upon virus rebound after structured treatment interruption
Donors from pairs 1 to 5	Chronic stage of infection
Recipients from pairs 1 to 5	Acute stage of infection
Donors from pairs 6 and 9	Recently infected (early stage of infection
Recipients from pairs 6 and 9	Acute stage of infection

Discussion

When studying heterosexual transmission of HIV-1, it became apparent that a single viral variant was responsible for establishing an infection within the recipient in the vast majority of cases[2]. The success of this single variant, referred to as the T/F virus, has been attributed to its unique phenotypic features that allow it to overcome a stringent transmission bottleneck that occurs during heterosexual transmission[2,5,12,14,35,36]. In this study, we have described a previously under-explored role of viral RNA in potentially enhancing transmission of the HIV-1 virus. We asked a simple question: is there a particular trait within the viral genome that can provide the virus with enhanced transmission fitness? We gathered information from multiple different infection and transmission data sets, that evaluated viral sequences from various anatomical locations – semen, cervix, donor blood and recipient blood– and we observed that there is a consistent trend where the virus appears to favour a high number of TNF-α inducing motifs for transmission. Through our analyses of the various cohorts, we have demonstrated a potentially unique feature to T/F viruses that can aid in understanding HIV transmission biology and be exploited in future vaccine development.

The majority of work done on understanding HIV-1 transmission has focused on the phenotypic differences of the Env glycoprotein. These differences include changes in the glycosylation pattern, changes in the length of variable loops, increased Env density on the surface of the virions and changes in co-receptor usage (CXCR4 to CCR5)[2,12–14,36]. Yet, it is still unclear what properties distinguishes a true T/F virus from the rest of the swarm. It seems that beyond the use of CCR5, there is no single major genetic or phenotypic signature that is consistent for T/F viruses. Rather, an array of potential phenotypic variations collectively or in some unclear combination is associated with enhanced virus transmission for subtypes A, C and possibly B. Our results describe a novel role for HIV-1 RNA, whereby the immunostimulatory potential of the viral RNA can enhance virus transmission during coital acts and should be considered along with the previously identified factors governing viral fitness.

Through the analysis of *env* sequences, our first key observation was derived from the study of transmission pairs. Here, we compared the numbers of immunostimulatory RNA

motifs found in viral sequences isolated from the serological compartment within donors and their respective recipients. The analysis indicated a trend where viral isolates within the recipient contained more TNF-α inducing motifs than that found in the donor (**Fig 3, 4**). Interestingly, Iyer *et al.* demonstrate that IFN resistance is a key determinant in HIV fitness; however, we see that the presence of TNF-α inducing motifs in viral *env* enhances viral transmission[29]. This provided some indication that the inflammatory nature of the viral RNA may in fact be playing a role in aiding transmission. Interestingly the donor had a larger spread in the immunostimulatory TNF-α inducing motif counts while the RNA from the recipient seemed to be more homogenous. This is interesting as it aligns with what we understand of HIV transmission. An accumulation of studies has shown that in 80% of heterosexual transmissions[2,5,37], 60% in MSM[38], 40% in IDUs[39,40] and 68% on maternal-infant transmission, depending on intrauterine vs intrapartum routes[41], that the subjects are productively infected by a single viral genome. While a limitation to this is the inability to track the T/F virus from the donor to the recipient, it does however display the homogenous nature of early infections in recipients. Indeed, the high degree of genetic diversity found in late or chronic viruses (i.e. the donor) should have more diversity in the immunostimulatory RNA motif counts, whereas the homogenous population of acute viruses, by virtue of the multiple genetic bottlenecks during transmission, should present a narrower count in TNF-α inducing motifs.

An important question however was where were the highly pro-inflammatory viral RNA sequences coming from? This leads us to the second key observation in our analysis where we evaluated the immune-stimulatory potential of viral RNA isolated from the semen vs blood and cervix vs blood of donors. Here the HIV-1 *env* sequences demonstrated that the majority of the viral RNA within the semen and cervix contains more pro-inflammatory TNF-α inducing motifs than the viral RNA in blood from within the same patient. This suggests that viral compartmentalization could be occurring in the different anatomical sites of the host. This has many implications for understanding HIV biology and for vaccine design as isolation and study of virus from the semen and cervix is the more likely source of transmitting virus during coital acts. However, it should be noted that we do notice that not all of the male donors' demonstrated this trend and it is possible that in some cases the

transmitted virus came from another compartment i.e. the blood. The notion that HIV-1 can be compartmentalized is controversial. While some studies have demonstrated compartmentalization between the semen and blood[34], there have also been studies that have shown a poor selective barrier and therefore the free flowing of virus between the two anatomical locations[42]. Similar to the semen, there is evidence that there is both intermittent and persistent compartmentalization in the FGT[43]. The presence of *env* sequences with higher pro-inflammatory motifs in the cervix of the female donor can be explained by reseeding of the FGT overtime in female donors[44]. Regardless, the virus that is likely to infiltrate the mucosal barrier during sexual transmission, is likely to arise from the semen or cervix rather than the blood.

Our third key observation is derived from the analysis of sequences from recipient cervix and blood. If the pro-inflammatory capability plays a role in transmission, we would expect that the viral RNA within the recipient's blood compared to their cervix to be more inflammatory. Indeed, our analysis of a cohort of women with primary HIV-1 infection demonstrated this trend. While the mucosal epithelium provides a robust barrier to HIV-1 transmission by preventing entry into submucosa, cytokine imbalances within this environment can provide an advantage for the virus. Interestingly, exogenous treatment of Langerhans cells (LCs) with TNF-α has shown an increase in HIV-1 replication within these cells –an important subset of migratory immune cells that have been implicated in trafficking HIV-1 particles to target CD4 T cells in secondary lymphoid organs[21]. TNF-α is a pro-inflammatory cytokine that can act in both a paracrine and endocrine fashion. In particular, the effect of TNF-α on LCs has been shown to up-regulate levels of co-stimulatory molecules such as CD40, CD54, CD86, maturation markers including CD83, DC-LAMP and lymph node homing receptors such as CCR7 all while down-regulating Langerin in a dose-dependent manner[45,46]. Interestingly, TNF-α has been shown to induce changes to the mucosal environment that increase susceptibility to recurrent infections. The presence of TNF-α within the mucosal milieu can induce expression of chemokines, such as CXCL1, that can recruit immune cells like neutrophils. While beneficial to the host to fight infections, pro-longed exposure of TNF-α enhances the ability of the virus to transmit through the mucosal membrane. The recruitment of innate immune cells, while prompted

to mediate viral clearance, may in fact aid the virus in crossing the submucosa. As such, it is plausible that viral RNA that contains more immunostimulatory TNF-α inducing motifs, can stimulate the production of more TNF-α, thereby creating a tissue environment conducive to successful infection. A recent publication by Deleage *et al.* further demonstrated that initial local replication at the mucosal portal of entry is important for successful systemic dissemination[44]. This further corroborates the results presented in this manuscript as it indicates a key moment where the RNA of the virus can play a role in altering the mucosal milieu. The precedence of replication to dissemination demonstrates that the virus does not move directly to the systemic compartment and that viral RNA exposure to the host during those early infectious events might be necessary for virus transmission[44]. The pro-inflammatory composition of the viral RNA could therefore play a significant role in enhancing infection and dissemination during the earliest stages of transmission. While this trend was observed for the majority of individuals, it was not observed for all. For the women that did not align with the trend, a plausible reason for this may be confounding factors such as the presence of clinical or subclinical sexually transmitted infections (STIs), or even the use of hormonal contraceptives, as these can all lower the barrier for transmission[5,37]. The presence of STIs, such as HPV, can abrogate the integrity of the mucosal barrier by breaking epithelial lining[47]. As a result, reduced epithelial barrier integrity can allow for more viral variants to cross the mucosal tissue and cause the systemic infection. In addition, the presence of STIs can increase inflammation through TNF-α production and increase the availability of target activated CD4 T cells and migration of APC populations. This further corroborates the potential role of inflammation and TNF-α as its presence enhances infection of HIV-1. Additional support for the role of inflammation on transmission comes from SIV models. The reduced expression of cytokines like IL-2, TNF-α and IFN-γ in sooty mangabeys may provide a protective niche, whereby non-pathogenic SIV and sooty mangeby can coexist[48]. On the contrary, humans and macaques produce a more immune activated environment leading to disease pathogenesis[49].

While most of the results reported here look at already published sequence data from various cohorts, the tissue explant model experiment provides additional supportive

evidence that the pro-inflammatory nature of the viral *env* sequence enhances infection. The acute/early and chronic constructs are based on the same HIV-1 backbone (lab adapted strain NL4-3) and differ only in the inserted *env* gene. This allowed us to measure differences in both the transmission fitness associated with Env and immunostimulatory potential of the *env* RNA. Regardless of any other pool of viruses within the competition, the early/acute *env* viruses B4, B7 and B9 constantly out competed all viruses in all competitions. These same viruses had the most TNF-α inducing motifs and therefore represent an *in vitro* model demonstrating how the antigenicity of the viral RNA can play a role in transmission success. It should be noted that all viruses in that assay had previously had their replicative levels assessed and were shown to be similar in PBMC cultures[50,51]. Therefore, the fact that B4, B7 and B9 were serial winners in the mucosal transmission assay, was not due to them having any replicative advantage over the other viruses, nor was it due to the fact that the other viruses had Env glycoproteins that had reduced functionality.

To our knowledge, the field has extensively worked on understanding changes in Env glycosylation in response to immune pressure. A common shift in PNLGs, due to the ensuing immune response, is the selection of viral variants with a N332 to N334 glycan shift as this provides resistance to V3-glycan antibodies[52,53]. However, when we attempted to align the location of these multiple 4-mer TNF-α inducing motifs to PNLGs and CpG motifs, we noticed that there was no overlap between the pro-inflammatory motifs and PNLGs, indicating that changes to motif number may not influence PNLGs (**Supplementary 2-3**). While we know that these motifs decrease overtime (**Fig 2-2)** and are present in higher amounts in viruses that are likely to transmit (**Fig 2-3, -4, -5, -6, -7, -8, -9**), we are unsure of the specific selection pressure patterns that drive a less-immunogenic RNA. Contrary to our assumption, the conserved regions, in particular C3, contained the highest frequency of motifs (**Supplementary 2-3**). With the majority of mutations accumulating within the variable region, the presence of these 4-mer motifs within conserved regions may present a different selection pressure. This implies that the TNF-α inducing motifs may be an independent variable that is under immune pressure within the host and that changes in these motifs may not align with alterations in the

glycosylation pattern of Env. The fact that that there is no overlap with CpG motifs (**Supplementary 2-3**) may show that mutations targeted by APOBEC present little to no harm in these motifs to mediate or lose their immunostimulatory affects. A paper by Geller *et al.* showed that G→A substitutions occurred 77% of the time (77/100 pooling *env* and *int-vif-vpr*) or C→T substitutions mediated by host A3G and A3D/F/H[54]. These spontaneous nonsynonymous substitutions, however, may help explain changes in the presence of TNF-α and IFN-α inducing motifs as these 4-mer sequence motifs are known to be AU and GU rich respectively. Changes to these known 4-mer motifs will have an impact on their ability to potently stimulate their respective TLRS as shown by Forsbach *et al*[26]. However, it is important to note that this analysis is limited to mapping these motifs on acute/early viruses. In the future, it would be important to map the location of TNF-α motifs in comparison to PNLGs and CpG motifs in sequences from later timepoints to assess any shifts in particular motifs or bias in location.

The data presented here provides a novel understanding of HIV-1 RNA in the context of fitness transmission. We show that viral RNA that contains more TNF-α inducing motifs is more likely to successfully transmit and cause the systemic infection. While the work presented is intriguing, it does have certain caveats that need to be considered. Firstly, this analysis has only been applied to HIV-1 and understanding transmission. The implications of viral RNA presented in this study could have interesting applications to other viral infections. For instance, it would be interesting to apply this analysis to other viral infections such as HTLV, Measles and EBV whereby understanding of pathogenesis or transmission can be garnered through extensive analysis of viral genomic sequences. Moreover, it would be interesting to evaluate the immunostimulatory motif numbers in the SIV models of infection, especially in light of the differences in infection severity in old world and new world non-human primates. Secondly, we believe that although the RNA motifs inducing TNF-α responses provide strong evidence for a role in transmission, it may be possible to strengthen these analyses by integrating the results from the IFN-α and CpG analyses. This could provide a more comprehensive analysis if all were taken together as a multifactorial process may be at play. Thirdly, although the motifs used in these analyses are separated into TNF-α and IFN-α biased motifs, very few of the motifs exclusively

produce one cytokine in the absence of the other. Hence the analyses could be made more rigorous by clearly identifying those motifs that play a key role in inflammation.

Overall, we provide a potential feature that distinguishes T/F viruses from the rest of the viral swarm. This we believe could help advance our understanding of HIV-1 transmission and also be used to help select and design the next generation vaccine candidates, particularly those directed against T/F viruses.

Methods

DNA sequences

Sequences from the different anatomical sites and transmission pairs were attained from the aforementioned studies in the manuscript and can be found summarized in **Table 2**. The whole genome of the common HIV-1 group M reference strains A, B, C and D were obtained from the Los Alamos National Laboratory HIV database (https://www.hiv.lanl.gov) (Accessions: Subtype A: DQ676872, AB253421, AB253429, AF2886237-8, GU201516; Subtype B: K03455 (HXB2), AY423387, AY173951, AY331295; Subtype C: U52953, U46016, AF067155, AY772699; Subtype D: K03454, AY371157, AY253311, U88824).

Bioinformatic analysis

To evaluate any differences between specific regions within the genome of HIV-1, all reference strains were aligned against HXB2 *gag* (790-2292 bp), *pol* (2085-5096 bp), *env* (6225-8795 bp), *vif-vpr*-5'*tat-vpu* (5041-6310 bp), and *nef* (8797-9417 bp), trimmed and de-gapped accordingly using the BioEdit software. The *env* sequences from the different cohorts (**Table 2**) were also aligned against HXB2 *env*, trimmed to ensure similar region of *env* analysis and de-gapped for subsequent analysis. After alignment and prior to trimming, we manually evaluated each sequence to identify and discard any sequences that did not share full *env* coverage. Of exception, cohorts analyzed from Shankarappa, *et al.* (1999) and Katja, *et al* (2018), did not contain full *env* coverage and regions C2-V5 and C2-V3, respectively, were analyzed.

To enumerate the presence of genetic sequences known to be immunostimulatory and triggers of innate pathogen recognition receptors, we developed an in-house Python script. This script is run on PyCharm CE and is able to identify and count the frequencies of known TNF-α (n=17; ttgt, tttc, tttt, tgtt, ctgt, tatt, gttt, attt, atgt, cttt, ttta, ttgc, tctt, attc, ttct, tctc, ttga), IFN-α (n=11; gttc, ttgt, tttt, tttc, gttt, tgtt, cttt, attt, tctt, ttta, attc) and CpG motifs in thousands of sequences at once. The resulting matrix was then graphed using GraphPad Prism8 to analyze trends.

To map the presence of TNF-α (n=17) and CpG motifs in the three acute viruses (B4, B7 and B9) and HXB2, we designated a unique number from 1 to 17 to each TNF-α motif sequence (designated unique number for CpG motifs = 18) and manually located each motif in the *env* sequences. To identify PNGLs (designated unique number = 19), the N-GlycoSite application from the Los Alamos National Laboratory HIV database (https://www.hiv.lanl.gov) was used. GraphPad Prism8 was then used to graph out the location of these motifs through the full env gene.

Ethics statement

Human penile and cervical tissue was collected with written consent according to the local research committee guidelines. All tissues were obtained under protocols approved by the Imperial College NHS Trust Tissue Bank and the National Research in accordance with the Human Tissue Act 2004.

Tissue Explant Model Experiment

Refer to the following publication for detailed account on performing the tissue explant model experiment (Katja, K *et al.* 2020). Briefly:

Creating the acute/early and chronic chimeric viruses: Eleven early/acute subtype B envelopes were obtained from the Center of HIV-1/AIDS Vaccine Immunology (CHAVI)[2]. Three chronic subtype B envelopes were obtained from a Belgium cohort[50,51]. Chimeric viruses were created using a previously established yeast-based recombination system[55], where the different envelope genes were inserted into pREC_nfl_NL4-

3_Δenv/URA3. As a result, the pREC constructs differed solely on the inserted *env* sequence. HEK293T cells were co-transfected with a rescue vector pCMV_cplt and virus was propagated in U87.CD4.CCR5 cells. The acute viruses are referred to as: B1, B2, B3, B4, B7, B8, B9, B14, B17, B19 and B20. The chronic viruses are referred to as: I10, K44 and Q0. Reverse transcriptase activity of each virus was used to quantify infectivity for future tissue infections.

Transmission fitness assay and Sequence data analysis: Frozen endocervical explants were acquired through the National Disease Research Interchange (NDRI) and cut in approximately 3mm X 3mm X 3mm pieces and placed in each well of a 48-well plate. Pools of four to five acute and three chronic Env chimeric viruses (150 IU per virus; as determined by reverse transcriptase activity) were mixed in culture containing 3-5 pieces of cut explant tissue for 3 hours at 37°C. Tissue was washed three times in complete RPMI 1640 media (Sigma). Migratory cells (MC) from the tissue were harvested 24 hours later and co-cultured with 10^5 PM1 CD4 T cells. On day 10, tissue was lysed with 300μL of 1% Triton-X100 supplemented with proteinase K over night at 56°C and cells from the MC-PM-1 T cell co-culture were lysed with 300μL of 1% Triton-X100. DNA was extracted using PureLink Pro 96 Genomic DNA kit (Invitrogen). Each multi-virus competition was performed in two donors in triplicate. Following DNA extraction, PCR amplification of the C2-V3 envelope region was performed and products were ligated with barcode oligonucleotides, purified and subject to Next-Generation-Sequencing (NGS) (Roche 454 GS Junior sequencer). Using SeekDeep, a custom pipeline that sorts sequences based on barcodes and aligns sequences within barcode bins, viruses within each competition were quantified.

Env Antigenicity

The Env gene from B4, B7 and B9 (acute/early) and Q0, K44, I10 (chronic) viruses were amplified using primers T7_E80 (TAATACGACTCACTATAGCCAATTCCCATACATTATTGTG) and E125 (CAATTTCTGGGTCCCCTCCTGAGG). The following PCR cycling conditions were used: cycle 1 consisted of a 98°C step for 2 minutes; cycle 2 consisted of 3 stages repeated

10 times and consisted of a 98°C step for 10 seconds, followed by a 55°C step for 30 seconds, and then 72°C for 8 minutes; cycle 3 consisted of 3 stages repeated 25 times at 98°C for 10 seconds, followed by 55°C for 30 seconds and finally 78°C for 8 minutes. Lastly, cycle 4 consisted of a 72°C step for 7 minutes before incubation at 4°C. The PCR amplification reaction consisted of: 1X Phusion buffer (Thermo Fisher), 10µM of dNTP mixture (Thermo Fisher), 1 U of Phusion DNA polymerase (Thermo Fisher), 0.5µM of forward and reverse primers.

Following gel extraction, the amplicons were used for downstream *in vitro* transcription using a T7 MEGAscript™ transcription kit (ThermoFisher) according to the manufacturer's instructions. Briefly, 2µL of each ribonucleotide, 2µL of 10X reaction buffer, 0.15µg of PCR product, 2µL of T7 enzyme mix and DNA/RNA free water (Sigma) was added to yield a total reaction volume of 20µL. The reaction was incubated at 37°C for 16 hours in a PCR Thermocycler to increase number of transcription initiation events.

To test the antigenicity of the various *env* RNA, THP-1 Dual™ reporter cells (Invitrogen) were seeded 10^5 cells/well in a 96-well plate and stimulated with 2.5µg of *env* RNA complexed with 5µg of dioleoyl-3-trimethylammonium propane (DOTAP) (Roche). Prior to stimulation, the RNA: DOTAP complex was incubated in incomplete RMPI for 15 minutes at room temperature. Cells were cultured with the stimulant for 24 hours at 37°C and 5% CO_2 before assessing the supernatant for IRF (via detection of secreted luciferase) and NFκB (via detection of secreted embryonic alkaline phosphatase (SEAP)) activation according to manufactures instructions. A Cytation5 reader was used to measure absorbance readings at 630 nm. Four independent stimulations were performed for the measure of IRF and NFκB activation.

Table 2-2: Table with the Cohorts used in the study.

Paper	https://www.hiv.lanl.gov	Shankarappa, *et al.* Journal of Virology (1999)	Iyer, *et al.* PNAS (2017)	Oberle, *et al.* Retrovirology (2016)	Anderson, *et al.* PLoS pathogens (2010)	Katija, *et al.* PLoS pathogens (2018)	https://www.hiv.lanl.gov
Purpose	Reference strain analysis	Longitudinal analysis of viral RNA	Transmission pair analysis: male-to-male (MSM) and female-to-male (HET) transmissions Semen vs Blood analysis	Transmission pair analysis: male-to-male (MSM) and female-to-male (HET) transmissions	Semen vs Blood analysis	Cervix vs Blood analysis	TNF-α changes over time
Subtype	A, B, C, D	B	B (N=3) and C (N=4)	B	B (N=4) and C (N=12)	A (N=7), C (N=5), D (N=4)	B
Region of Env analyzed	Whole genome	C2-V5 region of Env	Full Env	Full Env	Full Env	C2-V3 region of Env	Full Env
Cohort Size	A (N=6) B (N=4) C (N=4) D (N=4)	N=9	N=8 pairs	N=9 (N=6 MSM; N=3 HET)	N=14	N=16	N=2*
Data Set	Los Alamos	Multicenter AIDS Cohort Study (MACS)	CHAVI- 001 acute HIV-1 infection cohort	Zurich Primary HIV Infection study (ZPHI) and/or the Swiss HIV Cohort Study (SHCS)	men with chronic HIV-1 infection in Lilongwe, Malawi (N=12) or from the CHAVI 001 (N=4) clinical study	Hormonal Contraception and HIV-1 Genital Shedding and Disease Progression among Women with Primary HIV Infection (GS)	Los Alamos

*note $n = 2$ refers to the two Fiebig stages analyzed; total of 1850 full *env* sequences were analyzed (N= 1364 for Fiebig stage 5 and N=486 for Fiebig stage 6).

Acknowledgments

The authors would like to thank the authors of the cited studies for publishing their sequences in the database. Without free access to these sequences, we would not be able to do the analysis presented in this manuscript. We encourage continued publication of sequencing results. The authors would also like to thank Dr. Stephan Barr and Dr. Rodney DeKoter for their useful comments in improving the quality of the overall manuscript. We also like to extend our thanks to the participants who volunteered for their donations of blood and tissue; this study would not have been possible without their support.

Author contributions

R.P and J.F.S.M wrote the manuscript. J.F.S.M and C.N.W wrote the python script. R.P downloaded and performed the analysis of the various datasets. K.K performed the tissue explant competition assay. R.P and C.N.W performed the T7 *in-vitro* transcription and R.P performed the THP-1 stimulation experiments. E.J.A, R.T and K.K helped edit the manuscript.

Competing interests

The authors declare no competing interests.

Material and correspondence

Correspondence to Dr. Jamie F.S. Mann

References

1. Patel, P. *et al.* Estimating per-act HIV transmission risk: A systematic review. *AIDS* **28**, 1509–1519 (2014).
2. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7557 (2008).
3. Klein, K. *et al.* Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS Pathog.* **14**, e1006754 (2018).
4. Williams-Wietzikoski, C. A. *et al.* Comparisons of Human Immunodeficiency Virus Type 1 Envelope Variants in Blood and Genital Fluids near the Time of Male-to-Female Transmission. *J. Virol.* **93**, (2019).
5. Haaland, R. E. *et al.* Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1. *PLoS Pathog.* **5**, e1000274 (2009).
6. Carlson, J. M. *et al.* Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science (80-.).* **345**, (2014).
7. Wawer, M. J. *et al.* Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *J. Infect. Dis.* **191**, 1403–1409 (2005).
8. Quinn, T. C. *et al.* Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. *N. Engl. J. Med.* **342**, 921–929 (2000).
9. Chakraborty, H. *et al.* Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: A probabilistic empiric model. *AIDS* **15**, 621–627 (2001).
10. Freeman, E. E. *et al.* Herpes simplex virus 2 infection increases HIV acquisition in men and women: Systematic review and meta-analysis of longitudinal studies. *AIDS* **20**, 73–83 (2006).
11. Stekler, J. *et al.* HIV Dynamics in Seminal Plasma during Primary HIV Infection. *AIDS Res. Hum. Retroviruses* **24**, 1269–1274 (2008).
12. Derdeyn, C. A. *et al.* Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**, 2019–22 (2004).
13. Chohan, B. *et al.* Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J. Virol.* **79**, 6528–31 (2005).
14. Joseph, S. B., Swanstrom, R., Kashuba, A. D. M. & Cohen, M. S. Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. *Nature Reviews Microbiology* **13**, 414–425 (2015).
15. Poltronieri, P., Sun, B. & Mallardo, M. RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. *Curr. Genomics* **16**, 327–335 (2015).

16. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition by the innate immune system. *Int. Rev. Immunol.* **30**, 16–34 (2011).

17. Yoo, J. S., Kato, H. & Fujita, T. Sensing viral invasion by RIG-I like receptors. *Current Opinion in Microbiology* **20**, 131–138 (2014).

18. Lester, S. N. & Li, K. Toll-like receptors in antiviral innate immunity. *Journal of Molecular Biology* **426**, 1246–1264 (2014).

19. Heil, F. *et al.* Species-Specific Recognition of Single-Stranded RNA via Till-like Receptor 7 and 8. *Science (80-.).* **303**, 1526–1529 (2004).

20. Chen, L. *et al.* Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget* **9**, 7204–7218 (2018).

21. De Jong, M. A. W. P. *et al.* TNF-α and TLR agonists increase susceptibility to HIV-1 transmission by human Langerhans cells ex vivo. *J. Clin. Invest.* **118**, 3440–3452 (2008).

22. Matsuyama, T. *et al.* Enhancement of HIV Replication and Giant Cell Formation by Tumor Necrosis Factor. *AIDS Res. Hum. Retroviruses* **5**, 139–146 (1989).

23. Osborn, L., Kunkel, S. & Nabel, G. J. Tumor necrosis factor α and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor κB. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2336–2340 (1989).

24. Kimber, I. & Cumberbatch, M. Stimulation of Langerhans cell migration by tumor necrosis factor α (TNF-α). *J. Invest. Dermatol.* **99**, 48–50 (1992).

25. Vaidya, S. A. *et al.* Tumor necrosis factor α is associated with viral control and early disease progression in patients with HIV type 1 infection. *J. Infect. Dis.* **210**, 1042–1046 (2014).

26. Forsbach, A. *et al.* Identification of RNA Sequence Motifs Stimulating Sequence-Specific TLR8-Dependent Immune Responses. *J. Immunol.* **180**, 3729–3738 (2008).

27. Shankarappa, R. *et al.* Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* **73**, 10489–502 (1999).

28. Pinkerton, S. D. Probability of HIV transmission during acute infection in Rakai, Uganda. *AIDS Behav.* **12**, 677–684 (2008).

29. Iyer, S. S. *et al.* Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E590–E599 (2017).

30. Oberle, C. S. *et al.* Tracing HIV-1 transmission: Envelope traits of HIV-1 transmitter and recipient pairs. *Retrovirology* **13**, (2016).

31. Zhao, S., Zhu, W., Xue, S. & Han, D. Testicular defense systems: Immune privilege and innate immunity. *Cellular and Molecular Immunology* **11**, 428–437 (2014).

32. Fijak, M. & Meinhardt, A. The testis in immune privilege. *Immunological Reviews* **213**, 66–81 (2006).

33. Sabatté, J. *et al.* The role of semen in sexual transmission of HIV: Beyond a carrier for virus particles. *Microbes and Infection* **13**, 977–982 (2011).

34. Anderson, J. A. *et al.* HIV-1 populations in semen arise through multiple mechanisms. *PLoS Pathog.* **6**, 51–52 (2010).

35. Novitsky, V. *et al.* Transmission of Single and Multiple Viral Variants in Primary HIV-1 Subtype C Infection. *PLoS One* **6**, e16714 (2011).

36. Salazar-Gonzalez, J. F. *et al.* Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* **206**, 1273–1289 (2009).

37. Abrahams, M.-R. *et al.* Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants. *J. Virol.* **83**, 3556–3567 (2009).

38. Li, H. *et al.* High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathog.* **6**, 1–17 (2010).

39. Bar, K. J. *et al.* Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J. Virol.* **84**, 6241–7 (2010).

40. Masharsky, A. E. *et al.* A Substantial Transmission Bottleneck among Newly and Recently HIV-1–Infected Injection Drug Users in St Petersburg, Russia. *J. Infect. Dis.* **201**, 1697–1702 (2010).

41. Russell, E. S. *et al.* The Genetic Bottleneck in Vertical Transmission of Subtype C HIV-1 Is Not Driven by Selection of Especially Neutralization-Resistant Virus from the Maternal Viral Population. *J. Virol.* **85**, 8253–8262 (2011).

42. Miller, R. L. *et al.* HIV Diversity and Genetic Compartmentalization in Blood and Testes during Suppressive Antiretroviral Therapy. *J. Virol.* **93**, (2019).

43. Mabvakure, B. M. *et al.* Evidence for both Intermittent and Persistent Compartmentalization of HIV-1 in the Female Genital Tract Downloaded from. *jvi.asm.org 1 J. Virol.* **93**, 311–330 (2019).

44. Deleage, C. *et al.* Defining early SIV replication and dissemination dynamics following vaginal transmission. *Sci. Adv.* **5**, eaav7116 (2019).

45. Berthier-Vergnes, O. *et al.* TNF-α enhances phenotypic and functional maturation of human epidermal Langerhans cells and induces IL-12 p40 and IP-10/CXCL-10 production. *FEBS Lett.* **579**, 3660–3668 (2005).

46. De Witte, L. *et al.* Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nat. Med.* **13**, 367–371 (2007).

47. Shaw, G. M. & Hunter, E. HIV transmission. *Cold Spring Harb. Perspect. Med.* **2**, (2012).

48. Dunham, R. *et al.* The AIDS resistance of naturally SIV-infected sooty mangabeys is independent of cellular immunity to the virus. *Blood* **108**, 209–17 (2006).

49. Kornfeld, C. *et al.* Antiinflammatory profiles during primary SIV infection in African green monkeys are associated with protection against AIDS. *J. Clin. Invest.* **115**, 1082–1091 (2005).

50. King, D. F. L. *et al.* Mucosal tissue tropism and dissemination of HIV-1 subtype B acute envelope-expressing chimeric virus. *J. Virol.* **87**, 890–9 (2013).

51. Troyer, R. M. *et al.* Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J. Virol.* **79**, 9006–18 (2005).

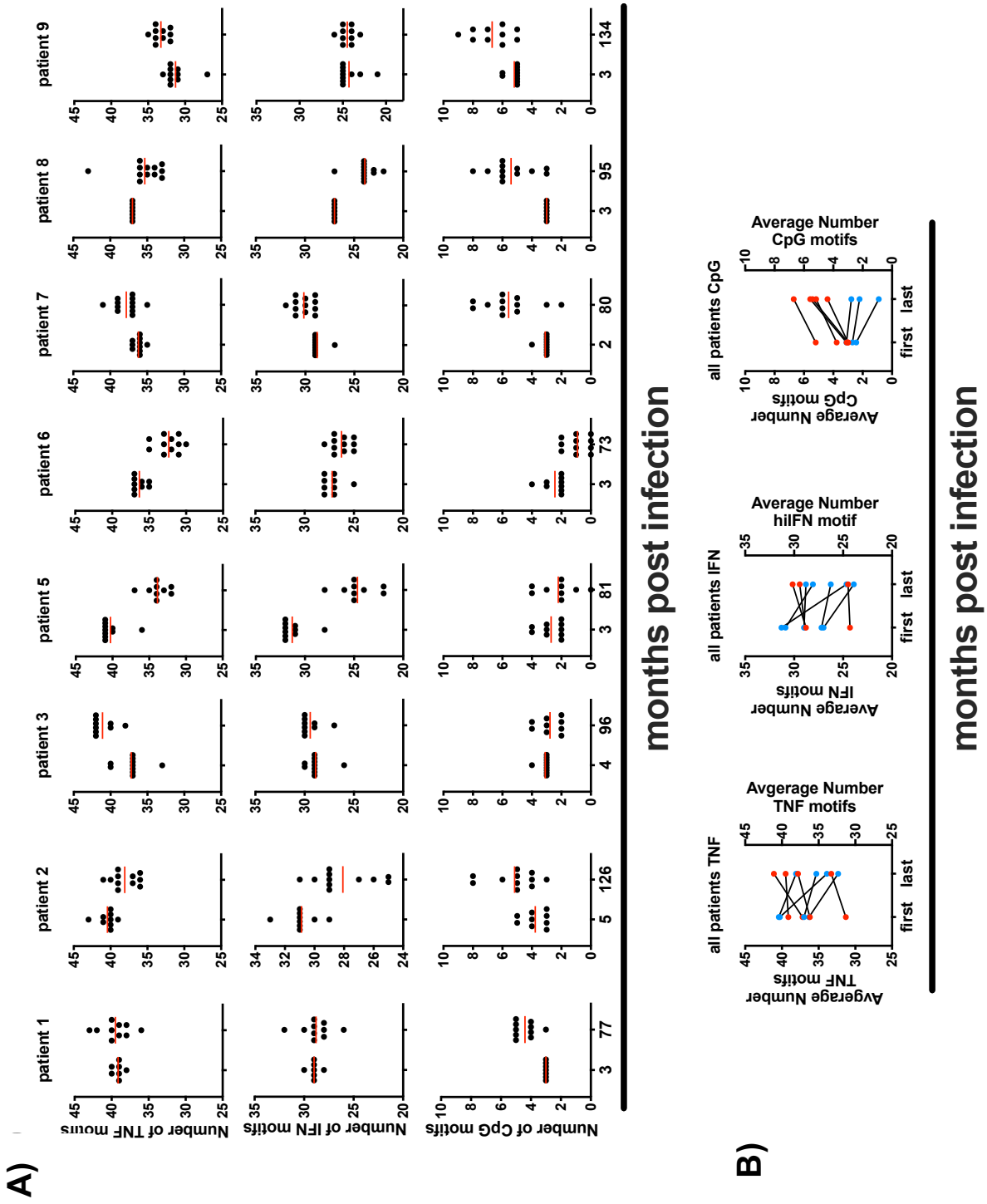
52. Moore, P. L. *et al.* Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nat. Med.* **18**, 1688–1692 (2012).

53. Freund, N. T. *et al.* Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. *Sci. Transl. Med.* **9**, (2017).

54. Geller, R. *et al.* The external domains of the HIV-1 envelope are a mutational cold spot. *Nat. Commun.* **6**, 1–9 (2015).

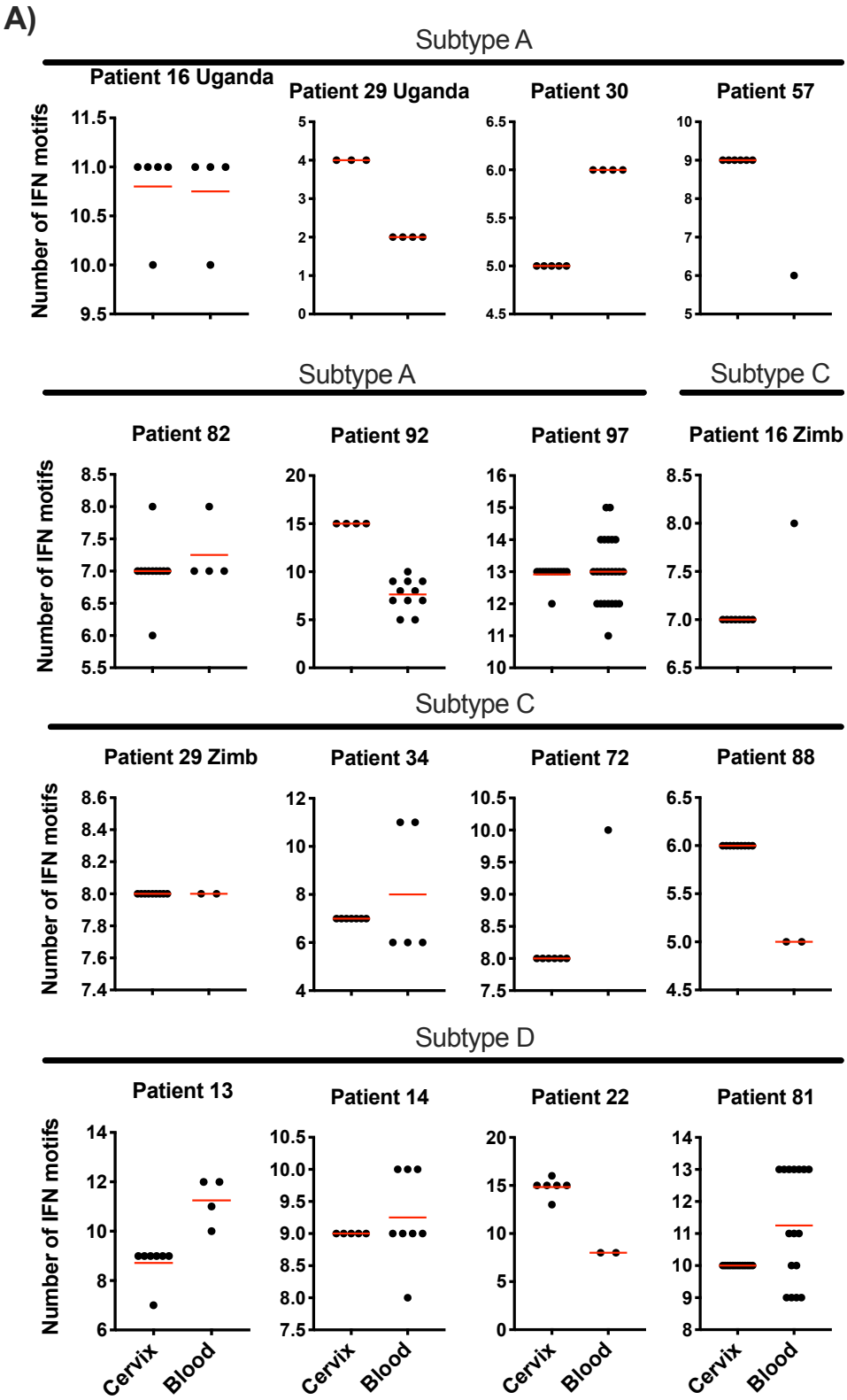
55. Moore, D. M., Arts, E. J., Gao, Y. & Marozsan, A. J. A yeast recombination-based cloning system to produce chimeric HIV-1 viruses and express HIV-1 genes. *Methods Mol. Biol.* **304**, 369–85 (2005).

Supplementary Figures

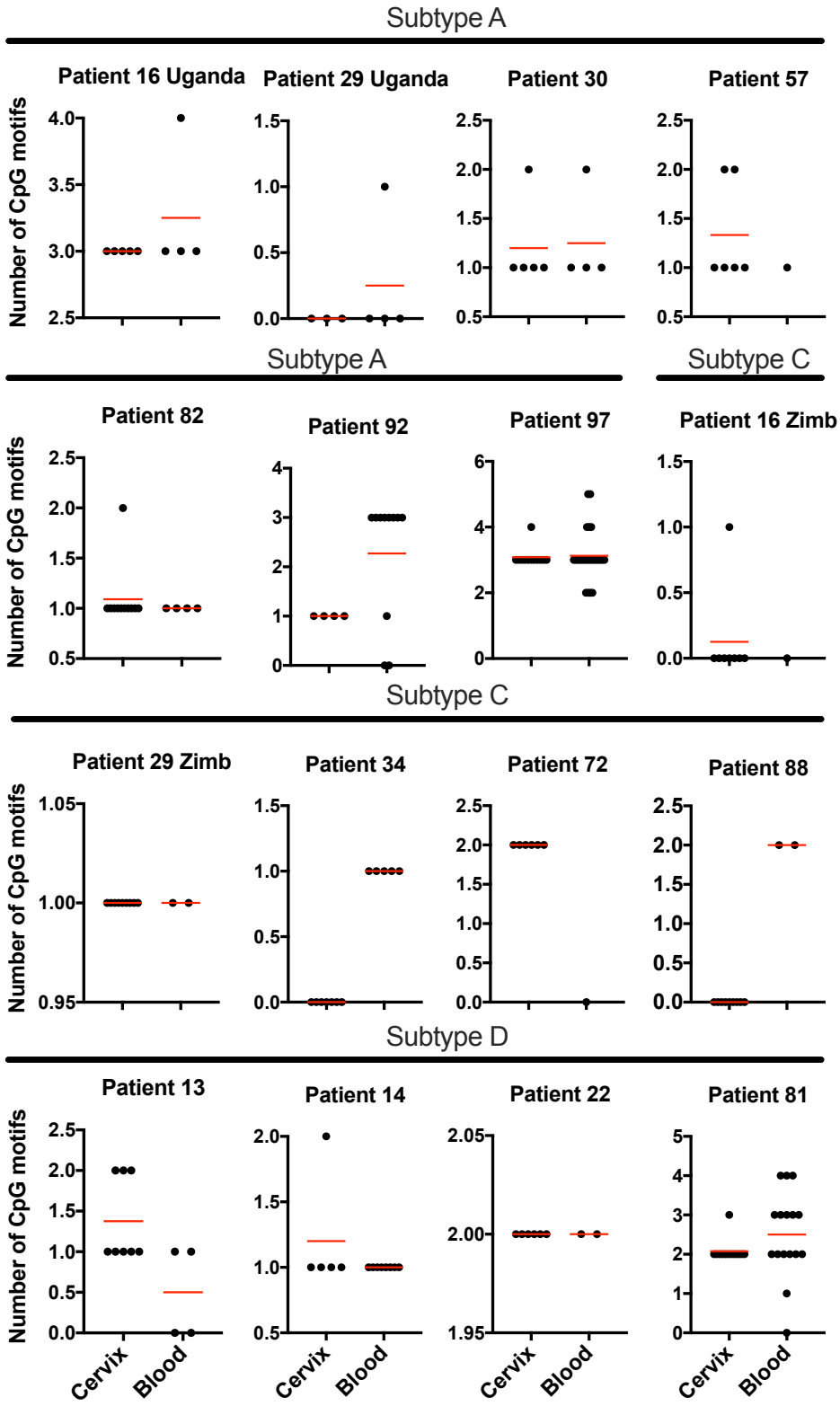


Supplementary Figure 2-1: HIV *env* sequences from the serum of 8 men overtime demonstrate no apparent trend with the number of TNF-α, IFN-α or CpG inducing motifs.

Sequences were attained from Shankarappa *et al.* (1999). Viral *env* sequences were isolated from the serum of 8 men at various time points through the individual's infection. This cohort contains *env* sequences from the C2 to V5 region (822 bp in length). Sequences were aligned to HXB2 and trimmed to ensure full C2-V5 coverage. (**A**) Viral sequences from the first time point of isolate and the last time point of isolate from the serum compartment were analyzed for 8 men. Graphs display the number of TNF-α, IFN-α and CpG motifs of all viral sequences isolated from patients from the first time point of isolation and last time point of isolation. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. Each dot represents a unique viral sequence. (**B**) Graphs display the average number of TNF-α, IFN-α and CpG motifs for all patients based on viral sequences isolated from the first and last time point of infection. Red dots from matched serum samples within each patient signifies increases in the number of immunostimulatory motifs while blue dots between matched serum samples within each patient signifies a decrease in the number of immunostimulatory motifs.

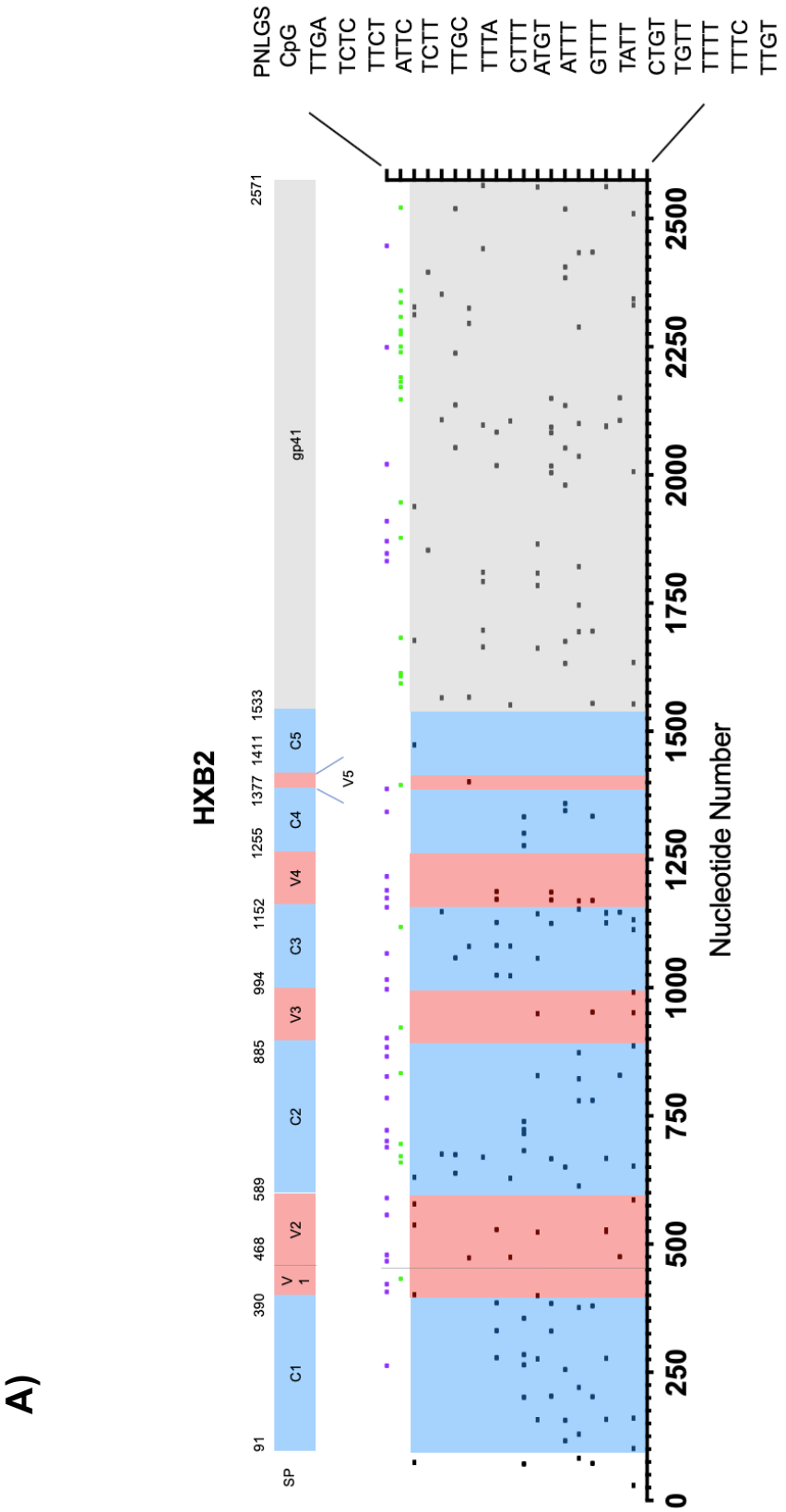

A)
Subtype A
Patient 16 Uganda
Patient 29 Uganda
Patient 30
Patient 57
Number of IFN motifs
Subtype A
Subtype C
Patient 82
Patient 92
Patient 97
Patient 16 Zimb
Number of IFN motifs
Subtype C
Patient 29 Zimb
Patient 34
Patient 72
Patient 88
Number of IFN motifs
Subtype D
Patient 13
Patient 14
Patient 22
Patient 81
Number of IFN motifs
Cervix
Blood

B)

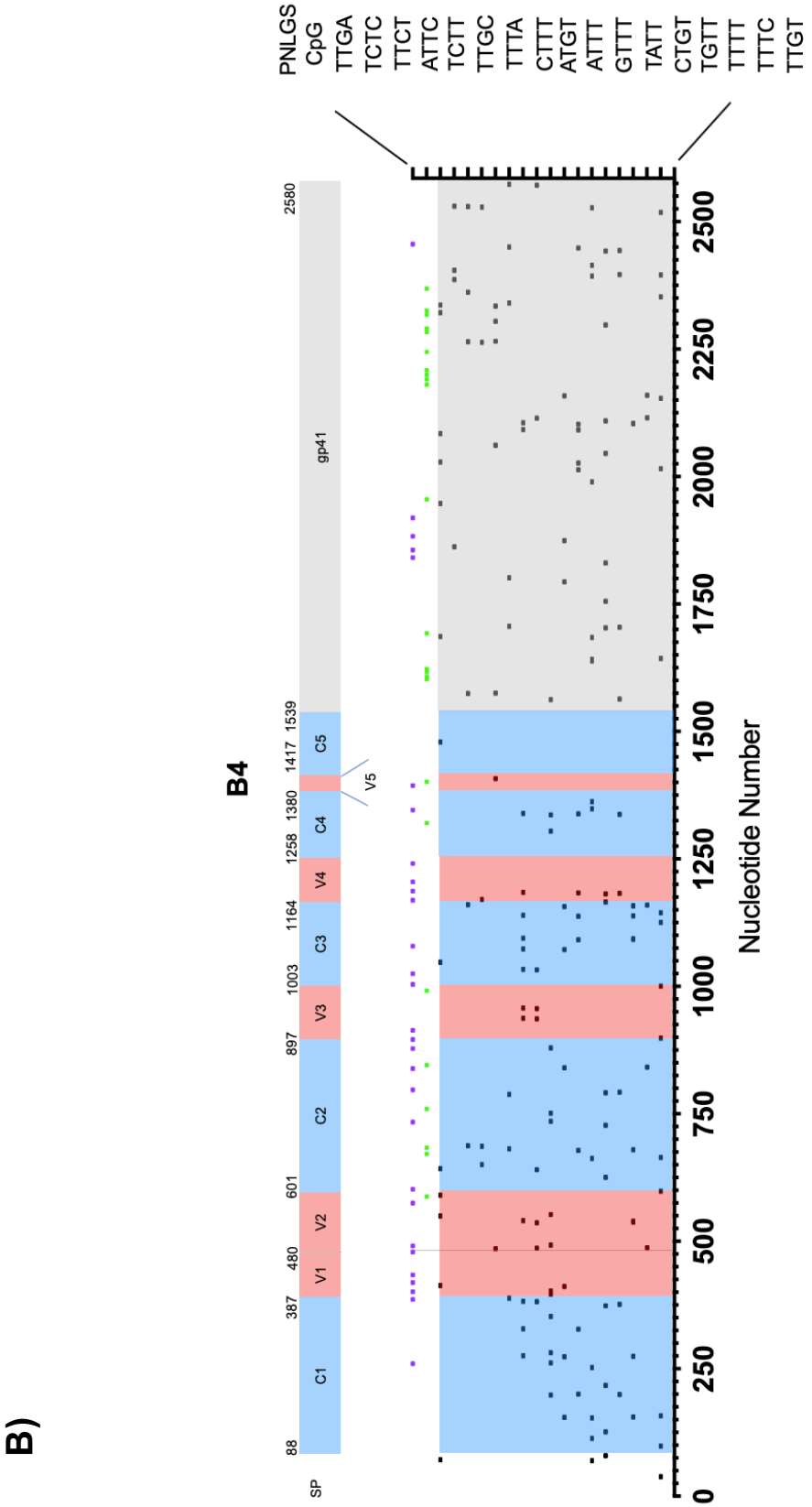

Subtype A
Patient 16 Uganda
Patient 29 Uganda
Patient 30
Patient 57
Number of CpG motifs
Subtype A
Subtype C
Patient 82
Patient 92
Patient 97
Patient 16 Zimb
Number of CpG motifs
Subtype C
Patient 29 Zimb
Patient 34
Patient 72
Patient 88
Number of CpG motifs
Subtype D
Patient 13
Patient 14
Patient 22
Patient 81
Number of CpG motifs
Cervix
Blood
Cervix
Blood
Cervix
Blood
Cervix
Blood

Supplementary Figure 2-2: The serological compartment demonstrates no apparent trend with the number of IFN-α inducing or CpG inducing motifs.

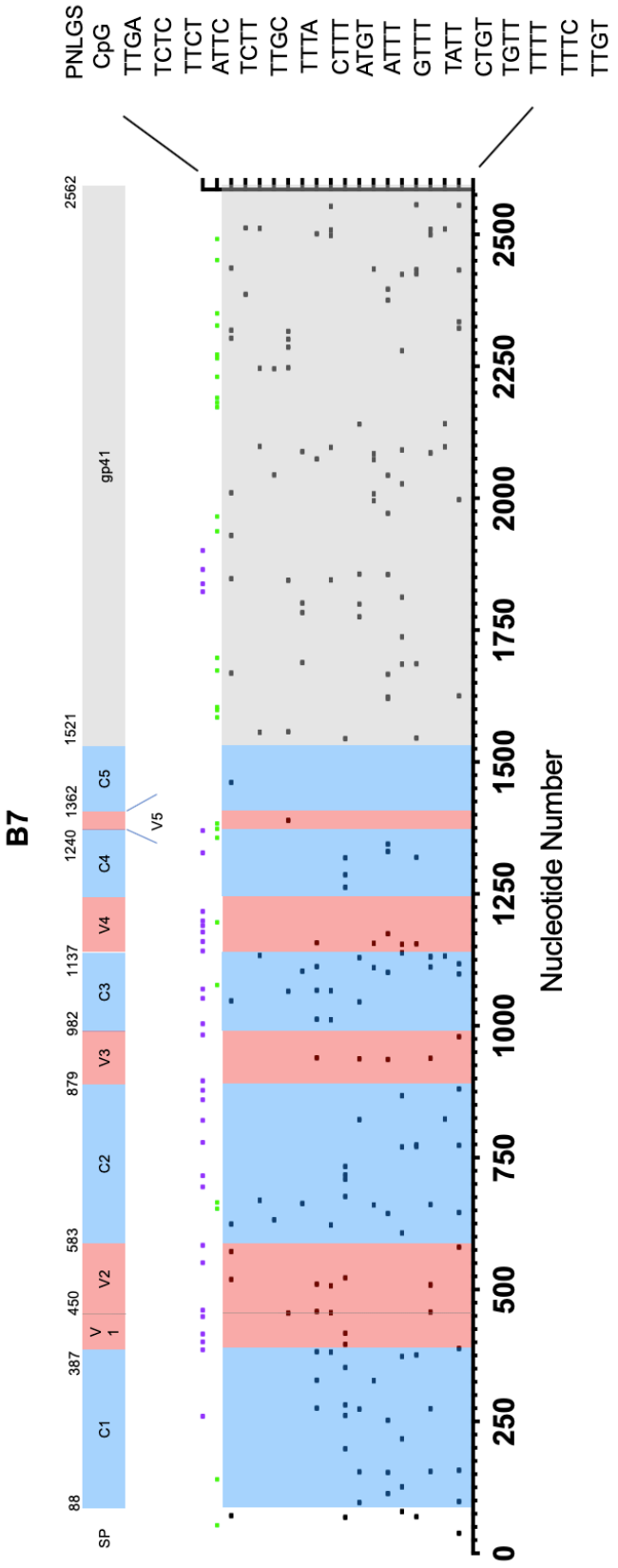
Sequences were attained from Klein *et al*. (2018). This cohort contains *env* sequences from the C2 to V3 region (405 bp in length). Sequences were aligned to HXB2 and trimmed so the region analyzed was the same length. Each dot represents a unique viral sequence. **(A)** Patients were stratified according to their subtype infection and then analyzed for the number of TNF-α motifs. Patients were infected with either subtype A, C or D. There were no AD recombinants. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences. **(B)** Patients were stratified according to their subtype infection and then analyzed for the number of CpG motifs. Red line represents the geometric mean number of immunostimulatory motifs in the HIV sequences.


A)
HXB2
SP
C1
V1
V2
C2
V3
C3
V4
C4
V5
C5
gp41
91
390
468
589
885
994
1152
1255
1377
1411
1533
2571
PNLGS
CpG
TTGA
TCTC
TTCT
ATTC
TCTT
TTGC
TTTA
CTTT
ATGT
ATTT
GTTT
TATT
CTGT
TGTT
TTTT
TTTC
TTGT
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Nucleotide Number

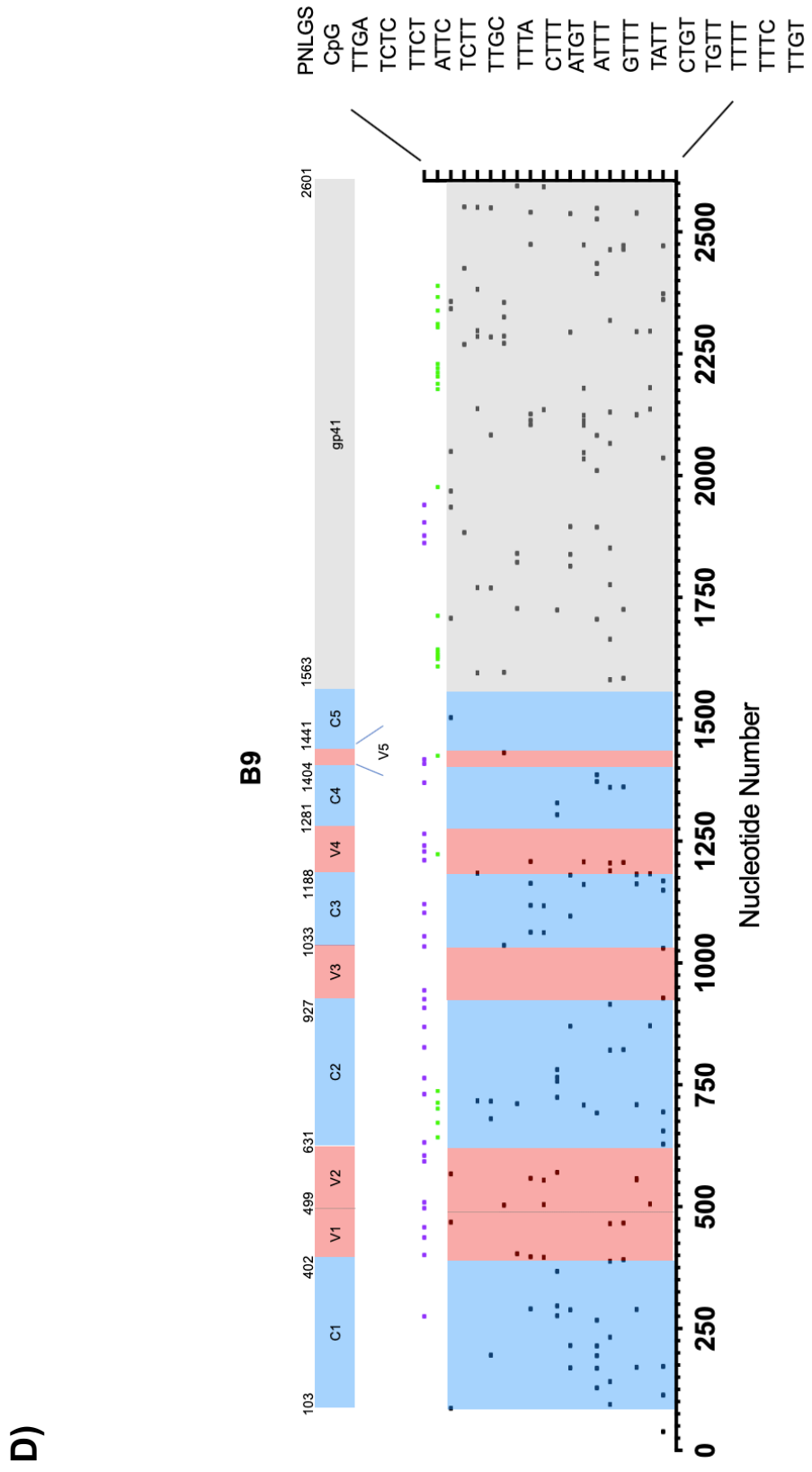
B)

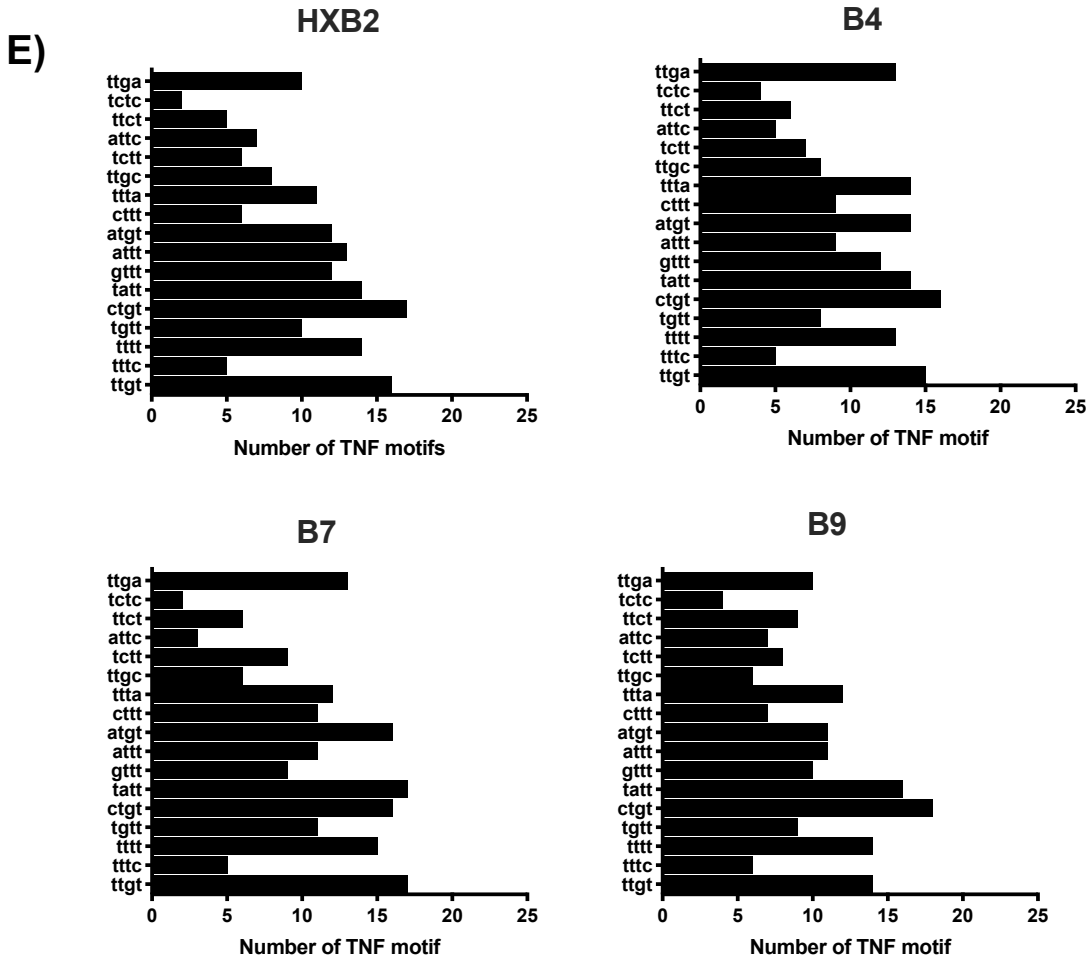

B4
PNLGS
CpG
TTGA
TCTC
TTCT
ATTC
TCTT
TTGC
TTTA
CTTT
ATGT
ATTT
GTTT
TATT
CTGT
TGTT
TTTT
TTTC
TTGT
SP
C1
V1
V2
C2
V3
C3
V4
C4
V5
C5
gp41
88
387
480
601
897
1003
1164
1258
1380
1417
1539
2580
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Nucleotide Number

C)


B7
PNLGS
CpG
TTGA
TCTC
TTCT
ATTC
TCTT
TTGC
TTTA
CTTT
ATGT
ATTT
GTTT
TATT
CTGT
TGTT
TTTT
TTTC
TTGT
SP
C1
V1
V2
C2
V3
C3
V4
C4
V5
C5
gp41
88
387
450
583
879
982
1137
1240
1362
1521
2562
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Nucleotide Number

D)


B9
PNLGS
CpG
TTGA
TCTC
TTCT
ATTC
TCTT
TTGC
TTTA
CTTT
ATGT
ATTT
GTTT
TATT
CTGT
TGTT
TTTT
TTTC
TTGT
gp41
C1
V1
V2
C2
V3
C3
V4
C4
V5
C5
103
402
499
631
927
1033
1188
1281
1404
1441
1563
2601
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Nucleotide Number



Supplementary Figure 2-3: There is no overlap between the various TNF-α stimulatory motifs and CpG motifs and PNLGs.

The number of TNF-α inducing motifs for the sequences **(A)** HXB2 **(B)** B4 **(C)** B7 **(D)** B9 **(E)** are shown. Full *env* sequences for the three predominant acute-early viruses were mapped for the various immunostimulatory motifs and compared to HXB2. The blue shaded areas represent the constant regions, while the pink shaded areas represent the variable regions on *env*. The grey shaded area represents the gp41 portion of Env. The x-axis represents the position on the Env gene using HXB2 numbering. TNF-α motifs labelled on the y-axis were assigned a number from 1 to 17 and mapped across the *env* sequences. The **black dots** represent the location of different TNF-α inducing motifs; the

green dots represent the location of CpG islands; the purple dots represent the location of PNLGs using the N-Glycosite predictor from the Los Alamos database.

Chapter 3 « Objective two preface »

To date, all HIV-1 vaccines that have advanced to human clinical trials have failed to demonstrate any level of protective efficacy[1–7]. Although the RV144 "Thai Trial" had been described to convey a modest 31.2% protection[3,8], a recently conducted follow-up study, HVTN 702, failed to repeat the previously reported success of RV144 and was stopped due to futility[9–11]. One of the major impediments in the development of an efficacious HIV-1 vaccine is the inability of current vaccine regimens to invoke protective neutralizing antibody responses[10,12]. Ultimately, this is due to a failure in both vaccine design and utilization of the correct vaccine antigens[12]. Over the last few years, we have been using novel *in silico* methodologies for identifying the next generation of vaccine antigens (See Chapter 2) and using molecular engineering to expose sites on antigens that are critical for neutralizing antibody development. The latter point being the major objective for the remainder of this thesis.

Our research team has been pioneering novel vaccine technologies that can be used as immunotherapies for HIV-1 cure[13]. Through that research, we have identified virus-like particles (VLPs) as promising vaccine candidates that can also be exploited for prophylactic HIV-1 vaccine development. The VLPs we generate in the lab morphologically resemble wild-type HIV and are believed to express Env in a structurally accurate conformation, thereby more accurately representing wild-type proteins[13]. Therefore, due to their particulate nature and ability to present Env antigens in natural conformations, we believe VLPs should serve as an ideal vaccine delivery mechanism to present native and functional Env antigens to immune cells. As, VLPs are non-infectious, genome-less, multiprotein structures, it also makes them intriguing B cell vaccine candidates as they should be capable of crosslinking BCRs. As N-linked glycans have previously been described to flank sites of vulnerability on HIV Env antigens, thereby masking these sites from the immune response[14–16], we believe that removing the putative N-linked glycans (PNLGs) surrounding the CD4 binding-site (bs) on the VLPs, may be able to stimulate production of CD4bs-directed broadly neutralizing antibodies (bnAb) expressing B cell lineages. We ultimately aim to use viral Env antigens identified as Transmitted/founder (T/F) viruses (See Chapter 2) as part of a sequential VLP-based

vaccination regimen, ending in the use of chronic Env antigens displayed on the surface of VLPs. Using VLPs expressing Env from different time points in infection, we are creating a vaccination protocol that imitates the time course of a natural infection. The latter enabling approximately 10-30% of HIV infected individuals to generate broadly neutralizing antibodies against diverse strains of HIV-1[17,18]. Using this sequential vaccination protocol, we hypothesize we may be able to stimulate the production of bnAb lineages by selecting and expanding bnAb germline precursor B cells and directing them to become bnAb producers via the necessary somatic hypermutation (SMH).

In the second half of this MSc thesis, we use the lab adapted strain NL4-3 as a model to identify which specific glycan mutations surrounding the CD4bs that will enhance viral neutralization. We noticed that by removing a single glycan and presenting the modified Env on the surface of an infectious virion, there was enhanced neutralization by CD4bs-directed Abs. Next, as a proof-of-principle, we introduced the most effective single glycan mutation into various putative T/F viral clones provided by the AIDS Reagent Program and evaluated the neutralization efficacy mediated by CD4-bs-directed Abs. Lastly, we devise a sequential vaccination protocol on female BALB/c mouse and VRC01 gH knock-in mice. We will carry out our sequential vaccination protocol with first, second and third vaccinations utilizing T/F, acute and chronic HIV-1 VLPs. Here we are comparing the wild-type (WT) of each VLP formulation with the glycan mutant formulation to assess the impact in CD4bs-directed Ab development.

References

1. Pitisuttithum, P. *et al.* Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J. Infect. Dis.* **194**, 1661–71 (2006).
2. Flynn, N. M. *et al.* Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* **191**, 654–65 (2005).
3. Edlefsen, P. T., Gilbert, P. B. & Rolland, M. Sieve analysis in HIV-1 vaccine efficacy trials. *Curr. Opin. HIV AIDS* **8**, 432–436 (2013).
4. Volk, J. E. *et al.* The HVTN503/Phambili HIV vaccine trial: a comparison of younger and older participants. *Int. J. STD AIDS* **25**, 332–40 (2014).
5. Emu, B. *et al.* HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. *J. Virol.* **82**, 5398–407 (2008).
6. Altfeld, M. *et al.* HLA Alleles Associated with Delayed Progression to AIDS

Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1. *PLoS Med.* **3**, e403 (2006).
7. Harrer, T. *et al.* Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection. *AIDS Res. Hum. Retroviruses* **12**, 585–92 (1996).
8. Kim, J. H., Excler, J.-L. & Michael, N. L. Lessons from the RV144 Thai phase III HIV-1 vaccine trial and the search for correlates of protection. *Annu. Rev. Med.* **66**, 423–37 (2015).
9. Large-Scale HIV Vaccine Trial to Launch in South Africa | NIH: National Institute of Allergy and Infectious Diseases. Available at: https://www.niaid.nih.gov/news-events/large-scale-hiv-vaccine-trial-launch-south-africa. (Accessed: 9th March 2020)
10. Robinson, H. L. HIV/AIDS Vaccines: 2018. *Clinical Pharmacology and Therapeutics* **104**, 1062–1073 (2018).
11. Experimental HIV Vaccine Regimen Ineffective in Preventing HIV | NIH: National Institute of Allergy and Infectious Diseases. Available at: https://www.niaid.nih.gov/news-events/experimental-hiv-vaccine-regimen-ineffective-preventing-hiv. (Accessed: 9th March 2020)
12. Gao, Y., McKay, P. F. & Mann, J. F. S. Advances in HIV-1 vaccine development. *Viruses* **10**, (2018).
13. Pankrac, J. *et al.* A heterogeneous human immunodeficiency virus-like particle (VLP) formulation produced by a novel vector system. *NPJ vaccines* **3**, 2 (2018).
14. Pritchard, L. K. *et al.* Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6**, 7479 (2015).
15. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10**, 14 (2013).
16. Montefiori, D. C., Robinson, W. E. & Mitchell, W. M. Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9248–52 (1988).
17. McCurley, N. P. *et al.* HIV transmitted/founder vaccines elicit autologous tier 2 neutralizing antibodies for the CD4 binding site. *PLoS One* **12**, e0177863 (2017).
18. Sagar, M., Wu, X., Lee, S. & Overbaugh, J. Human Immunodeficiency Virus Type 1 V1-V2 Envelope Loop Sequences Expand and Add Glycosylation Sites over the Course of Infection, and These Modifications Affect Antibody Neutralization Sensitivity. *J. Virol.* **80**, 9586–9598 (2006).

« Engineering HIV Virus-Like-Particles for Improved Env Antigenicity to CD4 binding-site Directed Neutralizing Antibodies »

Authors: Rahul Pawa[1], Seth Kibel[1], Renata Ceccacci[1], Katja Klein[1], Meijuan Tian[1], Eric J. Arts[1] and Jamie F. S. Mann[1,2].

Affiliations:

[1]Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, Canada N6A 5C1.

[2] School of Veterinary Sciences, University of Bristol, Bristol, United Kingdom.

*Correspondence to: J.F.S Mann, University of Bristol.

Manuscript in preparation

Abstract

Crucial for infectivity, HIV-1 Env is the only viral protein on the surface that is targeted by the immune response. The arms race between the virus and the immune response concomitantly selects for viral escape mutants and evolves the adaptive immune response to develop non-neutralizing, neutralizing and broadly neutralizing antibodies (bnAbs). The accumulation of glycans peripheral to immunogenic epitopes, such as the CD4 binding site (bs), make it difficult for the immune response to recognize antibody target sites and result in only a subset of individuals capable of developing bnAbs. Previous HIV-1 vaccines have failed to elicit a protective neutralizing antibody response with sufficient breadth, due to structural mismatches between vaccine antigens and the Env on infecting wildtype virus. Furthermore, current vaccine strategies largely fail to effectively tackle the genetic diversity that ironically selects for specific immune B cell lineages. As such, in this study we proposed to remove particular N-linked glycans peripheral to the CD4bs and follow a sequential vaccination protocol using virus-like-particles (VLP) from different time point of infection to select for CD4bs-directed antibodies. Using NL4-3, we present a proof-of-concept that removing a single glycan can enhance neutralization potential by anti-CD4bs antibodies. In particular, mutant N301A presented no impairment to Env structure and was the most susceptible to CD4bs-directed antibody neutralization. Overall, this study identified a target glycan mutation to introduce into Envs, to be displayed by VLPs, for use in a sequential vaccination protocol. By presenting antigens from different time points in their near native conformation, we may be able to induce production of protective antibodies. With the overwhelming burden of the AIDS pandemic, the best solution is an effective vaccine.

Introduction

The Human Immunodeficiency Virus (HIV) viral surface envelope glycoprotein (Env) is a heterotrimer of gp120 and gp41 protomers that are crucial for infectivity and immune sensing[1,2]. The Env glycoprotein is heavily glycosylated, with potential N-linked glycans (PNLGs) accounting for as much as 50% of the molecular weight of the protein[2]. These glycans are added as post-translational modifications mediated by the endoplasmic reticulum and Golgi networks from the host[3]. PNLGs are considered 'potential' as Env sequence variation plays an important role in the modification of glycosylation patterns by adding or removing Asparagine (Asn) amino acids. As a result, individual gp120 subunits have on average 18-33 PNGS which vary between hosts, between viral isolates, within hosts, and even across time within hosts[4–6]. Through mass spectrometry studies, ~4 have been found to be located in the inner domain, ~7-8 in the V1/V2 and V3 loops and the rest in the outer domain of gp120[7–9]. The error prone reverse transcriptase and the selective pressure from the immune response diversifies the viral RNA. A common shift in PNGS due to the ensuing immune response is the selection of viral variants with a N332 to N334 glycan shift as this provides resistance to V3-glycan antibodies[10,11]. Together, all these N-linked glycans restrict antibody access to 97% of the Env surface[12]. This poses a challenge for the generation of a vaccine against HIV-1, as the glycan shield[13] represents a major barrier to the development of protective immune responses by shielding vulnerable immunogenic sites on the viral Env[14–16].

During sexual transmission, it has been well accepted that in 70-80% of infections, a single viral variant is able to overcome a stringent genetic bottleneck by successfully evading mucosal secretions, overcoming the mucosal barrier, and circumventing immune mediated clearance to establish a disseminating systemic infection[17–21]. A continuously evolving arms race between the virus and the host immune system, drives viral diversification with neutralizing antibodies (bnAbs) arising in ~20-50% of HIV-1+ individuals[5,22–25]. After years of immune selection and viral escape, these nAbs become highly mutated, especially in the framework region, and they tend to have abnormally long and charged immunoglobulin heavy complementary-determining regions (CDR; region that binds to the antigen)[26,27]. This small subset of antibodies can neutralize various primary isolates because

they bind to conserved epitopes on Env[28] —sites that are functionally important for the virus and are referred to as broadly neutralizing ntibodies (bNabs). The HIV-1 trimer complex has five such sites that are vulnerable to bnAb recognition: the CD4 binding site (CD4bs) of gp120, PG9 and PG16 recognition site in gp120, V3 loop, the membrane proximal external region (MPER) in gp41 and the gp120/gp41 interface region[29]. While the presence of these bnAbs in HIV-1-infected patients does not decrease the viral burden, they are associated with longer duration infection and more effective CD4+ T cell help[30–33]. As the diverging viruses balance escape mutations with deleterious mutations through glycan modification of the Env protein, the various HIV-specific bnAbs-specific B cells evolve with the variants to effectively neutralize the virus. Constant evolution of Ab lineages can force strong divergence from their germline ancestors. As a result, these B cell ancestors to bnAb producing cells usually fail to interact with concealed and conserved epitopes of virus that has undergone years of evolution[34,35]. Nonetheless, how these bnAbs are generated or why they are only generated in less than 50% of HIV-1-infected patients are questions that remain unanswered.

The CD4 binding site (bs) is an example of a region that cannot be heavily modified by the virus as it will directly impact the ability of the virus to engage with its receptor. As a result, the virus mutates regions around the CD4bs to circumvent bnAbs that target this site for virus neutralization. Zhou *et al*. and Liang *et al*. have identified PNLGs surrounding the CD4bs that upon removal, significantly increase VRC01 bnAb binding. These sites include: N197, N276, N301, N363, N386 and N486[36,37]. Upon removal of the putative N-linked glycosylation sites, the binding affinity and neutralization breadth of VRC01 dramatically increased. Liang *et al*. were able to demonstrate that removal of the putative N-linked glycan-binding site 197, had little effect on the overall structure of the Env trimer[37]. More importantly, they revealed a steric clash to the CD4bs in the presence of the glycan which destabilizes certain Env-antibody complexes –such as Env-VRC01[37]. Collectively, this suggests future Env immunogens lacking certain PNLGs may be used to focus the immune response onto critically sensitive structural epitopes to increase the chance of eliciting a bnAb response.

Prior preclinical and clinical studies used a variety of recombinant monomeric and soluble oligomeric HIV Env immunogens that have lacked conformational similarity to native trimers on wild type infectious virus. These immunogens failed to elicit bnAb production or engage nAb germline precursor B cells[38] and thus allow for proper B cell ontogeny[15,38]. Lessons from previous Env iterations encouraged development of BG606 SOSIP.664 –a soluble HIV-1 Env trimer with a more native tertiary and quaternary structure[39]. While this recombinant protein achieves a high degree of structural integrity and conformational similarity to the native trimer, it fails to target the diversity of the wild-type Env protein and thus stimulate a protective bnAb response.

To address the deficiencies of previous Env immunogens and vaccine strategies at eliciting neutralizing antibody responses, our lab has pioneered methodologies for virus-like-particle (VLP) vaccine production[40]. The VLPs produced in our lab morphologically resemble wild-type HIV and are believed to express Env in a structurally accurate conformation, thereby more accurately representing wild-type proteins[40]. Therefore, the VLPs may serve as an ideal vaccine delivery mechanism and present native and functional Env spikes found on wild-type infectious virus to immune cells. By removing the PNLGs surrounding the CD4bs on the VLPs, we hypothesize that we may stimulate production of CD4bs-directed bnAb producing B cell lineages. By using VLPs from different time points in infection, we can implement a sequential vaccination protocol that imitates the time course of a natural infection and expose the immune response to ancestral strains of HIV-1 through to immune system-evolved HIV-1 seen in chronic stage of infection[4,41]. We believe, this may stimulate the production of bnAb lineages by selecting and expanding bnAb germline precursor B cells and allow them to attain the necessary somatic hyper mutations (SMH) and ideal length of complementary-determining region (CDR) necessary to exhibit bNAb functioning.

In this study, we use the lab adapted strain NL4-3 as a model to verify and select for a specific glycan mutation surrounding the CD4-bs that will enhance neutralization. This model served as proof-of-concept and showed that by removing a single glycan and presenting the modified Env on the surface of virions, we can enhance neutralization by CD4bs-directed antibodies. Furthermore, we introduced a single glycan mutation into

various putative T/F viral clones provided by the AIDS Reagent Program and evaluated changes in CD4-bs-drected neutralizing Abs. Lastly, we devised a sequential vaccination protocol on female BALB/c mouse and VRC01 gH knock-in mice. Here we carried out our sequential vaccination protocol with first, second, and third vaccinations utilizing T/F, acute and chronic Envs displayed on our VLP vaccine system.

Results

Molecular engineering of viral particles (VPs) created successful glycan mutants

Using our already established yeast-based recombination protocol, we mutated previously identified PNLGs flanking the CD4bs within *env.* It is important to note that the lab adapted strain NL4-3 was used to test the mutagenesis approach due to the clonal nature and the presence of the target PNLGs. Using the wild-type (WT) NL4-3 *env* gene, we converted PNLG 197 from an asparagine (N) to a glutamine (Q) while PNLGs 276, 301, 363 and 386 was changed from an N to an alanine (A). Liang *et al.* were able to demonstrate that removal of the putative N-linked glycan-binding site 197, and introduction of Q had little effect on the overall structure of the Env trimer and in turn enhanced VRC01 binding[37]. We also included the N425K mutation into the C4 region of gp120, as it had previously been reported that this mutation enhanced the sensitivity of viruses to antibody responses[44]. The N425K phenotype is found in <1% of all circulating HIV-1 strains, suggesting that despite its ability to enhance viral infectivity, it is an unfavorable mutation for the virus evolutionarily, possibly due to it making the virus more susceptible to neutralizing antibody mediated attack [44]. As such, we decided to explore the effect of this mutation in addition to the aforementioned PNLGs. Following PCR mutagenesis and yeast-based cloning, the pREC mutant constructs were sent for Sanger sequencing to verify that the site-directed mutagenesis cloning strategy was successful in the introduction of the desired mutation **(Figure 3-1A, B)**. The sequencing results indicated that all desired mutations were introduced in N197Q, N301A, N363A, N386A, N463A and N425K **(Supplementary Figure 3-1)**. On the other hand, the targeted mutagenesis of PNLG 276 resulted in an unexpected conversion from an N to serine (S), as opposed to the desired A. While we were unsure of the nature of this mutation to the overall Env trimer, we decided to explore

its effects as the resulting mutation should still prevent N-linked post-translational modification at this site.

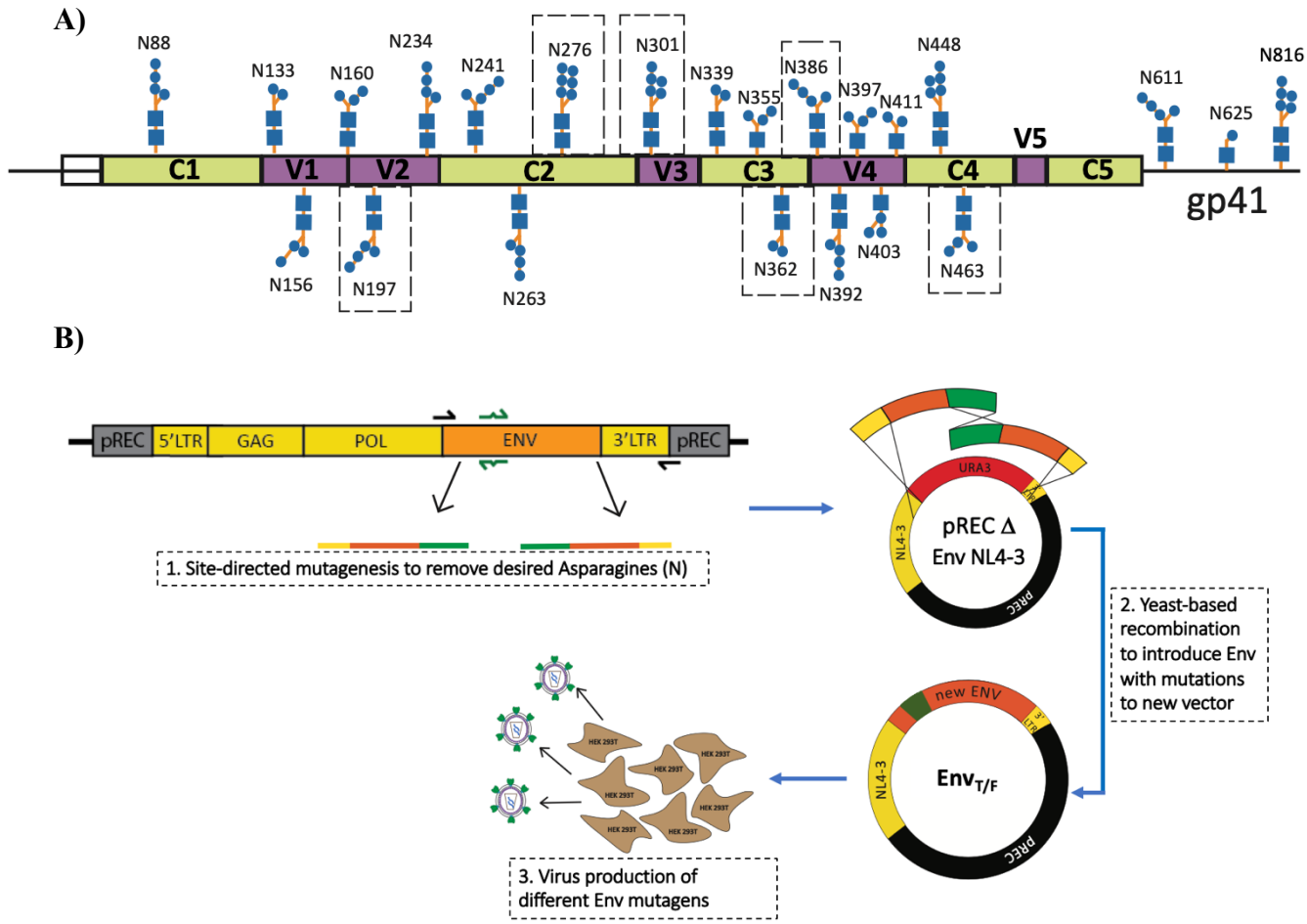
Glycan mutations did not impair overall virus production and structural integrity of Env

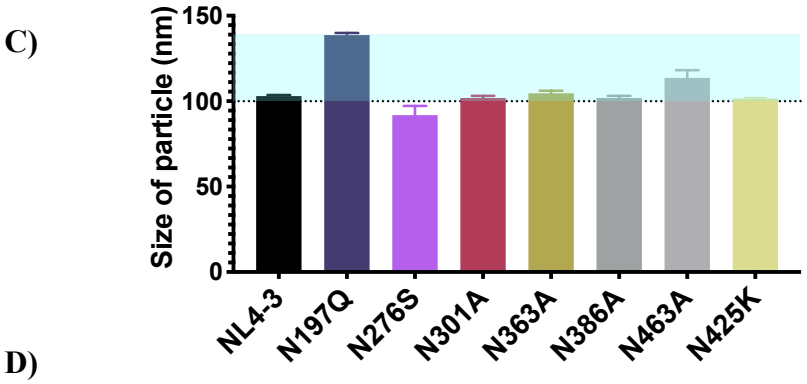
To test if the induced mutations had a negative impact on Env functioning, we made virus displaying the various Env mutants. We hypothesized that if the mutations were deleterious to the Env conformation, we would see reductions in the ability of the virus to infect susceptible cell lines. To generate viral particles harboring the desired PNLG deletions, HEK293T cells were transfected with the plasmids encoding N197Q, N276S, N301A, N363A, N386A, N463A and N425K, along with the plasmid encoding the WT genome (i.e. NL4-3 without any mutations) **(Figure 3-1B)**. Following a 3-day transfection, supernatant was collected and subject to VP purification. A p24 ELISA was performed to test for the production of viral protein Gag p24. Overall, the transfections and purifications yielded high amounts of Gag p24, similar to that seen with wildtype NL4-3, indicating that the various Env mutations did not impair overall virion production **(Figure 3-1D)**. The minimal differences that were observed in quantification are likely due to differences in transfection efficiencies and overall virus purification between the different mutant constructs. To ensure particle production created virions of the correct size, we measured the size distribution profile of each of the PNLG mutants using dynamic light scattering (DLS). The sizes of all of the viral particles were within the accepted range of a mature HIV-1 virion, 100-120 nm **(Figure 3-1C)**. However, while not statistically significant, the size of the N276S mutant was 92nm which is slightly lower that the accepted size of 100 nm **(Figure 3-1C)**. Whether this size discrepancy is related to the achieved N276S mutation rather than the expected N276A mutation is currently unknown.

Many studies have focused on high-throughput alanine scanning, where amino acids are systemically substituted for alanine at selected positions by site-directed mutagenesis, for the effect of removing target PNLGs. These studies have shown enhanced antibody binding and no overall impact on Env structure[9,37,45,46]. In particular, N197Q and N301Q have been heavily studied for their additive effect of facilitating VRC01 binding to the CD4bs[37,47].

These studies have also demonstrated no effect to the overall Env trimer structure. Since we introduced an A in position 301, we wanted to confirm that there was no overall impact to the Env trimer. Additionally, we wanted to assess the impact of the unexpected N276S mutation on the overall Env trimer structure. We decided to use the N197Q in addition to the WT control VP as it has been demonstrated previously that the N197Q mutant does not affect overall Env trimer structure. The anti-p24 antibody binds and thereby quantifies the production of the structural protein Gag and as such provides no indication of Env production. Production of Env was first detected using the gp140 ELISA, an ELISA that relies on lectins that bind the glycosylated Env and subsequent detection of Env using the 5F3 antibody. The 5F3 antibody is a non-neutralizing antibody that binds the epitope QNQQEKNE, located outside the MPER region and therefore only recognizes dimeric and trimeric forms of cross-linked gp140. As such, the ELISA is more likely to quantify surface Envs that are likely to resemble trimers. The gp140 ELISA demonstrated that the WT, N197Q mutant and N301A mutant VPs expressed oligomeric Env, while the N276S mutant VP showed no oligomeric Env production **(Supplementary Figure 3-2)**. To further validate the results of the gp140 ELSIA, we performed a b13 western blot. This antibody binds to a linear epitope on the CD4bs within the gp120 Env glycoprotein. This assay was used to determine whether the N197Q and N301A VPs expressed the gp120 Env glycoprotein as well as whether the introduced PNLG mutation altered the structure of the CD4bs. The western blot was normalized using data collected previously from the p24 ELISA, such that each well received the same amount of VP preparation corresponding to 100ng of p24 protein of the VP preparation. Overall, the western blot resolved bands at ~120kDa (indicating the presence of gp120 monomers), ~160kDa (indicating gp160 heterodimers of gp120 and gp41) and ~240kDa Env glycoproteins (indicating oligomers of gp120 homodimers) as excepted for the WT, N197Q mutant and N301A mutant **(Supplementary Figure 3-3)**. Consistent with the gp140 ELISA, the N276S mutant did not resolve any bands. Interestingly, this also aligns with the slightly lower size observed through DLS **(Supplementary Figure 3-2, -3, Figure 3-1C)**. This may imply that the N → S mutation may have had a deleterious impact on Env stability or production, despite production of viral particles (demonstrated through p24 and DLS). It is thus possible that the N276S mutant gives rise to non-enveloped HIV gag particles in the absence of surface

expressed Env. Hence, the N276S mutant was eliminated from further analysis as a vaccine candidate.


A)
N88
N133
N160
N234
N241
N276
N301
N339
N355
N386
N397
N411
N448
N611
N625
N816
V5
C1
V1
V2
C2
V3
C3
V4
C4
C5
gp41
N156
N197
N263
N362
N392
N403
N463
B)
pREC
5'LTR
GAG
POL
ENV
3'LTR
pREC
1. Site-directed mutagenesis to remove desired Asparagines (N)
URA3
NL4-3
pREC Δ
Env NL4-3
pREC
2. Yeast-based recombination to introduce Env with mutations to new vector
new ENV
NL4-3
Env T/F
pREC
HEK 293T
3. Virus production of different Env mutagens


C)
Size of particle (nm)
150
100
50
0
NL4-3
N197Q
N276S
N301A
N363A
N386A
N463A
N425K

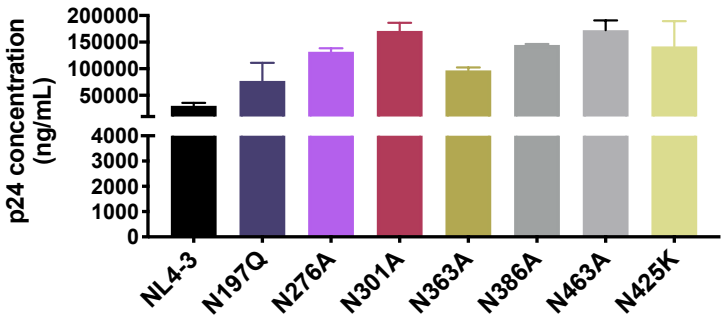

D)
p24 concentration (ng/mL)
200000
150000
100000
50000
4000
3000
2000
1000
0
NL4-3
N197Q
N276A
N301A
N363A
N386A
N463A
N425K

Figure 3-1: Single Glycan mutants do not negatively affect viral particle production.

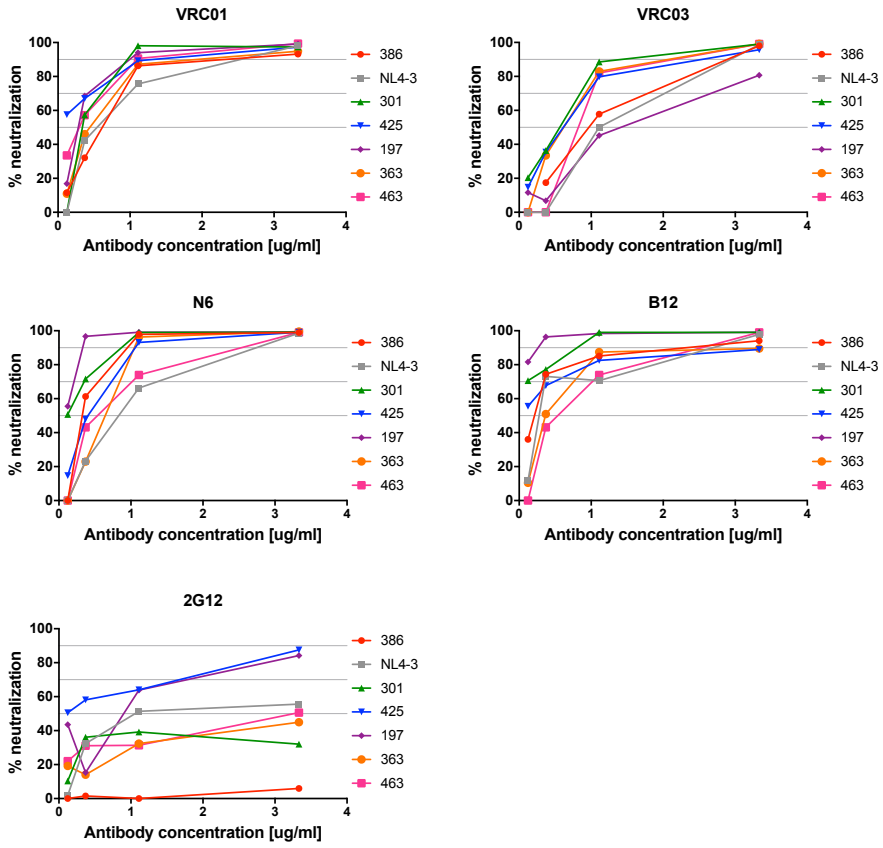
(A) The top schematic shows the constant regions and variable regions within *env.* The bottom schematic shows the locations of the most likely PNLGs found on NL4-3 *env.* The glycans surrounding by the dotted box are the glycans that sterically occlude the VRC01 bNAb from access to the CD4bs. Through our site-directed mutagenesis cloning strategy, we sought to mutate PNLGs outlined in red in order to increase VRC01 accessibility to the CD4bs. **(B)** Schematic of the site-directed mutagenesis cloning strategy. Mutagenic primers are drawn in green, whereas non-mutagenic primers are drawn in black. **(C)** Virion size demonstrated through DLS analysis. Briefly, a diluted VP preparation was loaded into a glass cuvette and analyzed using a DynaPro dynamic light scattering apparatus. Twenty independent read events were collected for each sample for analysis. **(D)** p24 ELISA of the various PNLG mutant constructs. Briefly HEK293T cells were transfected with the PNLG mutant plasmids using Xtremgene and VPs were purified as previously described. VPs were then quantified using a p24 ELISA that detects the production of the capsid protein Gag. Overall, all transfections resulted in sufficient production of VPs for use in future experiments. As the p24 ELISA was used to assess VP production, significance across different groups was not determined. Data represents mean ± SD. N=3.

N301A enhances CD4bs directed antibody neutralization efficiency.

Based on the previous analysis regarding no impairment in viral particle production and Env trimer integrity, we sought to test the effect of removing the aforementioned PNLGs in enhancing CD4bs-directed antibody neutralization. We decided to use the following anti-CD4bs antibodies: VRC01, VRC03, N6, B12 and the control 2G12. The 2G12 antibodies offered a valid positive control as it binds to a mannose cluster far from the CD4bs and should therefore expect little to no neutralization. As the existing PNLG mutant constructs lack a Δ5' LTR, prior to performing the neutralization assay, we had to perform a co-transfection of HEK293T cells using our mutant PNLGS constructs along with a complementing vector, pCMV_cplt. Upon transfection and after 3 days of culture, the transfection supernatant was used to infect U87.CD4.CXCR4 cells to propagate the viruses. After collection, a virus titration was performed to normalize virus input based on RLU for the subsequent neutralization assay **(data not shown)**. The normalized virus input along with the various neutralizing antibodies were incubated for 1 hour and then added to cell cultures of TZM-bl cells for 72 hours. Here were accessed the IC90 of the different CD4bs antibodies as we believe that displays a more stringent and relevant neutralization metric. Based on RLU, we observed that all glycan mutant viruses were more susceptible to neutralization by CD4bs targeting antibodies compared to the WT NL4-3 (without any glycan mutations) **(Figure 2A, B)**. In particular, the virus with the N301A glycan mutation was the most susceptible to neutralization **(Figure 2A, B)**. Consistent with neutralization potency, N6 demonstrated the highest degree of neutralization efficacy **(Figure 2A. B)**. Interestingly, VRC01 demonstrated a higher degree of neutralization potency targeting mutant N301A (IC90 0.5μg/mL) compared to N6 (0.8μg/mL) **(Figure 2A, B)**. This may be due to direct steric hindrance implicated by the glycan at 301 surrounding the CD4bs and VRC01 binding[36,37]. A derivative of VRC01, VRC03 demonstrated a similar degree of neutralization efficacy for all mutant viruses except for mutants N386A (IC90 1.2μg/mL) and N197Q (IC90 0μg/mL) **(Figure 2A, B)**. In the latter case, VRC03 demonstrated weak IC90 neutralization activity for N386A and no IC90 activity for N197Q. However, there was neutralization activity for VRC03 towards N197Q based on IC50 (1.1μg/mL) and IC70 (2.5 μg/mL) **(Figure 2A)**. This alludes to a weaker neutralization potency of VRC03

compared to VRC01. Overall, here we show a proof-of-concept that single glycan mutations, in particular N301A, render viruses more susceptible to neutralization.

A)



B)

		386	NL4-3	301	425	197	363	463
VRC01	IC90	1.3	2.2	0.5	1.2	0.8	1.3	1.1
VRC03	IC90	1.2	2.4	1.2	1.7	0	1.3	1.3
N6	IC90	0.4	2.0	0.8	0.8	0.3	0.8	0.5
B12	IC90	1.2	0	0.7	2.9	0.2	1.2	1.8
2G12	IC90	0	0	0	0	0	0	0

Figure 3-2: Single glycan mutants enhance neutralization mediated by CD4bs-directed antibodies.

Briefly, virus was produced by transfection of the pREC construct with a supplementary vector pCMV_cplt into HEK293T cells. Transfection supernatant was used to infect U87.CD4.CXCR4 cells to propagate the viruses. A virus titration was performed to normalize virus input for neutralization based on RLU. A titration of the following CD4-bs antibodies was used to test the effect of the glycan mutation: VRC01, VRC03, N6 and B12. The control antibody was 2G12. **(A)** Percent neutralization of mutant virus by different anti-CD4bs antibodies. A decrease in RLU is an indication of neutralization by the antibody. **(B)** The IC90 was then calculated for each antibody.

Discussion

In this study, we have begun evaluating the impact of removing single glycans peripheral to the CD4bs and the changes in neutralization efficiency of CD4bs-directed neutralizing antibodies. Using NL4-3, a lab adapted strain, we were able to establish a proof-of-concept that removing even a single glycan can increase neutralizing efficiency. In particular, the presence of the N301 glycan in Env demonstrated the highest degree of steric hindrance and reduces the ability of bnAbs to neutralize virus. Due to time constrains for this MSc thesis, I was not able to finish the rest of this project. The next step in this project was to recreate these findings using the Env sequences isolated from Transmitted/Founder (T/F) virus constructs provided from the NIH AIDS Research Program. To do that we planned on using the same yeast-based recombination strategy to create the mutant NL4-3 Env mutants. In the future, it would be interesting to merge our new *in silico* viral screening approach, to identify potential T/F viral sequences, and then use those T/F Envs as candidate vaccine antigens.

The distribution of the various glycans on Env can be mapped largely to antigenic epitopes. The outer domain contains the neutralizing face and the immunologically silent face[47]. While oligomannose glycans are found on the silent face, complex glycans are found primarily on the neutralizing face[48,49]. As such, their position infers the role of masking such immunological epitopes from nAbs. There are certain positions where the glycan type is conserved, such as the glycan at residue N301. The glycan at this position has consistently been found to be complex and has been heavily studied to serve a protective role from Abs that target the V3 loop. Similar to other studies in the literature, when we removed the glycan at position 301, we observed an enhanced neutralizing efficacy by CD4bs antibodies. While many studies have focused on N301Q mutants and have shown no effect on trimer structure or infectivity, we decided to introduce an Alanine (A) and noticed a similar affect where structure and infectivity were not affected. Alanine is a neutral amino acid and does not incur any post-translational modifications. The small size of the amino acid incurs little to no effect to the overall structure of the protein. The enhanced neutralization mediated by the removal of the glycan at 301 may be due to two reasons: (i) an altered steady-state conformation of Env or (ii) slower fusion kinetics that

may allow prolonged access to antibodies that target the V3-loop (such as anti-CD4bs Abs)[47]. Experiments that show increased sensitivity to monoclonal antibody (Mab) 15e and no change in sensitivity to MPER help explain the altered conformational stability of the Env trimer and the lack of steric hindrance for prolonged binding[50–52]. Additionally, the removal of a complex glycan at that position creates a "glycan hole" that further exposes the CD4bs epitope[50–52]. Considering the beneficial effect the glycan mutations had on the lab adapted strain NL4-3, we plan to evaluate if there is a similar advantageous effect of the N301A glycan mutation on the NIH T/F *env* constructs. Through the neutralization assays, we expect, that the N301A glycan would allow the CD4-bs-directed Abs to have a better neutralization capability –in particular, with VRC01.

An unexpected observation through this study was the negative impact of the N276S mutation. This unexpected conversion from A to S would have eliminated any potential for N-linked glycosylation at this site, however, that does not preclude the possibility of O-linked glycosylation instead – should Env have been produced and displayed on the surface of the virus[53]. The consequences of this substitution are difficult to determine, as patterns of O-linked glycosylation of the HIV-1 virion, as well as the implications in masking vulnerable epitopes, on Env are still unclear. As well, S and N possess different functional groups on their side chains, a hydroxyl and amide group respectively. As such, the introduction of S at this residue, regardless of glycosylation, could cause structural changes in Env that can alter expression or trimer formation. We observed that this mutation did not impact Gag production, based on the p24 ELISA; however, the loss of Env structure on the surface or poor Env stability may explain the small virion size through DLS and the lack of Env detection in the gp140 ELISA and B13 western blot.

The evolutionary arms race between the virus and the immune response concomitantly selects for escape viral variants and B cells that develop to neutralize the virus. The constant exposure of viral Env antigen to B cells constantly encourages accumulation of somatic hyper mutations that increase affinity to the antigen. Sagar *et al*. isolated *env* variable loops 1 and 2 (V1-V2) sequences from primary phase of infection and chronic phase (2-3 years) from nine women infected with HIV-1 clade C[54]. They demonstrated that the chronic viruses had longer V1-V2 loops and were immune to neutralization indicating that this

modification selected for escape mutants[54]. On the contrary, the early viruses succumbed to neutralization and had shorter V1-V2 loops and less PNLGs[54]. The initial Ab response is targeted to the early phase viruses and begins to evolve with the virus[54]. The co-evolution with the virus is thought to be the reason for the development of bnAbs. In particular, work by McCurley *et al*. demonstrated a proof-of-principle that sequential vaccination of evolving Envs presented on virus-like-particles can produce antibodies that can neutralize autologous tier 2 T/F virus[41]. Although there was no antibody breadth towards other tier 2 viruses despite eliciting anti-CD4bs nAbs, the study demonstrated that certain B cell lineages can be initiated through sequential antigen exposure and encourage production of Abs with a heterologous neutralization breadth[41]. The plan is to perform a sequential immunization study using T/F, early/acute and chronic VLPs in BALB/c mice. The mice will receive a total of three intramuscular (IM) vaccinations at three-week intervals. The mice will be organized into 4 experimental group with a respective PBS sham and NL4-3 VLP control group for each. The first group will use either T/F, early/acute or chronic VLP immunogens administered with and without the adjuvant AddaVax, a squalene-oil-in-water nano-emulsion[55]. The second group will be similar to the first, but will receive T/F, early/acute or chronic VLPs immunogens with the N301A glycan mutation. These two studies will help evaluate the ability of the single glycan mutation in selecting a CD4bs-lineage specific B cell and encourage development of CD4-bs-like Abs. The third group will follow a sequential vaccination schedule in an attempt to promote diversification of the B cell response. The first, second and third vaccinations will be T/F, acute/early and chronic VLPs respectively. In addition, these treatments arms will be tested with and without the AddaVax adjuvant. The fourth group will follow the same vaccination schedule but will utilize the N301A glycan mutant VLPs of T/F, acute/early and chronic. Again, this will test the ability of the glycan mutant in selecting for the CD4bs Ab lineage specific and encourage development of higher neutralization breadth. Blood samples will be collected at the beginning of the study and weekly to assess Abs avidity and neutralization breadth for each group.

Overall, we were able to demonstrate that removing the complex glycan present at position 301, enhanced neutralization of CD4bs directed antibodies. Due to time constraints, the

remainder of the experiments for this project were unable to be performed. Regardless, the constructs required to perform the subsequent assays have been complete.

Methods

Molecular engineering of virus like particles (VLPs) with glycan mutants

To incorporate the single glycan mutations, envelope from near full-length (nfl) NL4-3 was amplified in two fragments. The 5' fragment was amplified using EnvB and a reverse mutagenic primer. The 3' fragment was amplified using a matched forward mutagenic primer and 47M. Refer to **Table 1 for primer sequences**. The following PCR cycling conditions were used: cycle 1 consisted of a 98°C step for 2 minutes. Cycle 2 consisted of 3 stages repeated 10 times and consisted of a 98°C step for 10 seconds, followed by a 55°C step for 30 seconds, and then 72°C for 8 minutes. Cycles 3 consisted of 3 stages repeated 25 times at 98°C for 10 seconds, followed by 55°C for 30 seconds and finally 78°C for 8 minutes. Lastly, cycle 4 consisted of a 72°C step for 7 minutes before incubation at 4°C. The PCR amplification reaction consisted of: 1X Phusion buffer (Thermo Fisher), 10µM of dNTP mixture (Thermo Fisher), 1 U of Phusion DNA polymerase (Thermo Fisher), 0.5µM of forward and reverse primers.

Glycan deleted Env VLPs were constructed using a yeast-based recombination cloning method as previously described[42]. Briefly, the amplified fragments were recombined into pREC_nfl_NL4-3_Δenv/URA3 following transfection into *Saccharomyces cerevisiae.* VLP was produced by transfection of the pREC construct into HEK293T cells using Xtremegene HP DNA transfection reagent (Sigma-Aldrich Canada) according to manufactures instructions. Note, a 1:4 ratio of total DNA to Xtremegene was used. The cells were incubated in a CO_2 incubator at 37°C for 72 hours.

Transfected cell culture supernatant was decanted into 50 mL conical tubes and centrifuged at 422G for 10 minutes to pellet any cells. Subsequently, the supernatant was decanted into a Stericup Filter unit with a pore size of 0.45µm (Millipore Germany). Supernatants

containing the VLPs were then added to a Centrion Ultra-15 Centrifuge Filter (100,000 NMWL) (Millipore Germany) and centrifuged at 2571 G for 15 minutes. The retentate was poured into 1.5mL centrifuge tubes and then ultra-centrifuged at 40,000 G for 1 hour at 4°C using a Beckman Avanti J-301. The VLP pellet was resuspended in 250µL of sterile PBS and stored at -20°C. VLPs were quantified by p24 ELISA (AIDS Vaccine Program, National Cancer Institute) according to the manufacturers protocol.

Dynamic light Scattering

Purified VLP preparations were diluted 100-fold in 1X PBS and 10µL of the diluted sample was loaded into a glass cuvette at 25°C. The cuvette was then loaded and analyzed using a DynaPro dynamic light scattering apparatus (Wyatt Technology, CA). Twenty independent read events were collected for each sample for analysis.

Neutralization Assay

A neutralization assay measures neutralization in TZM-bl cells as a function of a reduction in Tat-induced luciferase (Luc) reporter gene expression after a single round of infection. The assay was performed as previously described by the Montefiori Laboratory at Duke University[43]. Briefly, virus was produced by transfection of the pREC construct with a supplementary vector pCMV_cplt into HEK293T cells using Xtremegene HP DNA transfection reagent (Sigma-Aldrich Canada) according to manufactures instructions. Note, a 1:4 ratio of total DNA to Xtremegene was used. The cells were incubated in a CO_2 incubator at 37°C for 3 nights. Half of the transfection supernatant was used to infect U87.CD4.CXCR4 cells to propagate the viruses. Virus was initially collected 48 hours post-syncytia formation and then collected every 3 days after that. A virus titration was performed to normalize virus input for neutralization based on RLU. A titration of the following CD4-bs antibodies was used to test the effect of the glycan mutation: VRC01, VRC03, N6 and B12. The control antibody was 2G12. The IC50, IC70 and IC90 was then calculated for each antibody. A decrease in RLU is an indication of neutralization by the antibody.

Western Blot

For analysis of HIV-1 Env protein production a b13 western blot was performed. VLPs were normalized to 100ng and lysed using 0.0031M Triton Z-100 (Thermo Fisher) by incubating at 37°C for 30 minutes. LDS Sample Buffer (4X) (Thermo Fisher) was added such that the final amount was 1X. Samples were loaded onto a 3-8% Tris-Acetate Gel (Thermo Fisher) submerged in 1X Novex Tris-Acetate Running Buffer and run at 150V for 1 hour on ice. Resolved proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane using 1X Novex Tris-Acetate Transfer Buffer. The protein transfer was carried out at 30V for 90 minutes. The PVDF membrane was then blocked in 4% milk on a shaker for 1 hour at room temperature. The membrane was subsequently washed three times for 10 minutes each using 0.05% PBS-Tween (Sigma-Aldrich Canada). The PVDF membrane was then incubated overnight at 4°C with clarified hybridoma supernatant containing mouse anti-gp12- (B13) mAb (Research and Reference Reagent Program, NIAIS, NIH, contributed by Geroge Lewis Institute of Human Virology, Baltimore, MD, and Bruce Chesebro, NIAID, Hamilton, MT). Membranes were washed and incubated for 1.5 hours with species-specific HRP-conjugated antibody (1:3000 for B13, Thermo Scientific). Blot was developed and quantified using ECL substrate and a C-DiGit chemiluminescence Western blot scanner (LI-COR Biosciences)

GNA gp140 ELISA

The lectin GNA from *Galanthus nivalis* was diluted to a final concentration of 5µg/mL in coating buffer consisting of $NaHCO_3/Na_2CO_3$. 100µL of the final solution was added to a Greiner high binding 96-well plate (Sigma) overnight and sealed at 4°C. The plate was washed 3X in wash buffer consisting of PBS and 0.05% Tween 20 after which 100µL of dilution buffer containing 10% FBS (Corning) was added to each well and placed on a shaker for 2 hours. The standard recombinant gp140 SF162 (NIH AIDS Reagent Database) was diluted to 100ng/mL in dilution buffer and serially diluted 2-fold to 7.8ng/mL. Samples were serially diluted 3-fold from 100ng to 5.08×10^{-4} based on p24. Samples were lysed in 1% Triton-X solution. After blocking, the plate was washed as previously described and 100µL of standard and samples were added to the plate for 1-1.5 hours on a shaker at room temperature. Plate was washed and 100µL of monoclonal anti-gp41 Ab 5FE (NIH AIDS

Reagent Database) diluted to a final concentration of 1μg/mL was added to each well for 1 hour on a shaker at room temperature. The plate was washed again and 100μL of the secondary polyclonal Ab GAH-IgG conjugated with peroxidase diluted 1:5000 was added to each well for 1 hour on a shaker at room temperature. The plate was washed and 100μL of room temperature TMB color substrate (Thermo Fisher) was added to each well and incubated for 10-15 minutes in the dark. After the incubation, 100μL of 2.5M H_2SO_4 was added to stop the colorimetric reaction and subsequently read using a Cytation5 reader (Biotek) at 450nm. Binding affinity of glycan mutant VLP was calculated based on the standard.

Creating chimeric Transmitted/founder (T/F) virial constructs, $Env_{T/F}$

$Env_{T/F}$ were constructed from ten Transmitted/founder (T/F) subtype B full-length genomes obtained from the NIH Aids Reagent Program. Briefly, full-length Env from each T/F plasmid was amplified using primers EnvB and EnvM **(primer sequences Table 3-1)**. The following PCR cycling conditions were used: cycle 1 consisted of a 98°C step for 2 minutes. Cycle 2 consisted of 3 stages repeated 10 times and consisted of a 98°C step for 10 seconds, followed by a 55°C step for 30 seconds, and then 72°C for 8 minutes. Cycles 3 consisted of 3 stages repeated 25 times at 98°C for 10 seconds, followed by 55°C for 30 seconds and finally 78°C for 8 minutes. Lastly, cycle 4 consisted of a 72°C step for 7 minutes before incubation at 4°C. The PCR amplification reaction consisted of: 1X Phusion buffer (Thermo Fisher), 10μM of dNTP mixture (Thermo Fisher), 1 U of Phusion DNA polymerase (Thermo Fisher), 0.5μM of forward and reverse primers.

Env chimeric viruses were constructed using a yeast-based recombination cloning method as previously described. Briefly, the amplified envelope gene from the ten different T/F clones were recombined into pREC_nfl_NL4-3_Δenv/URA3 following transfection into *Saccharomyces cerevisiae.* Virus was produced by transfection of the pREC construct with a supplementary vector pCMV_cplt into HEK293T cells using Xtremegene HP DNA transfection reagent (Sigma-Aldrich Canada) according to manufactures instructions. Note, a 1:4 ratio of total DNA to Xtremegene was used. The cells were incubated in a CO_2 incubator at 37°C for 3 nights. Half of the transfection supernatant was used to infect

U87.CD4.CCR5 cells to propagate the viruses. Virus was initially collected 48 hours post-syncytia formation and then collected every 3 days after that.

Creating chimeric Transmitted/founder (T/F) virial constructs with N301A glycan mutation, $Env_{T/F-301}$

To incorporate the single glycan mutations, envelope from $Env_{T/F}$ was amplified in two fragments. The 5' fragment was amplified using EnvB and the reverse N301A mutagenic primer. The 3' fragment was amplified using the forward N301A mutagenic primer and EnvM. Refer to **Table 3-1 for primer sequences**. Yeast recombination and virus production was performed as mentioned previously (**Creating chimeric Transmitted/founder (T/F) virial constructs, $Env_{T/F}$)**

Ethics statement

Human penile and cervical tissue was collected with written consent according to the local research committee guidelines. All tissues were obtained under protocols approved by the Imperial College NHS Trust Tissue Bank and the National Research in accordance with the Human Tissue Act 2004.

Tissue Explant Model Experiment

Refer to the following publication for detailed account on performing the tissue explant model experiment (Katja, K *et al.* 2020). Briefly:

Creating the acute/early and chronic chimeric viruses: Eleven early/acute subtype B envelopes were obtained from the Center of HIV-1/AIDS Vaccine Immunology (CHAVI)[2]. Three chronic subtype B envelopes were obtained from a Belgium cohort[50,51]. Chimeric viruses were created using a previously established yeast-based recombination system[55], where the different envelope genes were inserted into pREC_nfl_NL4-3_Δenv/URA3. As a result, the pREC constructs differed solely on the inserted *env* sequence. HEK293T cells were co-transfected with a rescue vector pCMV_cplt and virus was propagated in U87.CD4.CCR5 cells. The acute viruses are referred to as: B1, B2, B3, B4, B7, B8, B9, B14, B17, B19 and B20. The chronic viruses are referred to as: I10, K44

and Q0. Reverse transcriptase activity of each virus was used to quantify infectivity for future tissue infections.

Transmission fitness assay and Sequence data analysis: Frozen endocervical explants were acquired through the National Disease Research Interchange (NDRI) and cut in approximately 3mm X 3mm X 3mm pieces and placed in each well of a 48-well plate. Pools of four to five acute and three chronic Env chimeric viruses (150 IU per virus; as determined by reverse transcriptase activity) were mixed in culture containing 3-5 pieces of cut explant tissue for 3 hours at 37°C. Tissue was washed three times in complete RPMI 1640 media (Sigma). Migratory cells (MC) from the tissue were harvested 24 hours later and co-cultured with 10^5 PM1 CD4 T cells. On day 10, tissue was lysed with 300μL of 1% Triton-X100 supplemented with proteinase K over night at 56°C and cells from the MC-PM-1 T cell co-culture were lysed with 300μL of 1% Triton-X100. DNA was extracted using PureLink Pro 96 Genomic DNA kit (Invitrogen). Each multi-virus competition was performed in two donors in triplicate. Following DNA extraction, PCR amplification of the C2-V3 envelope region was performed and products were ligated with barcode oligonucleotides, purified and subject to Next-Generation-Sequencing (NGS) (Roche 454 GS Junior sequencer). Using SeekDeep, a custom pipeline that sorts sequences based on barcodes and aligns sequences within barcode bins, viruses within each competition were quantified.

Table 3-1: Primer sequences used in study.

Primer name	Sequence
EnvB	agaaagagcagaagacagtggcaatga
EnvM	tagcccttccagtccccccttttctttta
47M	tccctagttagccagagagctccc
IntF2	tggaaaggtgaaggggcagtagtaatacaagataatagtgacataaaagtagtgccagcagcacacgcaaagatcatc
IntR2	caatcatcacctgccatctgttttccataatccctgatgatctttgcgtgtgctgctggcactacttttatgtc
N197Q F	agctataggttgataagttgtcagacctcagtcattacacaggcc
N197Q R	ggcctgtgtaatgactgaggtctgacaacttatcaacctatagct
N276A F	gatgtagtaattagatctgccgctttcacagacaatgctaaaacc
N276A R	ggttttagcattgtctgtgaacgtggcagatctaattactacatc
N301A F	attaattgtacaagacccaacgccaatacaagaaaaagtatccgt
N301A R	attaattgtacaagacccaacgccaatacaagaaaaagtatccgt
N363A F	aataataaaacaataatctttgcgcaatcctcaggaggggaccca
N363A R	tgggtcccctcctgaggattgcgcaaagattattgttttattat
N386A F	ggaggggaatttttctactgtgcttcaacacaactgtttaatagt
N386A R	actattaaacagttgtgttgaagcacagtagaaaaattcccctcc
N463A F	agagatggtggtaataacaacgctgggtccgagatcttcagacct
N463A R	aggtctgaagatctcggacccagcgttgttattaccaccatctct
N425K F	tgcagaataaaacaatttataaagatgtggcaggaagtaggaaaa
N425K R	ttttcctacttcctgccacatctttataaattgttttattctgca

References

1. West, A. P. *et al.* Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* **156**, 633–48 (2014).
2. Munro, J. B. & Mothes, W. Structure and Dynamics of the Native HIV-1 Env Trimer. *J. Virol.* **89**, 5752–5 (2015).
3. Wagh, K. *et al.* Completeness of HIV-1 Envelope Glycan Shield at Transmission Determines Neutralization Breadth. *Cell Rep.* **25**, 893-908.e7 (2018).
4. Bonsignori, M. *et al.* Staged induction of HIV-1 glycan–dependent broadly neutralizing antibodies. *Sci. Transl. Med.* **9**, eaai7514 (2017).
5. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–76 (2013).
6. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–12 (2003).
7. Zhang, M. *et al.* Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* **14**, 1229–1246 (2004).
8. Pritchard, L. K. *et al.* Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat. Commun.* **6**, 7479 (2015).
9. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10**, 14 (2013).
10. Freund, N. T. *et al.* Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. *Sci. Transl. Med.* **9**, (2017).
11. Moore, P. L. *et al.* Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nat. Med.* **18**, 1688–1692 (2012).
12. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
13. Zhou, T., Doria-Rose, N. A., Mcdermott, A. B., Mascola, J. R. & Kwong, P. D. Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. (2017). doi:10.1016/j.celrep.2017.04.013
14. Klasse, P. J., Sanders, R. W., Cerutti, A. & Moore, J. P. How can HIV-type-1-Env immunogenicity be improved to facilitate antibody-based vaccine development? *AIDS Res. Hum. Retroviruses* **28**, 1–15 (2012).
15. Moir, S. & Fauci, A. S. B-cell responses to HIV infection. *Immunol. Rev.* **275**, 33–48 (2017).
16. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–19 (2011).
17. Abrahams, M.-R. *et al.* Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* **83**, 3556–67 (2009).
18. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7 (2008).
19. Gnanakaran, S. *et al.* Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *PLoS Pathog.* **7**,

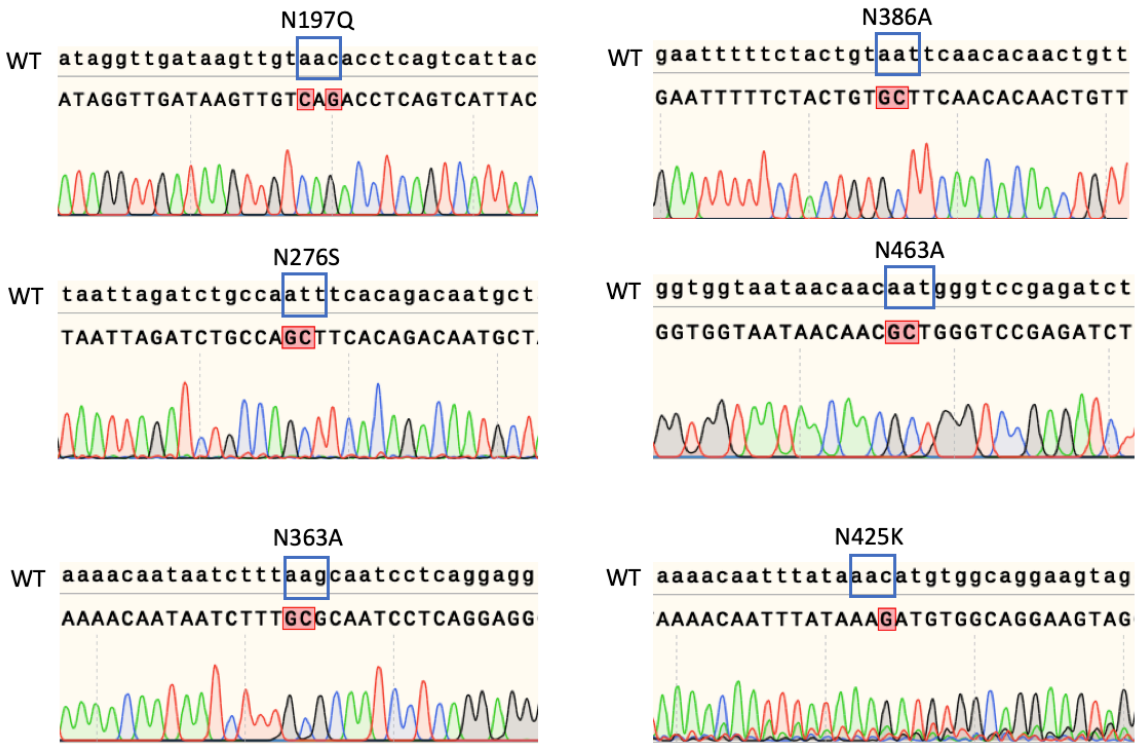
e1002209 (2011).
20. Tully, D. C. *et al.* Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathog.* **12**, e1005619 (2016).
21. Klein, K. *et al.* Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS Pathog.* **14**, e1006754 (2018).
22. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–19 (2011).
23. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–9 (2009).
24. Doria-Rose, N. A. *et al.* Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables. *J. Virol.* **84**, 1631–6 (2010).
25. Euler, Z. *et al.* Longitudinal analysis of early HIV-1-specific neutralizing activity in an elite neutralizer and in five patients who developed cross-reactive neutralizing activity. *J. Virol.* **86**, 2045–55 (2012).
26. Klein, F. *et al.* Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization. *Cell* **153**, 126–138 (2013).
27. Mouquet, H. *et al.* Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* **467**, 591–595 (2010).
28. Kwong, P. D. & Mascola, J. R. Human Antibodies that Neutralize HIV-1: Identification, Structures, and B Cell Ontogenies. *Immunity* **37**, 412–425 (2012).
29. Shcherbakov, D. N., Bakulina, A. Y., Karpenko, L. I. & Ilyichev, A. A. Broadly Neutralizing Antibodies against HIV-1 As a Novel Aspect of the Immune Response. *Acta Naturae* **7**, 11–21
30. Moody, M. A. *et al.* Immune perturbations in HIV-1–infected individuals who make broadly neutralizing antibodies. *Sci. Immunol.* **1**, aag0851 (2016).
31. Moore, P. L., Williamson, C. & Morris, L. Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends in Microbiology* **23**, 204–211 (2015).
32. Piantadosi, A. *et al.* Breadth of neutralizing antibody response to human immunodeficiency virus type 1 is affected by factors early in infection but does not influence disease progression. *J. Virol.* **83**, 10269–74 (2009).
33. Rusert, P. *et al.* Determinants of HIV-1 broadly neutralizing antibody induction. *Nat. Med.* **22**, 1260–1267 (2016).
34. Hoot, S. *et al.* Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4bs bNAbs. *PLoS Pathog.* **9**, e1003106 (2013).
35. Xiao, X. *et al.* Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: Implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun.* **390**, 404–409 (2009).
36. Zhou, T. *et al.* Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep.* **19**, 719–732 (2017).
37. Liang, Y. *et al.* Changes in Structure and Antigenicity of HIV-1 Env Trimers Resulting from Removal of a Conserved CD4 Binding Site-Proximal Glycan. *J.*

Virol. **90**, 9224–36 (2016).
38. Hoot, S. *et al.* Recombinant HIV envelope proteins fail to engage germline versions of anti-CD4bs bNAbs. *PLoS Pathog.* **9**, e1003106 (2013).
39. Do Kwon, Y. *et al.* Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol.* **22**, 522–531 (2015).
40. Pankrac, J. *et al.* A heterogeneous human immunodeficiency virus-like particle (VLP) formulation produced by a novel vector system. *NPJ vaccines* **3**, 2 (2018).
41. McCurley, N. P. *et al.* HIV transmitted/founder vaccines elicit autologous tier 2 neutralizing antibodies for the CD4 binding site. *PLoS One* **12**, e0177863 (2017).
42. Moore, D. M., Arts, E. J., Gao, Y. & Marozsan, A. J. A yeast recombination-based cloning system to produce chimeric HIV-1 viruses and express HIV-1 genes. *Methods Mol. Biol.* **304**, 369–85 (2005).
43. Montefiori, D. C. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **485**, 395–405 (2009).
44. Ratcliff, A. N., Shi, W. & Arts, E. J. HIV-1 Resistance to Maraviroc Conferred by a CD4 Binding Site Mutation in the Envelope Glycoprotein gp120. *J. Virol.* **87**, 923–934 (2013).
45. Dubrovskaya, V. *et al.* Targeted N-glycan deletion at the receptor-binding site retains HIV Env NFL trimer integrity and accelerates the elicited antibody response. *PLOS Pathog.* **13**, e1006614 (2017).
46. Zhou, T. *et al.* Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep.* **19**, 719–732 (2017).
47. Binley, J. M. *et al.* Role of Complex Carbohydrates in Human Immunodeficiency Virus Type 1 Infection and Resistance to Antibody Neutralization. *J. Virol.* **84**, 5637–5655 (2010).
48. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433**, 834–841 (2005).
49. Zhu, X., Borchers, C., Bienstock, R. J. & Tomer, K. B. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* **39**, 11194–11204 (2000).
50. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–82 (2002).
51. McCaffrey, R. A., Saunders, C., Hensel, M. & Stamatatos, L. N-Linked Glycosylation of the V3 Loop and the Immunologically Silent Face of gp120 Protects Human Immunodeficiency Virus Type 1 SF162 from Neutralization by Anti-gp120 and Anti-gp41 Antibodies. *J. Virol.* **78**, 3279–3295 (2004).
52. Schønning, K., Jansson, B., Olofsson, S., Nielsen, J. O. & Hansen, J. E. S. Resistance to V3-directed neutralization caused by an N-linked oligosaccharide depends on the quaternary structure of the HIV-1 envelope oligomer. *Virology* **218**, 134–140 (1996).
53. Bernstein, H. B., Tucker, S. P., Hunter, E., Schutzbach, J. S. & Compans, R. W. Human immunodeficiency virus type 1 envelope glycoprotein is modified by O-linked oligosaccharides. *J. Virol.* **68**, 463–468 (1994).
54. Sagar, M., Wu, X., Lee, S. & Overbaugh, J. Human Immunodeficiency Virus Type 1 V1-V2 Envelope Loop Sequences Expand and Add Glycosylation Sites over the

Course of Infection, and These Modifications Affect Antibody Neutralization Sensitivity. *J. Virol.* **80**, 9586–9598 (2006).

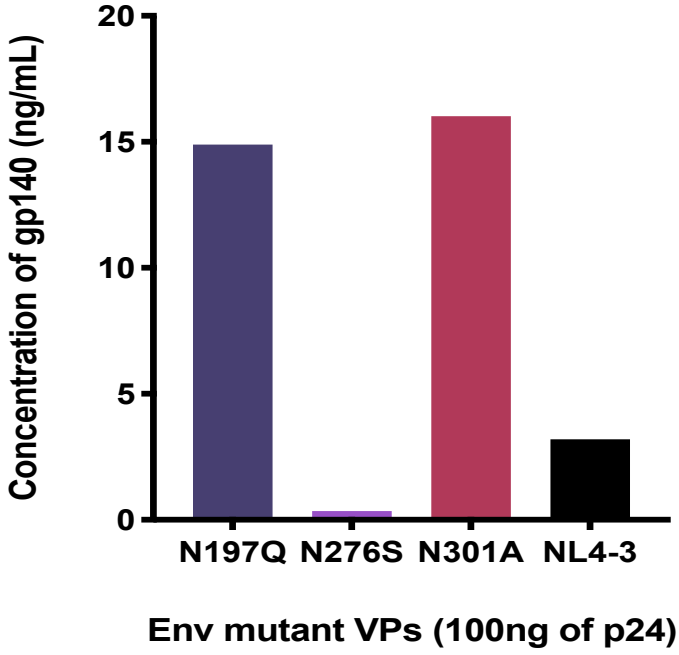
55. Cao, W. *et al.* An oil-in-water nanoemulsion enhances immunogenicity of H5N1 vaccine in mice. *Nanomedicine Nanotechnology, Biol. Med.* **12**, 1909–1917 (2016).

Supplementary Figure



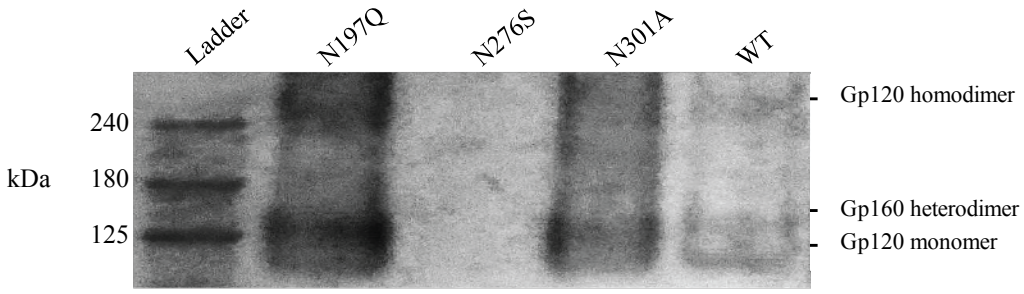
Supplementary Figure 3-1: Site-directed mutagenesis to remove PNLGs surrounding the CD4bs.

Plasmids generated from the site-directed mutagenesis cloning strategy were verified using Sanger sequencing. Data represents a chromatogram generated from the sequencing run aligned with the WT NL4-3 sequence at the specified PNLG site. For each chromatogram, nucleotides in red boxes indicate the nucleotide mutation that was introduced and differ from the WT NL4-3 *Env* sequence. For each chromatogram green peaks indicate adenine (A), red peaks indicate thymine (T), **black peaks** indicate guanine (G), and blue peaks indicate cytosine (C). Blue boxes identify the three nucleotides forming the codon at each PNLG site and their associated amino acid. A detailed description of the directed mutagenesis cloning strategy can be found in the materials and methods section of this study.



Supplementary Figure 3-2: N197Q and N301A single glycan mutants do not impact oligomeric Env glycoprotein, with the exception of N276S.

HEK293T cells were transfected and VLPs were purified as previously described. Production of oligomeric envelope glycoprotein was quantified using a gp140 ELISA. The gp140 ELISA was normalized to 200ng of p24 for each VLP. Oligomeric envelope glycoproteins were detected using the F53 antibody. Production of oligomeric envelope glycoproteins was detected from all VLPs, with the exception of VLP N276S. Optical density values at a wavelength of 450nm (OD_{450}) were read using a Cytation5 reader (Biotek) and binding affinity of glycan mutant VLP was calculated based on the standard. Statistical significance was not determined due to sample size. N=1.



Supplementary Figure 3-3: N197Q and N301A single glycan mutants do not impact gp120 Env glycoprotein, with the exception of N276S.

HEK293T cells were transfected and VPs were purified as previously described. Purified VPs were then quantified using a p24 ELISA and normalized (100ng of p24) for b13 western blot analyses. Briefly, the PVDF membrane was then incubated overnight at 4°C with clarified hybridoma supernatant containing mouse anti-gp12- (B13) mAb and incubated for 1.5 hours with species-specific HRP-conjugated antibody (1:3000 for B13). Blot was developed and quantified using ECL substrate and a C-DiGit chemiluminescence Western blot scanner. Bands seen below the 125kDa marker likely represent monomeric gp120(~120kDa), and bands seen above this marker likely represent gp120-gp41 heterodimers (~160kDa). Mutant VLPs N197Q and N301A and the WT VLP showed sufficient production of gp120 envelope glycoprotein, with the exception of VLP N276S, which showed no expression of gp120 envelope glycoprotein. Blot is a representative; N=2.

Chapter 4 « Overall Discussion »

The introduction of HIV-1 into the human population is a result of the zoonotic transmission of SIV[1,2]. Identified as the etiological agent behind AIDS in the early 1980's, HIV has continued to spread globally and become a major health concern. It has become clear that natural immunity is not protective, and that HAART treatment does not cure an individual[3,4]. As a result, an effective and safe vaccine against HIV-1 is necessary to tackle the AIDS pandemic. To date, there is no vaccine that has been able to elicit a protective antibody response[5–11]. Knowledge from previous HIV-1 vaccine studies has identified several potential reasons why vaccines to date have been ineffective. Reasons such as (1) lack of structural homology to proteins on wild-type virus; (2) inability to elicit protective neutralizing antibodies responses; and (3) inability to target the diversity found circulating in the human population, have become a recent focus while designing novel HIV-1 antigens. Towards this goal, the focus of this MSc thesis was to understand HIV-1 transmission and Env glycosylation in order to design a novel HIV-1 vaccine antigen. The premise of our vaccine design relies on following a sequential regiment whereby the immune system evolves with the virus and develops a dominant antibody response to the exposed CD4bs[12,13]. Using our VLP-based formulation[14], we would start with a diverse pool of T/F VLPs and follow with additional boosts using pools of acute/early VLPs and chronic VLPs. The Env's on these VLPs would all contain a glycan hole that in theory would allow the CD4bs to be the immunodominant epitope. VLPs from different time points of infection would target the genetic diversity found in circulating HIV. This in turn may encourage HIV-specific B cell to evolve with the virus and create potent neutralizing antibodies. The goal would be to create bnAbs or bnAbs-like antibodies that possess a similar degree of potency to wild-type virus. As T/F are notoriously difficult to characterize, very few sequences exist from which a vaccine can be generated. To perform this study, the first goal was to therefore develop methodologies that could identify T/F viruses, so that they could be exploited for vaccine development. The second goal was to identify a glycan mutation that would increase neutralization susceptibility of HIV-1 VLPs to anti-CD4bs antibodies, such as VRC01.

In Chapter 2, we presented novel evidence that the immunostimulatory potential of the viral genome of T/F viruses may play an important role in transmission success. In particular, we show evidence that viruses containing more TNF-α inducing motifs can exit the cervix more efficiently compared to viruses with lower numbers (**Fig 2-7, 2-9**). In addition, we show that the highly pro-inflammatory viral RNA may be compartmentalized within the semen and cervix of the donor (**Fig 2-5, 2-6, 2-8**). This data is significant as understanding what makes a T/F virus successful at the time of infection serves two purposes. First, it advances our understanding of HIV-1 biology and the key players that drive successful transmission. Secondly, it provides evidence for the development of an efficacious vaccine against HIV-1. Acknowledging that, in 80% of sexual transmissions, the T/F variant causes the productive infection in the recipient, an effective vaccine should therefore target this viral variant to prevent successful transmission[15–19]. Additionally, observing the differences in immunostimulatory potential in compartments may encourage isolation of viruses from sites that play a relevant role in sexual transmission. Future studies looking at SIV transmission and differences in infection rates between new world and old-world monkeys can provide additional understanding of the impact the viral RNA may have on transmission.

In Chapter 3, we demonstrated a proof-of-concept using NL4-3, where mutant viruses lacking a single glycan peripheral to the CD4bs are highly susceptible to neutralization mediated by CD4bs directed antibodies (**Fig 3-2**). In particular, deletion of glycan N301 demonstrated the highest degree of steric hindrance mediated by the presence of the complex glycan at that position. While the glycans on Env provide structural integrity to the quaternary structure of Env, it also creates a glycan mask whereby vulnerable epitopes are hidden from the immune system. The selective pressure from the ensuing immune response selects for certain glycosylated variants thereby limiting proper humoral immune stimulation[20–26]. Due to time constraints, further analysis regarding the mutations was not possible. In the future, it will be of interest to assess the impact of removing a combination of glycans on Env structure integrity and neutralization susceptibility. Additionally, it is of great interest to evaluate the impact of a N301-glycan deficient Env on T/F viruses and the degree of susceptibility it poses to CD4bs directed antibodies.

The main goal of this thesis was to create a novel vaccine formulation that can elicit production of protective antibodies. Our aim was to combine knowledge of HIV-1 transmission with glycosylation patterns to create a VLP-based antigens that can specifically target the T/F variant and prevent a productive infection. As this thesis potentially identifies a novel contributor to HIV-1 T/F virus success and neutralization susceptibility through targeted glycan removal, investigating the downstream effects of these VLP constructs in a sequential vaccine formulation is an important avenue of future research for HIV-1 vaccine development.

References

1. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. doi:10.1101/cshperspect.a006841
2. WATANABE, M. E. Origins of HIV: The Interrelationship between Nonhuman Primates and the Virus. *Bioscience* **54**, 810 (2004).
3. Kulkosky, J. & Bray, S. HAART-persistent HIV-1 latent reservoirs: their origin, mechanisms of stability and potential strategies for eradication. *Curr. HIV Res.* **4**, 199–208 (2006).
4. Huyghe, J., Magdalena, S. & Vandekerckhove, L. Fight fire with fire: Gene therapy strategies to cure HIV. *Expert Rev. Anti. Infect. Ther.* **15**, 747–758 (2017).
5. Pitisuttithum, P. *et al.* Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J. Infect. Dis.* **194**, 1661–71 (2006).
6. Edlefsen, P. T., Gilbert, P. B. & Rolland, M. Sieve analysis in HIV-1 vaccine efficacy trials. *Curr. Opin. HIV AIDS* **8**, 432–436 (2013).
7. Volk, J. E. *et al.* The HVTN503/Phambili HIV vaccine trial: a comparison of younger and older participants. *Int. J. STD AIDS* **25**, 332–40 (2014).
8. Emu, B. *et al.* HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. *J. Virol.* **82**, 5398–407 (2008).
9. Altfeld, M. *et al.* HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1. *PLoS Med.* **3**, e403 (2006).
10. Harrer, T. *et al.* Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection. *AIDS Res. Hum. Retroviruses* **12**, 585–92 (1996).
11. Kim, J. H., Excler, J.-L. & Michael, N. L. Lessons from the RV144 Thai phase III HIV-1 vaccine trial and the search for correlates of protection. *Annu. Rev. Med.* **66**, 423–37 (2015).
12. McCurley, N. P. *et al.* HIV transmitted/founder vaccines elicit autologous tier 2 neutralizing antibodies for the CD4 binding site. *PLoS One* **12**, e0177863 (2017).
13. Sagar, M., Wu, X., Lee, S. & Overbaugh, J. Human Immunodeficiency Virus Type 1 V1-V2 Envelope Loop Sequences Expand and Add Glycosylation Sites over the

Course of Infection, and These Modifications Affect Antibody Neutralization Sensitivity. *J. Virol.* **80**, 9586–9598 (2006).
14. Pankrac, J. *et al.* A heterogeneous human immunodeficiency virus-like particle (VLP) formulation produced by a novel vector system. *NPJ vaccines* **3**, 2 (2018).
15. Abrahams, M.-R. *et al.* Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* **83**, 3556–67 (2009).
16. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7 (2008).
17. Gnanakaran, S. *et al.* Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *PLoS Pathog.* **7**, e1002209 (2011).
18. Tully, D. C. *et al.* Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathog.* **12**, e1005619 (2016).
19. Klein, K. *et al.* Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. *PLoS Pathog.* **14**, e1006754 (2018).
20. Freund, N. T. *et al.* Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. *Sci. Transl. Med.* **9**, (2017).
21. Moore, P. L. *et al.* Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nat. Med.* **18**, 1688–1692 (2012).
22. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
23. Zhou, T. *et al.* Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Rep.* **19**, 719–732 (2017).
24. Klasse, P. J., Sanders, R. W., Cerutti, A. & Moore, J. P. How can HIV-type-1-Env immunogenicity be improved to facilitate antibody-based vaccine development? *AIDS Res. Hum. Retroviruses* **28**, 1–15 (2012).
25. Moir, S. & Fauci, A. S. B-cell responses to HIV infection. *Immunol. Rev.* **275**, 33–48 (2017).
26. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–19 (2011).

Curriculum Vitae

Name:	Rahul Pawa
Post-secondary Education and Degrees:	University of Western Ontario London, Ontario, Canada 2014-2018 B.A University of Western Ontario London, Ontario, Canada 2018-2020 MSc
Honours and Awards:	RGE Murray Graduate Scholarship 2018-2019, 2019-2020 FW Luney Graduate Fellowship 2018 Western Continuing Admission Scholarship 2014-2018
Related Work Experience	Summer Undergraduate Research Fellowship University of Western Ontario 2017

Publications:

Joshua Pankrac, Katja Klein, Paul F. McKay, Deborah F.L. King, Katie Bain, Jason Knapp, Tsigereda Biru, Chanuka N. Wijewardhana, **Rahul Pawa**, David H. Canaday, Yong Gao, Sarah Fidler, Robin J. Shattock, Eric J. Arts and Jamie F.S. Mann. A heterogeneous human immunodeficiency virus-like particle (VLP) formulation produced by a novel vector system. *npj Vaccines* (2018)

Jamie F.S. Mann, Joshua Pankrac, Katja Klein, Paul F. McKay, Deborah F.L. King, Richard Gibson, Chanuka N. Wijewardhana, **Rahul Pawa**, Jodi Meyerowitz, Yong Gao, David H. Canaday, Mariano Avino, Art F.Y. Poon, Caroline Foster, Sarah Fidler, Robin J. Shattock and Eric J. Arts. A Targeted Reactivation of latnet HIV-1 Using AN Activator Vector. *EBio Med* (2020) *manuscript under review*

Katja Klein and **Rahul Pawa**, Chanuka N. Wijewardhana, Ryan Troyer, Eric J. Arts and Jamie F.S. Mann. The antigenicty of HIV-1 viral RNA contributes to its transmission fitness. *manuscript in prepapration*