MALE GENITAL TRACT VERSUS BLOOD HIV-1 COMPARTMENTALIZATION AND SELECTION: THE FIRST STEP OF THE TRANSMISSION BOTTLENECK?

SAMUEL MUNDIA KARIUKI
KRKSAM004

This thesis is submitted in fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY in the Division of Immunology, Department of Pathology in the Faculty of Health Sciences at the University of Cape Town

Supervisor: Associate Professor Jeffrey Robert Dorfman
Co-supervisor: Professor Carolyn Williamson

OCTOBER 2019

This PhD thesis is dedicated to my mother; Nancy Wanjiru: It has been 26 years since you left. I have felt all alone for the better part of these years but knowing that you are watching over me; because of the faith you had in me; gives me the courage to soldier on.
Thengio mûno mum. Nîndagûciarîire tûcûcû.

Table of Contents

Thesis summary

Introduction

Sexual transmission of HIV-1 accounts for more than 80% of all the transmissions globally. After transmission, approximately 80% of the newly disseminated infections can be traced to a single variant, which comes from the minor HIV-1 population within the transmitting donor. This has led to the widely accepted idea of an HIV-1 transmission bottleneck. The nature of this bottleneck is not fully understood. Many studies working on understanding the nature of the transmitted virus have reported discordant traits of transmitted/founder viruses compared to viral isolates from chronically infected individuals. Such studies therefore lacked analysis of the intermediate step between these two populations: HIV-1 from the genital tract of the donors: where viruses are on their way to transmission to a new recipient. Importantly, numerous prior studies have shown that there is compartmentalization of HIV-1 populations between the general circulation and the genital tract, raising the possibility that the genital tract is an important selective environment. Collectively, prior studies of genital tract compartmentalization in males detected compartmentalization in about half of the donors studied, although by using techniques with limited depth of sampling than that employed in our study.

The virus that establishes disseminated infection in a new recipient is selected. However, the extent to which this selection occurs before, during or after crossing the mucosal surfaces of the recipient is less clear – the period during which transmission selection could extend far back to the donor, and the donor's genital tract. In other words, it is not clear as to what extent the transmission bottleneck occurs during compartmentalization of viral populations in the genital tract tissues. The design of an effective vaccine and other intervention strategies will rely upon the understanding the nature of the transmitted virus as this is the virus that must be targeted.

This thesis

Compartmentalization of minor variants cannot be tracked by techniques previously used to describe compartmentalization between the genital tract and the blood circulation. We

therefore used deep sequencing-based techniques to further understand compartmentalization of viral populations between blood and the male genital tract. In addition, we tested the sensitivity of variants to a range of entry and other inhibitors in order to explore possible changes in function that may arise when viral variants grow in the shifted selective milieu of the genital tract. We further hypothesized that this change of selective milieu as HIV-1 moves from blood into the genital tract may lead to viral variants in semen that are sensitive to autologous neutralization because such sensitive variants may be able to grow in the genital tract, which is presumably partially or completely shielded from antibodies. Because the viral populations in semen comes from a site that may be relatively protected from antibodies, they may be permitted to evolve differently in the relative absence of antibody pressure. It is possible that evolution of the virus within the genital tract is a significant part of the change the virus undergoes on its way to establishment of a new disseminated infection in the new recipient. We considered this possibility because even some small molecules like those of some antiretroviral drugs do not penetrate the genital tract effectively under some circumstances, raising the possibility that antibodies might not always penetrate in all areas of the genital tract.

This thesis had three objectives:

1. To evaluate HIV-1 compartmentalization in blood and the male genital tract using next generation sequencing to understand the nature of viral populations in these anatomical sites in greater detail.
2. To identify the differences in sensitivity of blood and semen variants to entry inhibitors to obtain information about differences in function between HIV-1 populations in blood vs the male genital tract.
3. To compare neutralization sensitivities of viral variants compartmentalized in blood and semen by testing their sensitivity to neutralization by autologous antibodies. As a control, we measured sensitivity to a pool of clade-matched heterologous sera to determine if any observed difference was due to global changes in neutralization sensitivity.

Methods

Study participants

Forty-four HIV-1 seropositive men were recruited and then requested to donate blood and semen samples at ANOVA Health's Ivan Toms clinics at Woodstock and Green Point or through their mobile clinic in Khayelitsha, all in Cape Town, South Africa. Viral loads from blood and semen and $CD4^+$ T cell counts from blood were measured. HIV-1 was enriched from semen using a Nycodenz gradient, and then concentrated using ultracentrifugation.

Chapter 2: HIV-1 Compartmentalization in blood and semen

Next generation sequencing on Illumina paired-end Miseq platform was performed. To our knowledge there is no published study that has used this technique to study male genital tract HIV-1 variants in chronically infected male donors, although there is one that does so for acutely infected donors. We argue here that this is a superior method of sampling populations in blood and the male genital tract. In particular, it allowed us to more accurately track minor populations within each compartment. Additionally, the use of PrimerID approach allowed us to more clearly identify clonal amplification events in the HIV-1 populations. Sequencing was performed on either the V3 or C3-V5 region of the HIV-1 envelope gene from paired blood and semen samples from 11 donors.

To evaluate compartmentalization, populations from blood and semen were compared using three standard techniques, Slatkin Maddison Test (SMT), Wrights measure of population subdivision (F_{ST}) and nearest neighbour statistic (S_{nn}). Clonal amplification and results of modelling a lower depth of sampling are also presented.

Chapter 3: Sensitivity of blood and semen variants to entry inhibitors and changes in function

From three subjects who exhibited the highest extent of compartmentalization, full-length envelope clones derived from semen and blood RNA were made using limiting dilution PCR (single genome amplification), which provided the advantage of minimizing PCR-based artificial recombination. A high fidelity Taq polymerase was also used to minimize base-substitution errors. An average of 10 clones were isolated per compartment. Pseudoviruses were then constructed from the full-length envelope clones from blood and semen.

The sensitivities of these pseudoviruses were tested against HIV-1 entry inhibitors; Maraviroc, PSC-RANTES, enfurvirtide (T-20) and JM2987. Maraviroc and PSC-RANTES are CCR5 inhibitors while JM2987 is a CXCR4 inhibitor. Enfurvirtide (T-20) is a fusion inhibitor blocking the virus from entering cells. The full-length clones used to make the pseudoviruses were also sequenced and genomic variations in variable loop characteristics (length and number of potential glycosylation sites) between blood and semen compared.

Chapter 4: Sensitivity of blood and semen variants to autologous and heterologous antibodies

To study the differences in sensitivity of blood and semen variants to antibodies, pseudoviruses cloned from semen RNA and blood RNA (above) were tested for their sensitivities to donor antibodies collected at the same time the samples were collected or to a pool of HIV-1 subtype C sera.

Results

Objective 1: Viral compartmentalization via next generation sequencing

HIV-1 populations were compartmentalized in all the 11 donors studied but to varying extents. Donor SVB043 had the most compartmentalized viral populations between blood and the male genital compartment using all the three measures of compartmentalization. Further analysis of the phylogenetic trees revealed that some clusters contained either purely blood or semen sequences, even in trees generated from analysis of donors with weakest compartmentalization. This might explain the viral compartmentalization signal in these weakly compartmentalized donors.

To mimic reduced sampling depth, subsampling of the Illumina Miseq data with a small number of sequences was done. This analysis revealed that viral compartmentalization between blood and male genital tract would likely (>50% estimated likelihood) have been detected in only 5/11 (45%) of the donors, a proportion which is very similar to the aggregate proportion from previous studies that had used single genome amplification (SGA) analysis. This means that the difference in detecting HIV-1 compartmentalization in this thesis vs previous studies can be explained by the depth of sequencing achieved here and that there is no evidence that the dynamics of the viral populations studied in this thesis were different from those previously studied.

In addition, the most recent common ancestor of semen variants was mostly located in blood, indicating the male genital tract was seeded by incoming variants from blood. Clonal amplification was also observed in all the 11 study participants and it was a characteristic of variants from blood and the male genital tract and its frequency did not obviously correlate with the severity of compartmentalization.

In sum, blood and male genital tract HIV-1 compartmentalization and clonal amplification is present in most or all HIV-1 infected males but was not detected in all individuals in previous studies when using techniques with lower depth of sampling.

Objective 2: Sensitivities of blood and semen variants to entry inhibitors and variable loop characteristics

Viral variants from the most compartmentalized donors had variations in sensitivities to entry inhibitors; although the direction of the difference was inconsistent. Donor SVB043 who had the most severely compartmentalized viral populations between blood and semen, had semen viruses that were 1.67 (95%CI 1.08 – 2.56) times more *resistant* to maraviroc (p=0.024) while SVB008 which was the second most compartmentalized donor, had semen isolates that were 4.8 (95%CI 2.76 – 8.28) times more *sensitive* to inhibition by maraviroc ($p < 0.0001$). The meaning of this discrepancy is not entirely clear. It could mean that trait(s) that are selected for in genital tract variants over blood circulation variants are linked to the CCR5 binding region, and that the linked CCR5 genotype was carried along with the selected trait(s). There were no differences in sensitivity to maraviroc between blood and semen clones for donor SVB049 (p=0.847); although this donor on further investigation was found to have functional levels of efavirenz in his blood (3μg/ml, which were within the therapeutic range of 1-4μg/ml) indicating that he was likely on antiretroviral therapy (ART). This was not known to the clinic staff at the clinic at which he was known to receive care and was recruited to this study.

The direction of sensitivities to PSC_RANTES (another CCR5 inhibitor) was concordant to that observed for maraviroc for donors SVB008 and SVB049 but not for donor SVB043 where semen variants were 1.67 (95%CI 1.08 – 2.56) times less sensitive than blood variants to maraviroc, with no detected difference in sensitivity to PSC_RANTES (p = 0.783). This

discrepancy for donor SVB043 probably reflects the difference in mode of action between Maraviroc and PSC_RANTES. The change in envelope conformation over movement from blood into the genital tract presumably affected the maraviroc binding site and not PSC_RANTES. All the clones from blood and semen for the three most viral compartmentalized donors were resistant to CXCR4 inhibitor suggesting that they were all R5 tropic viruses. There were no differences in sensitivities of blood and semen viruses to fusion inhibitor T-20.

The findings here suggest a changed viral envelope conformation/structure for the viruses in the male genital tract. The discordance suggests that the selected trait over movement of virus from blood into genital tract is linked or close to CCR5 binding site but itself does not involve binding to CCR5 co-receptor.

Differences in length and number of glycosylation sites were found between variants from blood and those from the genital tract but the direction of the difference was also inconsistent. Donor SVB043 who had the most compartmentalized blood and seminal variants had semen variants that had longer and more glycosylated envelopes. Donor SVB008 who had the second most compartmentalized blood and semen variants had no difference in variable loop length, but semen variants were less glycosylated. This therefore shows that selection for some of the previous reported traits of acute viral isolates may have started in the genital tract in a subset of the donors.

Objective 3: Sensitivities of blood and semen variants to autologous and heterologous antibodies

Viral populations compartmentalized in blood and the male genital compartment displayed a range of sensitivities to autologous and heterologous neutralization. Donor SVB043 who had the most compartmentalized viral populations between blood and the genital tract; had semen clones that were 1.75 (95%CI 1.11-2.78) times more sensitive to autologous neutralization compared to blood clones ($p = 0.018$). In contrast, donors SVB008 and SVB049 who exhibited substantial compartmentalization, but to a lesser extent than that found in donor SVB043, showed no differences in sensitivities of blood and semen variants to autologous serum. Neutralization sensitivity to a pool of heterologous subtype-matched sera

revealed no differences in sensitivities between clones from blood and semen for donors SVB043 and SVB008. Interpretation of results from donor SVB049 are clouded by the donor's ART use. Overall, these results suggest that, in some individuals, a shift in selective milieu of the genital tract virus occurs. This is presumably due to partial or complete shielding of the genital tract tissue from circulating antibodies, and this shielding shape the populations of HIV-1 variants available for transmission from some but not all individuals.

Overall conclusions

Our data add to the existing knowledge of existence of distinct viral populations between blood and the male genital tract of chronically HIV-1 infected donors. Importantly, and for the first time, we present evidence that HIV-1 compartmentalization between blood and the male genital tract is present in most or all donors, and that some clones are severely compartmentalized even in donors who exhibit very mild compartmentalization. It appears that viral compartmentalization and clonal amplification in these anatomical sites may be present in most individuals but remained undetected in some individuals in previous studies due to the lower depth of sampling applied.

We observed a discordance in entry inhibitor sensitivities and variable loop characteristics between blood vs semen variants among different donors. This may suggest a changed envelope conformation over importation of virus from blood into the genital tract. This change seems to be near or linked to the co-receptor binding site but does not appear to directly involve the co-receptor binding tested in this thesis. This interpretation also may explain the discordance in viral characteristics for the virus establishing infection reported in other studies. This thesis also shows that some of these traits of the transmitted/founder virus relating to neutralization sensitivity, sensitivity to entry inhibitors and variable loop characteristics may originate in and/or be enhanced by transition through the genital tract on the way to becoming a founder virus.

These results are important in understanding how the populations in the genital compartment are selected, giving rise to the population of HIV-1 that is available for transmission to a new individual. An understanding of the dynamics of HIV-1 populations prior to and during transmission is important for vaccine design and other intervention strategies.

This work has been presented at the following meetings/conferences/workshops: (reverse chronological order)

Kariuki SM, Selhorst P, Abrahams M-R, Rebe K, Arien KK, Martin P, Williamson C, Dorfman JR. **Compartmentalization and Clonal Amplification of HIV-1 from the Male Genital Tract Identified using Next Generation Sequencing**. Keystone Symposia on HIV vaccines, Canada, 2018. (Poster presentation)

Kariuki SM, Selhorst P, Rebe K, Arien KK, Martin P, Williamson C, Dorfman JR. **Evaluating Compartmentalization of HIV-1 between male genital tract and blood using next generation sequencing.** International Society for Computational Biology - Africa (ISCB-Africa) conference on Bioinformatics, Uganda, 2017 (Poster presentation)

Kariuki SM, Selhorst P, Abrahams M-R, Rebe K, Arien KK, Martin P, Williamson C, Dorfman JR. **Evaluating Compartmentalization of HIV-1 between male genital tract and blood using next generation sequencing**. 6th conference of South African Immunology Society, South Africa, 2017 (Oral presentation)

Kariuki SM, Selhorst P, Rebe K, Arien KK, Dorfman JR. **In Vitro manipulations of HIV-1 from semen**. International Centre for Genetic Engineering and Biotechnology (ICGEB) summer school, South Africa, 2015 (Poster presentation)

List of abbreviations

AIDS	Acquired Immunodeficiency virus
ART	Antiretroviral therapy
CCR5	Chemokine receptor type 5
cDNA	Complementary DNA
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CXCR4	Chemokine receptor type 4
DEAE	Diethylaminoethyl
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleic tri-phosphate
EC_{50}	Effective concentration 50: inhibitor concentration that produces half maximal response
EDTA	Ethylenediamine tetraacetic acid
env	HIV-1 envelope gene
Env	HIV-1 envelope protein
FBS	Fetal bovine serum
HEK	Human embryonic kidney
HIV-1	Human Immunodeficiency virus type 1
IC_{50}	antibody concentration able to reduce infection by 50%
ID_{50}	Inhibitory dilution of serum or blood plasma able to reduce infection by 50%
IMC	Infectious molecular clone
IFN	Interferon
IgG	Immunoglobulin G
IQR	interquartile range
LTR	Long terminal repeat
Min	Minute
ml	Mililitre
MPER	Membrane proximal region
ng	Nanogram
NGS	Next generation sequencing
NICD	National Institute of Communicable Diseases
NIH	National Institute of Health
NSC	Non seminal cells
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEP	Post-exposure prophylaxis
PNGS	Potential N-Linked glycosylation site
PreP	pre-exposure prephylaxis
RLU	Relative light units
RT	Reverse transcriptase
RNA	Ribonucleic acid
SGA	Single genome amplification
SIV	Simian immunodeficiency virus

STI	Sexually transmitted infection
T/F	Transmitted/founder virus
VL	Viral load
μg	Microgram
μl	Microlitre

List of figures

List of tables

Declaration

I, **Samuel Mundia Kariuki,** do declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be presented for another degree in this university or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever

Signature: **Signed by the student**

Date: 18th October 2019

Acknowledgment

The work in this thesis would not have been possible without the wonderful people who participated in one way or another to bring the success.

Frank Onyambu, you were the one who forwarded me the ICGEB pre-doctoral fellowship call via my Facebook inbox. Thank you for opening the door for me.

To my Supervisor **Jeffrey Dorfman**, there are many ways I am grateful to you. Mostly for holding my hand all the through my PhD journey, for providing sound scientific advice, for believing in me, for your teaching and for challenging me to become a better scientist. I am amazed by your vast knowledge of science. I remember you came to pick me at Kenilworth on my first day at UCT. **Prof. Carolyn Williamson**, my co-supervisor. Thank you for picking me up midway and providing the conducive environment for completion of this work.

Prof. Kevin Ariën, thank you for introducing me to HIV transmission biology and being there every step of the way from masters to PhD. **Dr. Philippe Selhorst**, Thank you for all the meaningful discussions we had concerning this project, from inception to completion. Your thoroughness always amazes me. **Dr. Melissa Abraham**, thank you for being a 'cushion', balance and the best trainer. I am forever grateful. My lab colleagues, **Dr. Thandeka Moyo**, **Fatma Guled, Dr. Marcel Tongo, Dr. Michelle Fisher** and **Zarinah Sonday**. Thank you for everything and for the moments beside lab work.

I would also like to thank the whole **HIV diversity group** headed by Prof. **Carolyn Williamson**. You have been part of me from the start. I used your facility for the bulk of the work. Your training, motivation, guidance and advice is appreciated. You provided an audience for presentation of my work. I particularly want to single out **Nonku** and **Jinny** for training me the SGA technique, **Dave** and **Colin** for help with bioinformatics part of the project.

I wish also to thank Prof. Darren Martin for the advice concerning Bayesian phylogenetic analysis.

Special appreciation goes to the staff of ANOVA's Ivan Tom's clinic, Cape Town for their dedication in study participant recruitment and sample collection. I particularly want to single out **Dr. Kevin Rebe**, **Dr. Johan Hugo**, **Ms Melphy Bebe** and **Ms Lois Suka**.

I am greaful to all the study participants who willingly donated samples for this project.

Much gratitude goes to the organizations that funded this project either in student support or research funds. **International Centre for Genetic Engineering and Biotechnology (ICGEB),** I am grateful for the pre-doctoral fellowship and for the research funds, **National Research Foundation (NRF)** for the research funds and **Poliomyelitis Research Foundation (PRF)** for the research funds and student bursary.

Last but not least, **Betty**, **Alan** and **Mike**, without you, this thesis would not have been possible.

1 CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 General Introduction to HIV

1.1.1 Background

Despite the concerted effort to manage HIV-1, it continues to be a devastating disease and one of the leading contributors to life years lost due to disability and premature deaths [1]. Since the beginning of the epidemic about 80 million people have been infected and about 35 million have died from AIDS-related complications [2]. The disease continues to infect more people each year and in 2017 (the present statistics at the time of writing this thesis) 1.8 million became infected and around 1 million people died of AIDS-related illnesses [2].

Presently there is increased use of anti-retroviral therapy (ART) that has markedly reduced the number of people dying from AIDS-related causes [3, 4]. The number increased from 8 million in 2010 to about 22 million in 2017 which is a significant progress [2]. Despite these impressive results, HIV-1 severity in sub-Saharan Africa where the mode of transmission is primarily heterosexual [5] remains high and accounts for 66% of all transmissions [2]. The good news is that nearly all the countries in sub-Saharan Africa have adopted the test and treat strategy [6] and with the help of a broad range of stakeholders more lives will be saved.

There is no HIV-1 vaccine yet and the only vaccine tested to date with positive efficacy (RV144) had only a modest efficacy of 31%, with the lower confidence interval extending to only 3% protection. The vaccine is being improved and is now being tested in a predominantly subtype C – infected population in South Africa under HVTN 702 phase III trial (NCT02968849).

These challenges means that more work is needed to be done to reduce the HIV-1 infections. Understanding the HIV-1 transmission biology may help in design an effective vaccine and other intervention strategies.

1.1.2 Discovery of HIV/AIDS

The reports of symptoms of a debilitating disease which later came to be referred to as acquired immunodeficiency syndrome (AIDS) started around 1981 when the first patients with AIDS were identified in USA among men who have sex with men (MSM) [7]. It was later

found in intravenous drug users (IVDU), haemophiliacs [8], refugees, and Haitians [9]. It appeared as a long-lasting disease with an unusually long lag phase before the symptoms manifested. The early signs were weight-loss, fever, diarrhoea and lymphadenopathy. Later signs included opportunistic infections such as *Pneumocystis carinii* pneumonia, *Mycobacterium avium intracellular* and cytomegalovirus (CMV) [7, 10-14]. The profound state of immune suppression made the patients present with these opportunistic infections. Little was known of retroviruses that had clinical effects upon humans. Importantly, assays for the presences of HIV-1 were poorly developed and of low sensitivity. Additionally, in the beginning, AIDS itself was defined by symptoms, leading to initial scepticism that it was caused by a single infectious organism, or even if it had an infectious cause at all [15]. Many other factors were proposed as causes; withholding of antiretroviral treatment after discovery of antiretroviral drugs; on the basis of these arguments ultimately lead to death of hundreds of thousands of people, particularly in South Africa [16].

It was particularly notable that there was a decrease in a particular subset of T cells that expressed CD4 surface antigen [17, 18], which had earlier been discovered by Milstein and Kohler using specific monoclonal antibodies [19]. This gave evidence that the agent concerned must be targeting the $CD4^+$ T cells. At this time some laboratory techniques of isolating and growing lymphocytes in culture had been in place. In addition, some scientists at National Institute of Health (NIH) had discovered the T-cell growth factor (interleukin-2) which sufficiently helped increase time for growth of lymphocytes to allow expression of putative latent retroviruses. These techniques helped the few labs that were working with retroviruses at the time. Among these labs were those of Robert Gallo and Luc Montagnier. The two scientists discovered the human T-cell leukaemia virus (HTLV) type 1 and type 2 [20].

Luc Montagnier's lab is also credited with the discovery of HIV-1 as well [21], and one-half of the 2008 Nobel Prize in Physiology/Medicine was shared between Prof Barré Sinoussi and Prof Luc Montagnier for their discovery of HIV [22]. Human Immunodeficiency Virus and HTLV caused symptoms that were overlapping, sometimes leading to confusion [23]. For instance, some of the animal models for HTLV-1 and HTLV-2 not only showed leukaemia-like symptoms and lymphoma but also weight loss conditions which resembled those of AIDS. In addition, HTLV was transmitted through contaminated blood, sex and mother to child which were

consistent with the early epidemiology of AIDS. In fact, some of the early patients studied, were infected by both HTLV and HIV-1 which complicated interpretation further [13, 24].

The evidence that AIDS was caused by a virus come from Centres for Disease Control (CDC) reports with haemophilia patients who presented with AIDS-like symptoms after receiving only filtered clotting factors [25]. The filtration eliminated other microorganisms that were larger than a virus. Subsequently, the first clear isolate of HIV was done at the Pasteur Institute in Paris in early 1983 [21]. The virus for the first time showed clear difference from HTLV in morphology and antigenicity. It was obtained from a patient presenting with lymphadenopathy which at the time was seen as a precursor to AIDS [21]. The lymph node as a site of discovery of HIV is not surprising since it is well known today that the virus thrives there during the asymptomatic period of infection [26].

Working with an early transmitted virus (the virus was isolated from lymph nodes and it is known today that they are infected quite early after transmission) seemed problematic since it could only grow in fresh lymphocytes and not in permanent cell lines [26]. This delayed full characterization of the virus. In addition, the virus at this early stage of discovery was thought to be causative since in acutely-infected patients the immunosuppression was only mild. Breakthrough only came when the virus was isolated from a patient with Kaposi's sarcoma in Paris which could now grow in permanent cell lines possibly because it came from a chronic infection. The long incubation periods before physical symptoms complicated the characterization of HIV-1 unlike other acute viruses whose symptoms manifest almost immediately. In addition, during symptomatic phase other opportunistic infections set in making it exceptionally very hard to link the causative agent to the disease. Later on, blood tests that were sufficiently sensitive were made which could now be used to directly link HIV to AIDS [27, 28]. With the advancement of laboratory techniques it was later possible to show that HIV-1 infection depended on certain chemokine receptors [29, 30]. Moreover, the surprising efficacy of the drugs against the reverse transcriptase enzyme of the virus all contributed as an indirect evidence of HIV being a causative agent of AIDS [31, 32]. These diagnostic tools opened the avenue for development of antiretroviral drugs and strong policies aimed at preventing the acquisition of HIV [33].

1.1.3 Classification of HIV-1 and HIV-2

Human Immunodeficiency Virus (HIV) is classified into two species, HIV-1 and HIV-2 according to differences in nucleotide sequences [34, 35]. In addition to the two types, HIV-1 is further classified into various groups: M (Main/major), O (Outlier), N (non-M, non-O), and P (Putative) [36]. HIV-1 group M (HIV-1M) is further subdivided into 9 subtypes annotated with letters A-D, F-H, J, K and 98 circulating recombinant forms (CRFs) so far (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html). The CRFs are presumed to be a result of recombination between subtypes. It must be observed in at least three epidemiologically unlinked subjects, otherwise it is not presumed to have circulated, in which case it is called a unique recombinant form (URF) [37]. The CRFs are given names comprising of a number which is serialised according to the order of discovery and two letters indicating subtypes presumed to be present in the mosaic e.g. CRF01_AE (previously subtype E [38]) was the first and comprises of a mosaic of subtype A and E [37]. If more subtypes are presumed to be involved in making the mosaic, cpx (complex) is used to denote that CRF [37]. A recent study has shown that some of the sequence tracts within the CRFs might have been derived from older unclassified parental lineages rather than from the current classified HIV-1M subtypes [39]. These parental lineages may have been less infectious at the time to spread to many subjects or they may not have been present at strategic points during the spread of the HIV-1 to global epidemic [40]. Indeed these findings are consistent with the high number of subtypes and high degree of genetic diversity observed among the subtypes coming from Congo basin, the epicentre of HIV epidemic [41, 42].

HIV-2 is inferred to have resulted from at least eight transmissions from sooty mangabeys (Cercocebus atys), a non-human primate found in west Africa [43]. These make up the 8 groups of HIV-2 (A-H) and each is inferred to represent an independent cross-species monkey-to-human transmission event. The epidemic has remained restricted in West Africa, with Mali, Guinea Bissau, Gambia, Mauritania, Nigeria, Sierra Leone, Cape Verde and Ivory coast having more than 1% of the general population infected . It is less virulent, less transmittable and usually characterized with lower viral loads compared to HIV-1 [44-47].

1.1.4 Origin of HIV/AIDS

Human immunodeficiency virus type 1 (HIV-1) has its roots in central Africa and it is thought that the avalanche of factors including railway transport (Figure 1) constructed by Belgian colonial rule, urban growth and change in sex lifestyles in Kinshasa fueled the spread of the virus to the global scale [48]. By the time HIV/AIDS was being discovered in the US, it had already been well established in central [49] and eastern [50] Africa. Indeed, the oldest HIV-infected human samples to have been discovered were both from central Africa; that of 1959 African plasma (ZR59) that was found to have been HIV-1 seropositive [51] and a Bouin's-fixed paraffin-embedded lymph node biopsy specimen [52]. These two early samples come from Léopoldville, Belgian Congo (now Kinshasa, Democratic Republic of Congo (DRC)) [52].

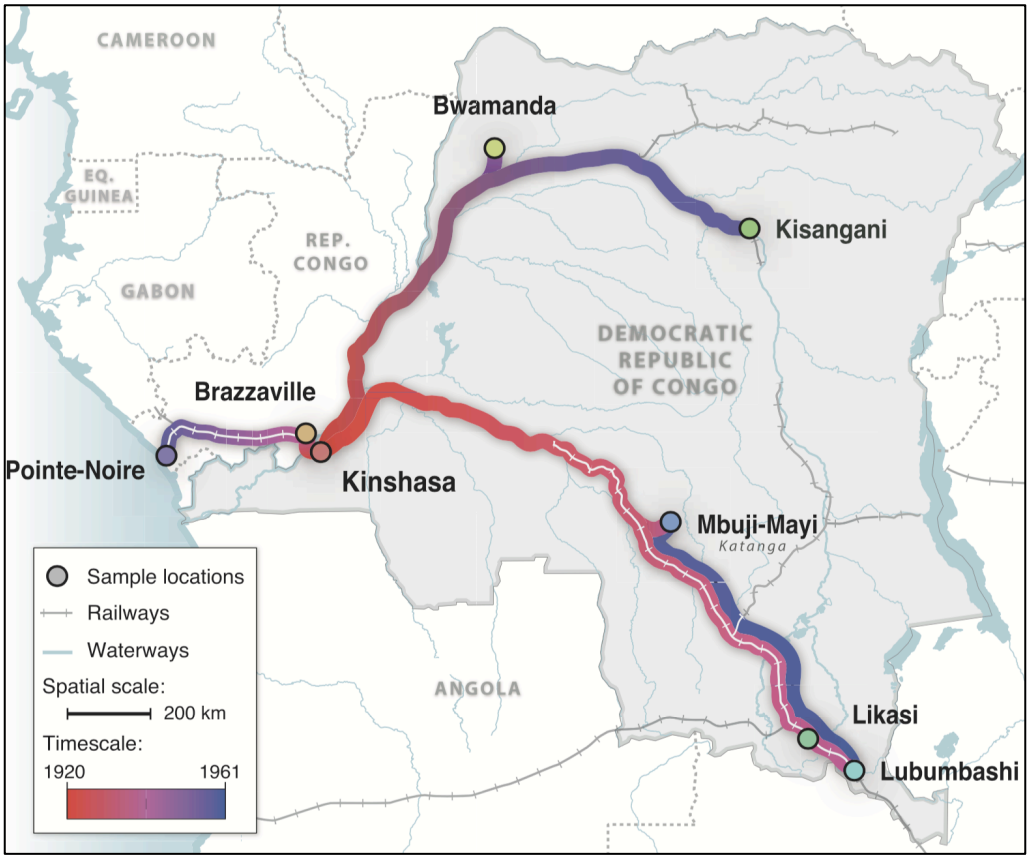


Figure 1: Factors that lead to exportation of HIV-1 group M from Congo basin include transportation network (railway and waterways). This map adopted from [48] shows the spatial movement of HIV-1 group M from Kinshasa following these transportation networks with gradient colours indicating the time scale of the movements.

The relatives of HIV-1 and HIV-2 are found in non-human primates (NHP) who have an equivalent family of viruses called Simian Immunodeficiency Virus (SIV) (Figure 2). HIV is thought to have resulted from multiple zoonotic introductions into the human population

from these relatives possibly due to bush meat practice [53] and keeping NHP as pets [54]. Bites and contact with blood and other bodily fluids from the NHP during hunting and slaughter is the most plausible explanation for the cross-species transfer.

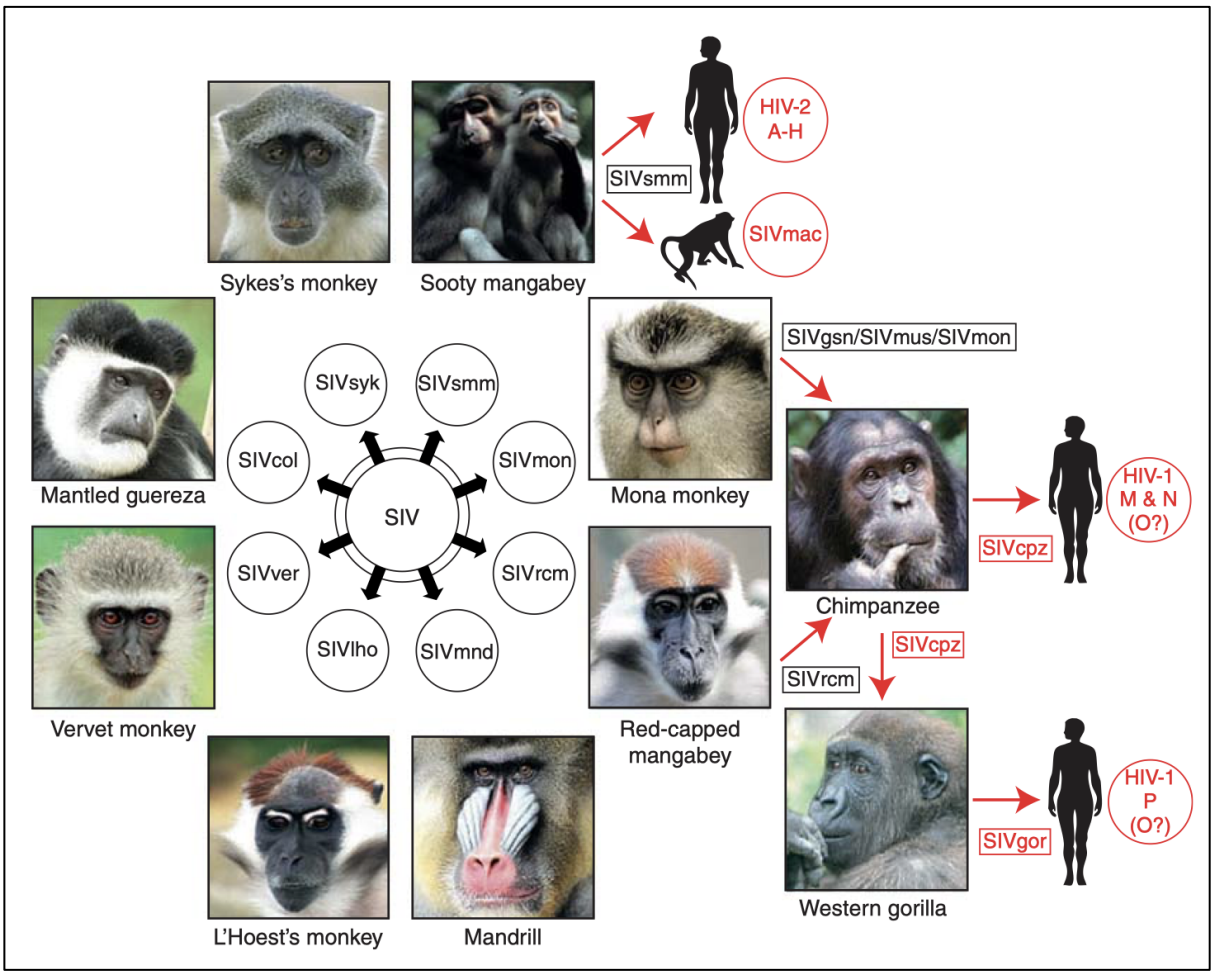


Figure 2: HIV-1 and 2 involve a simian origin. These old world monkeys are naturally infected by simian immunodeficiency viruses (SIVs) which have crossed the species barrier to infect great apes and humans. (adapted from [55])

Each of the four groups of HIV-1 involve separate cross-species transfer from a particular NHP (Figure 3). HIV-1 group M (HIV-1M) has close SIV relatives in chimpanzees [56]. The SIV species that is closer to HIV-1M is SIVcpz isolated from chimpanzees *Pan troglodytes troglodytes* [53]. Through sequencing of *gag* and *env* genes of HIV-1, it was possible to identify that the last common ancestor of HIV-1M existed around 1930 [57].

Conversely, the HIV-1 group O (HIV-1O) has a SIV relative in *Gorilla gorilla* [58, 59]. Typical variants of HIV-1O (ANT70, MVP5180 and VAU) were first discovered in 1990s from

Cameroonian natives residing in Europe [60-62]. Today, HIV-1O is less widespread and is localized in Cameroon where it has infected one percent of the population. It is thought that HIV-1M and HIV-1O had similar epidemic histories in the 1960s, but HIV-1M managed to cause the global epidemic [40].

The HIV-1 group N (HIV-1N) and HIV-1 group P (HIV-1P) are rare than HIV-1M and HIV-1O. The HIV-1N, (YFB30) was first discovered in 1998 from a Cameroonian woman who had died of AIDS three years earlier [63]. Her serum reacted to the envelope antigen of SIV isolated from chimpanzee and not antigens from HIV-1M and HIV-1O leading to classification of a new HIV-1 group. The typical variant of HIV-1P (RBF168) was discovered in yet another Cameroonian national arriving in France [36]. There are only three cases of HIV-1P so far discovered in the world [64].

HIV-2 on the other hand has a closest SIV relative called SIVsm from sooty mangabeys whose geographic regions still overlap in West Africa [55, 65].

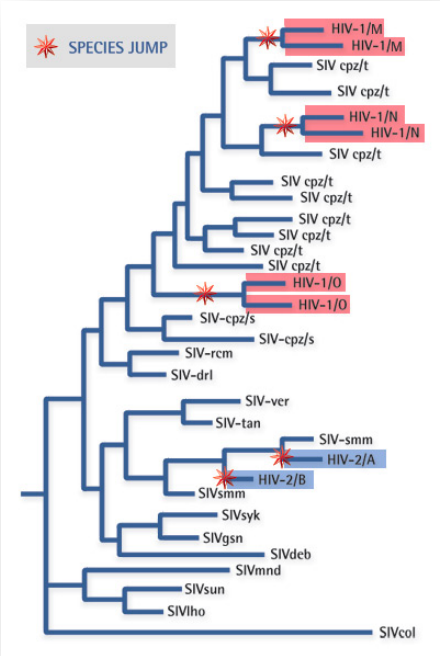


Figure 3: Phylogenetic tree indicating the origin of the two types of human immunodeficiency virus (HIV-1 and HIV-2). Photo obtained from https://www.pbs.org/wgbh/pages/frontline/aids/virus/tree.html

1.1.5 Epidemiology of HIV-1 group M

Since the discovery of HIV-1 at around 1980, approximately 78 million people have been infected by the disease and 35 million so far succumbed to AIDS [66]. According to the UNAIDS estimates of 2018 [66], approximately 36.9 million people in all age groups were living with HIV in 2017 (lower bound estimate, 31.1 million, upper bound estimate, 43.9 million). Among these, 35.1 million were adults and 1.8 million children. Globally, approximately half of the people living with HIV were males and half were females. Overall, currently the number of new infections have been reduced by approximately 18% since 2010 (Figure 4). Sub-Saharan Africa is the most affected contributing to approximately 66% of all infections among people in all ages, and more women being infected (56%) than men. This notwithstanding, more men living with HIV die more than women mainly attributable to higher diagnosis rate and higher treatment coverage among women than men [67].

With the advent of Highly Active Antiretroviral Treatment (HAART), there has been a steady decline in AIDS-related deaths globally with less than a million dying per year (approximately 34% decline since 2010) [66]. This reduction in number of deaths since 2010 have been witnessed in many regions with the greatest decline in eastern and southern Africa (42% reduction). Other regions with sustained decline were; Asia and Pacific (39% reduction), western and central Europe and north America (36% reduction), western and central Africa (24% reduction) and Caribbean (23% reduction). Latin America had the lowest decline in the number of death (12% reduction), but this is mainly due to their long term programme of antiretroviral therapy. On the other hand, eastern Europe and central Asia had no reductions in AIDS-related deaths since 2010 while the number actually increased by 11% in North Africa and Middle East [66]. The production and use of opiate drugs in these two regions, which has led to increased HIV-1 infection rates within vulnerable groups, has been suggested to be the cause as well as gaps in the treatment cascade [66, 68]. In addition, lack of HIV-1 surveillance systems, little access to HIV-care facilities as well as the geopolitical situation of the region have contributed to deaths from HIV-1 related complications [69].

The UNAIDS 2018 estimates [66] shows that around 22 million people are currently on treatment which is a net improvement from 2.3 million people since 2016. There has also

been an 18% overall decline in new infections since 2010 [66]. Again the greatest decline was achieved in sub-Saharan Africa with eastern and southern Africa showing a 30% decline. Other regions also achieved reduction, Caribbean (18% decline), Asia and Pacific (14% decline), western and central Africa (8% decline) and western and central Europe and north America region having a 1% decline. Middle East and North Africa and eastern Europe and central Asia, the incidence has gone up instead [66]. Men account for most of the new HIV infections globally but in sub-Saharan Africa, more adult women account for new HIV infections than men (59%) [66]. The fact that the decline in number of AIDS-related deaths does not correspond with a similar decline in new HIV-1 infections means that the number of people living with HIV-1 will continue to grow and more resources are needed to keep them on treatment.

An ambitious concept have been coined, branded “90-90-90” – with a target goal of year 2020. This concept intends that by 2020, 90% of the people living with HIV will know their status, 90% of the diagnosed cases will be put on therapy and 90% of those on therapy will have viral suppression. This idea is part of the United Nation’s programme on HIV/AIDS ambitious goal to achieve a 90% reduction in new HIV infections and death by 2030 based on 2010 baselines. It is based on “test and treat” principal in that if people can know their HIV status early enough, they can go on treatment immediately which in turn prevent onward transmission of the virus. This will have a net effect of reduction of new cases of HIV at the population level.

With the current estimates of 36.9 million people living with HIV [66], it means 33.2 million will need need to be diagnosed, 29.9 be put on therapy and 26.9 achieve viral suppression to meet the 90-90-90 goals globally. Current estimates [70] show that 75% of the people living with HIV knew their status in 2017, of which 79% (59% of all people living with HIV) were accessing treatment and among these accessing treatment 81% (47% of all people living with HIV) were virally suppressed [2]. Around 9 million people still remain undiagnosed and among the diagnosed, 5.8 million have not yet initiated treatment. Around 20% of the people on treatment are not achieving viral suppression. This will have an effect on the onward transmission of HIV and risk of development of drug resistant mutations thereby affecting the choice of future treatments.

Key populations at risk of HIV needs a special focus for an effective HIV eradication to be achieved. These special groups include men who have sex with men (MSM), intravenous drug users (IVDU), sex workers, transgender communities, prisoners among others. More recently immigrants have also been considered as part of the key HIV population with some of the people migrating due to persecution of their sexual identity (e.g. being gay or MSM) [71]. These groups have been shown to disproportionately concentrate the sub-epidemic of HIV in areas where overall prevalence within the general population is low [72]. In fact, a survey showed that new global HIV infections in 2017, were related to key populations and their sexual partners [73] with the epidemic driven by these groups in some regions than others. More than 75% of the new HIV infections in eastern and central Asia, middle east and north Africa, western and central Europe and north America, Latin America, Asia and the pacific, and the Caribbean were driven by these key populations and their sexual partners [66]. Notably, in 2017, intravenous drug users were the main drivers of the epidemic in eastern Europe and central Asia and in the Middle East and north Africa. In these two regions the HIV incidence has doubled in the last 20 years [66, 71]. In eastern and southern Africa, 16% percent of the new HIV infections were also driven by the key populations [70]. Therefore targeting these groups for diagnosis and treatment may be an important step in HIV eradication.

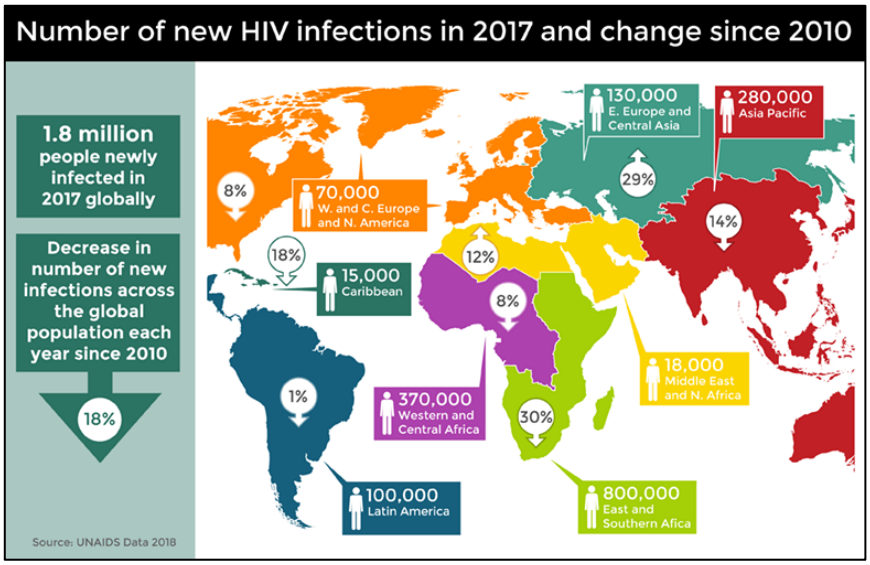


Figure 4: The new HIV in 2017 and the percentage change since 2010. Photo obtained from UNAIDS via https://www.avert.org/

1.1.6 Distribution of subtypes of HIV-1 group M

HIV-1 group M subtypes and CRFs are distributed in diverse geographical regions of the world. Subtype C is the most abundant representing close to 50% of all global infections [74]. It is followed by subtype A at 18%, B at 11%, CRF02_AG at 8%, CRF01_AE at 5%, G at 5% and D at 2% [74] (Figure 5). The rest of the subtypes have infected only a minority (<1%) of the population. All the CRFs and URFs represent a proportion of 20% of all the HIV-1 infections in the world [74]. The differential distribution of HIV-1 globally is as a result of founder effects, travel and migrations [75-77]. Southern African countries are almost entirely dominated by HIV-1 subtype C virus infecting countries like South Africa, Zimbabwe, Botswana, Zambia, Malawi and Swaziland. Subtype C also dominates in Burundi and Ethiopia which form part of Eastern Africa. In the other Eastern African countries; Tanzania, Rwanda and Kenya, subtype A predominates although C and D and their recombinants are also available. West African countries are dominated by CRF02_AG representing 50-80% of all infections, although CRF06_cpx represent up to 50% of the infections in Nigeria, Niger, Burkina Faso and Togo. North America and Western Europe countries are dominated by subtype B while in Eastern Europe subtype A as well as CRF03_AB dominate especially among the IVDUs [74]. In India, subtype C dominates although some regions are affected by subtype A as well. South American countries are dominated by subtype B and F while in Southeast Asia subtype B and CRF01_AE are prevalent. Oceania is dominated by subtype C while in Australia subtype B dominates. In China, CRF01_AE, CRF07_BC and CRF08_BC predominate [74].

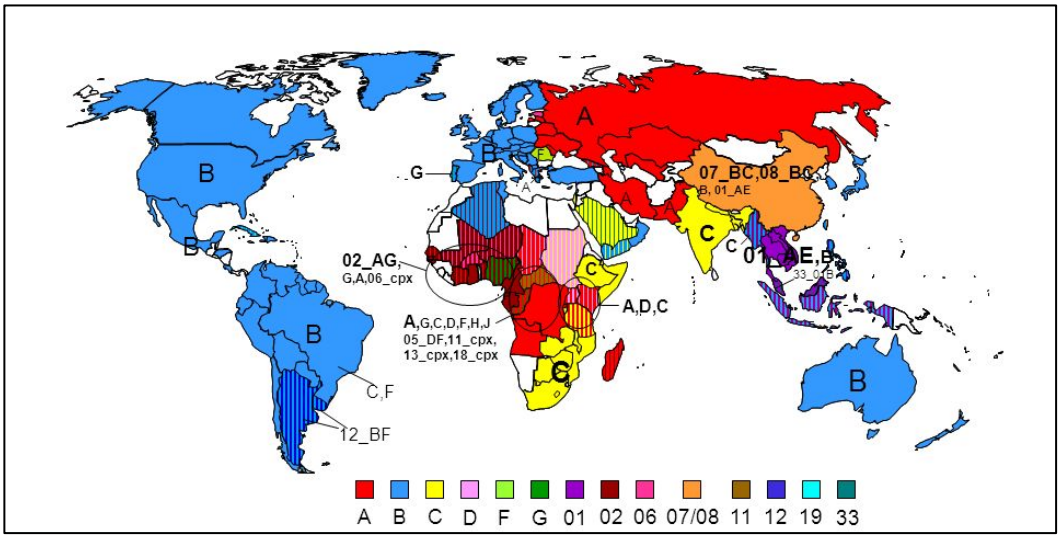


Figure 5: Geographic distribution of HIV-1 group M (Major) subtypes and CRFs. Figure adapted from WHO-UNAIDS HIV Initiative

1.1.7 Pathogenesis

The natural course of infection of HIV-1 is composed of three phases, acute, asymptomatic and symptomatic phase [78] (Figure 6) .

1.1.7.1 Acute phase of HIV-1 infection

The acute phase of infection starts with a single variant infecting a small subset of $CD4^+$ T cells, macrophages and dendritic cells at the portal of entry and lasts between 3-6 weeks [79, 80]. HIV uses a combination of pathways for establishments at the portal of entry some of which include transcytosis, endocytosis and capture by mannose C-type lectins receptors such as DC-SIGN found on the surface of dendritic cells and macrophages [81]. The initial amplification happens at the draining lymph nodes before the variants enter the blood stream and secondary amplification happens at the gastrointestinal tract [26, 82]. The infected $CD4^+$ T cells express a homing receptor called $\alpha4\beta7$ which is known to facilitate the migration of infected cells into the gut-associated lymphoid tissue (GALT) [83, 84]. The $CD4^+$ T cells are massively depleted in the GALT which also leads to accumulation of high viral loads in blood [85].

This is followed in approximately half of the infected individuals by some flu-like symptoms (fatigue, night sweats, fever, chills, headache, muscle aches, sore throat, diarrhoea, joint aches, swollen lymph nodes, skin rash, etc) [86]. This condition is also known as primary HIV infection or acute retroviral syndrome [87]. This is also the most contagious phase of the infection since most individual are unaware of their infection status and the initial indicators are usually ignored or confused with other mild infections [88]. The symptoms resolve on their own in a few weeks. This is the phase where peak viremia is observed at 10^6 - 10^7 viral RNA copies per millilitre (cp/ml) of blood which is one or two orders of magnitude higher than what is observed later during the chronic stage of infection [89, 90]. This phase is also accompanied by low prevalence of $CD4^+$ T cells and there are usually no detectable antibodies meaning antibody-based tests are not able to detect the virus at this stage [89].

1.1.7.2 Asymptomatic phase of HIV-1 infection

During the asymptomatic phase, viral loads in plasma decline to a set point which comes to characterize chronic infection [91]. This reduction is mainly attributed to HIV-1 specific $CD8^+$ cytotoxic cells which eliminate virally-infected target cells [92, 93]. In addition, antibodies against the virus start to build up and can now be detected in plasma [89]. The combined effort of humoral and cellular responses leads to partial recovery of $CD4^+$ T-cells and their numbers are seen to slightly increase [91]. Despite the absence of symptoms, viral replication is still ongoing. The ongoing replication and regeneration of the $CD4^+$ T-cells establishes a dynamic equilibrium hence a steady state viremia . The length of this phase varies; 1 – 2 years for fast progressors while slow progressors can take several years [78, 94].

1.1.7.3 Symptomatic phase of HIV-1 infection

Over time due to imbalanced replication of the virus and decreased regeneration of the $CD4^+$ T-cells, a gradual decrease in the latter is observed if the patient is not on antiretroviral drugs [94]. This leads to the symptomatic phase of the infection [95]. At first mild symptoms are experienced with severity increasing with further reduction in $CD4^+$ T-cells. This eventually leads to loss of immune function and an increase in viral replication at which Acquired Immune Deficiency Syndrome (AIDS) sets in. Opportunistic infections and rare cancers characterises this stage which eventually leads to fatalities [7].

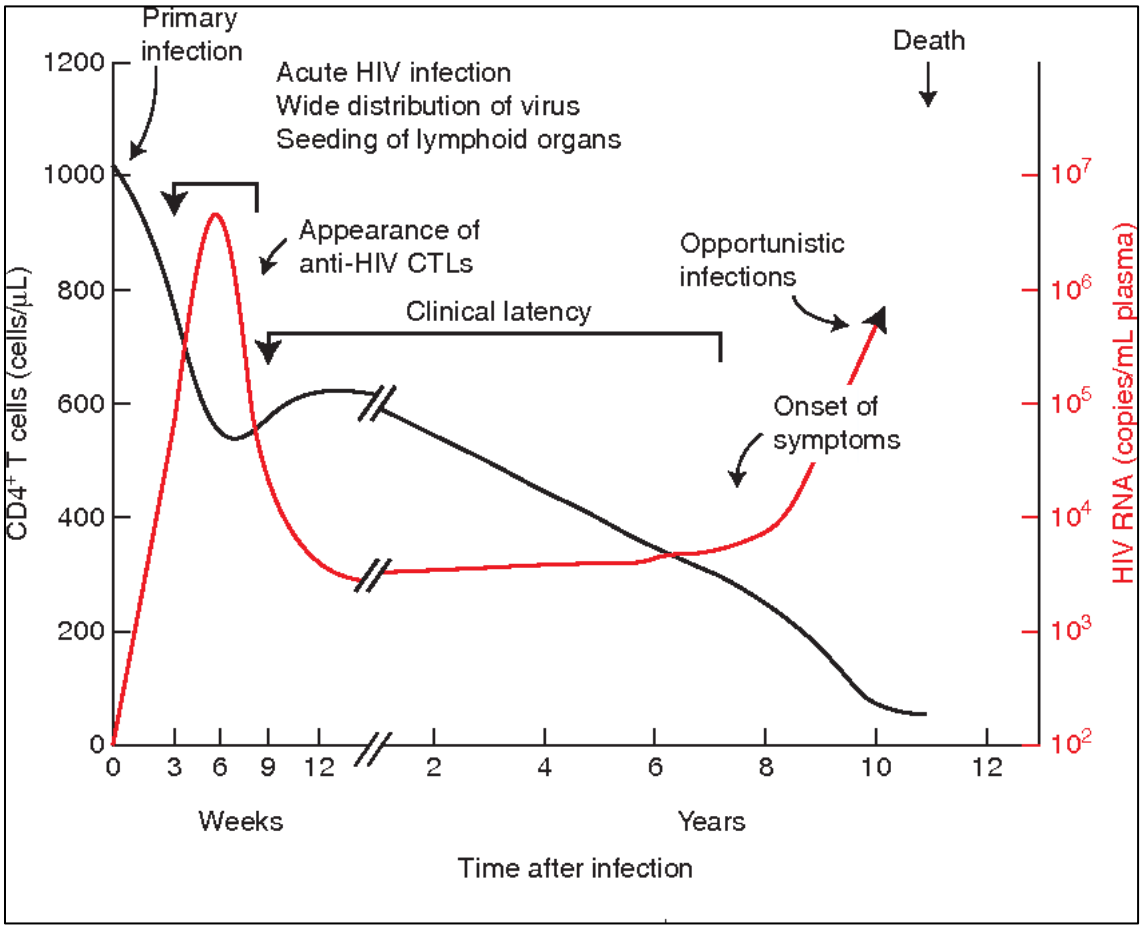


Figure 6: The dynamics of CD4+ cells and viral loads in a therapy naive HIV-1 infected subject over time. HIV pathogenesis sis characterized by four phases; primary infection, dissemination of virions in the body organs, clinical latency and symptomatic phase. (Image from Fauci and Desrosiers 1997 used with permission from Cold Spring Harbor Laboratory Press.)

1.1.8 Virological characteristics of HIV-1

1.1.8.1 The HIV-1 particle

The HIV-1 virion has been studied extensively using the electron microscopy and consists of an envelope, a core and matrix (Figure 7), [96, 97]. The envelope is derived from the cell membrane of the infected cell and encompasses the matrix and the capsid [95] (Figure 7). The envelope is studded with Env glycoprotein projections (gp120) that are used for attachment onto the host target cell during fusion and entry. These glycoproteins are also an important target for the host immune system. The capsid on the other hand houses two identical positive sense RNA strands complexed with p7 nucleocapsid protein, several enzymes necessary for establishment in an infected cell as well as some accessory proteins (Figure 7).

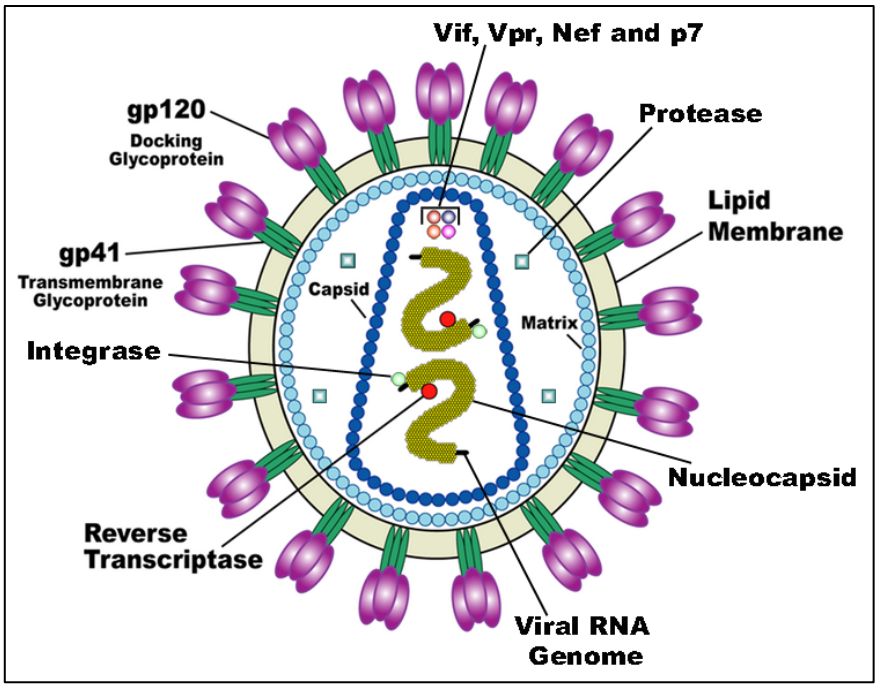


Figure 7: The structure of the HIV-1 particle. The virion capsid is encompassed by the envelope. The capsid in turn houses the viral genome as well as enzymes that are necessary for establishment in an infected cell. The figure was obtained from https://commons.wikimedia.org

1.1.8.2 HIV-1 genome organization and roles

The HIV-1 genome is ~9.8 kilobases [98] and encodes nine genes; *gag*, *pol*, *vif*, *vpr*, *vpu*, *tat*, *env*, *rev*, *nef*) (Figure 8). The Long Terminal Repeats or LTR are regions flanking the viral genome in its DNA form and they contain transcription initiation and polyadenylation regulatory regions. The genes are organized into structural (gag and env), regulatory (tat, rev) and accessory genes (vif, vpu, vpr, nef).

The structural proteins are first encoded into polyproteins that are then cleaved by protease to produce functional genes. Gag is the product of the HIV genome gag that encodes the capsid proteins (group-specific antigens). The precursor is the p55, a 55kDA myristoylated [99] protein that associates with the plasma membrane during the viral assembly. It (gag) is cleaved by viral protease into p7 (Nucleocapsid), p17 (Matrix), p24 (capsid) and p6 proteins [100]. The matrix particles are usually associated with the inner lipid bilayers of the viral envelope. In some particles however, it is also found deeper in the virion and participate in transport of the viral DNA into the host cell nucleus [101]. p24 forms the conical core of the

virus and is found associated with cycophilin A [102]. p6 mediates the interaction of p55 gag and the vpr accessory protein, aiding its inclusion into the virus particle [103]. It also helps efficient release of budding virus particles from host cells.

The HIV-1 pol is another structural gene and encodes the viral enzymes, reverse transcriptase (RT), RNase H, integrase (IN), and protease (PR). The precursor is produced as a Gag-Pol [104] by frame shifting of the ribosome on the 3´end of gag [105]. HIV-1 protease occurs as a dimer to cleave the Gag and Gag-Pol polyproteins during the viral maturation. The RT has both RNA- and DNA-dependent DNA polymerase activities and forms the double-stranded DNA from the dimer of RNA present in the virion within the first 6 hours of infection. The RT polymerase does not have proofreading capacity and therefore introduces point mutations in the viral genome during each replication cycle at a rate of about 1 substitution for each 3 genomes synthesized. Integrase is a 34kDa protein consisting of three domains as analysed through X-ray crystallography or NMR; the N-terminal domain, which chelates zinc, the core domain responsible the catalytic function and the C-terminal domain that binds DNA non-specifically. Integrase regulates the insertion of HIV provirus into the host genome, in complex with several host cell proteins. Integrase has three activities, an exonuclease activity by which it removes two nucleotides from the 3' ends of the provirus, an endonuclease activity by which it cleaves the host DNA at the integration site and a process where the provirus is covalently linked to the ends of host DNA (strand-transfer reaction) [106].

The third structural gene is *env* which encodes the Env protein. Env is the viral glycoprotein produced as a gp160 precursor from *env* gene then processed into two non-covalently bound proteins: a trans membrane part gp41 and an external part called gp120 (Figure 7). These non-covalently bound proteins form a trimer [107] on the viral envelope. The gp120 interacts with the CD4 [108] and chemokine receptors [109] on the surface of host cells during the infection process. The selective pressure of the antibody-mediated immune system of the host against the virus manifests against the Env [110]. As part of the adaptation of the virus, there is a co-receptor switch from CCR5 to CXCR4, which occur in 50% of the HIV patients infected with HIV sub-type B [111]. This switch, determined by the V3 region of the *env* gene is characterized with increased depletion of T-lymphocytes and rapid development of AIDS [110].

The HIV-1 genome has two regulatory genes *tat* and *rev* which encode the Tat and Rev proteinds respectively. Tat is the transactivator of HIV expression [112], one of the two important viral factors for the regulation of HIV gene expression. Tat binds to the transactivation response (TAR) in the RNA to activate elongation and as such prevents the 5´LTR poly A signal from premature termination of transcription [113-115]. Short transcripts of around 100 nucleotides are produced in the absence of Tat. They occur persistently at low levels in nucleus or nucleolus of infected cells in two forms, the minor and the major forms. The minor form is a 72 amino acid protein also called Tat-1 exon while the major form is an 86 amino acid protein also called the Tat-2 exon.

Rev is the other important regulatory factor of HIV expression after Tat. Rev is a sequence-specific phosphoprotein that helps in nuclear export and maintenance of the viral mRNA containing the Rev response elements (RRE) regions [116]. Rev binds to a bubble formed from non-Watson-Crick G-G base pairs [117] and transports unspliced or incompletely spliced mRNA from the nucleus to the cytoplasm which would ordinarily be retained in the nucleus. It is a 19kDA protein that binds to RRE, and like Tat, it is localized in the nucleus or nucleolus and mediates the transition from early to late phases of HIV gene expression. Rev contains three domains [118], arginine-rich region that binds RRE, multimerization domain [119] and effector domain with a nuclear export signal (NES) [120].

The HIV-1 genome also contains several accessory genes that encodes the accessory proteins Vif, Vpu, Vpr and Nef. Vif is the Viral Infectivity Factor, which exists as a soluble cytosolic protein. This 23-kDA protein is important for viral replication in that it inhibits antiviral activity of a group of cellular enzymes, APOBEC (a cytidine deaminase that mutates viral nucleic acids) by targeting it for ubiquitination and cellular degradation. Most cell lines are permissive to Vif mutants of HIV and although the virus produced is able to infect non-permissive cell lines, the virus produced from the latter are non-infectious.

Vpr is the Viral Protein R, a 14kDA protein incorporated in virions [121] by interacting with the carboxyl terminal, p6 of Gag [122]. Several roles have been proposed for this protein, which include, nuclear transport of pre-integrating complexes, which facilitates infection of

non-dividing cells [123]. This is achieved by tethering of the viral genome to the nuclear pore [124] facilitating nucleocytoplasmic transport. Other roles of this protein include the transactivation of cellular genes, the induction of cellular differentiation and cell growth arrest [125]. Vpu is the Viral Protein U, which is a type 1 integral membrane protein [126] that functions in degradation of CD4 in the endoplasmic reticulum and enhancement of release of the virion from the plasma membrane of the infected cells [127]. This 24kDA protein has been found to antagonize host cells through binding to host proteins such as CD4 and targeting them for degradation through recruiting of cellular multi-subunit ubiquitin ligases. Additionally, it has also been seen to make ion channels on host cell membranes [128]. A new target has been found to be an interferon induced trans membrane protein Tetherin/BST2 which allows retention of the nascent virions on the cell membrane and thereby restricting viral replication [128].

Nef (an acronym of negative factor, which is an early error of describing its activity) is a myristoylated protein usually found associated with the plasma membrane using the myristoyl portion although cytoplasmic variants have also been observed [129]. When found in the nucleus, it is usually associated with the cytoskeleton. Nef is required in the maintenance of high viral loads and individuals with defective Nef emerge to be long time survivors. This 27kDA protein does this by down regulation of MHC-I reducing the killing of the infected cell by cytotoxic T cells [130] and down regulation of CD4 receptors by endocytosis and lysosomal degradation [131, 132]. Vpx is a 12kDA protein found in HIV-2 and some SIVs but not in HIV-1. It is a homolog of *vpr* gene in HIV-1. Its role is not yet fully elucidated but like Vpr it interacts with p6 of Gag.

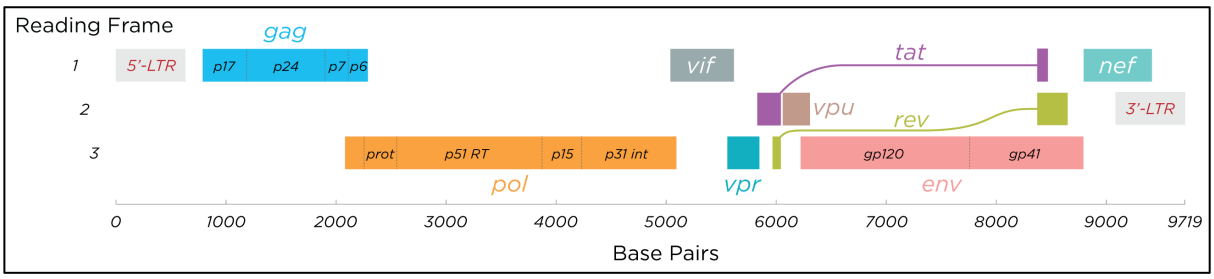


Figure 8: HIV-1 genome organization. Source: https://commons.wikimedia.org/

1.1.8.3 The replication cycle

HIV-1 lifecycle involves a series of seven steps; binding, fusion, reverse transcription into double-stranded DNA, integration into the host genome, viral RNA synthesis and synthesis of the proteins, assembly and budding off infected cells (Figure 9). HIV-1 infection lifecycle starts when the virus attaches and enters the target cells [133]. HIV-1 needs a high density of CD4 to infect cells and therefore, the main target cells are the $CD4^+$ T cells [134]. HIV-1 has a trimeric GP120 on the envelope of the virus which initiates infection by binding to the CD4 receptor on the surface of HIV target cells [135]. The binding of the viral glycoprotein leads to a conformational change of the glycoprotein hence exposing additional binding site for the host cell coreceptors [133]. HIV-1 can use either the CCR5 or the CXCR4 coreceptors available on the surface of the target cells. Based on the co-receptor utilization therefore HIV-1 can be classified as R5, X4 or R5/X4 dual-tropic viruses, utilizing CCR5, CXCR4 or both respectively [136]. CCR5 coreceptor utilization has been shown to be a characteristic of primary infecting viruses and also during chronic disease while CXCR4 are mainly observed during late infections [137]. This may be due to severe depletion of CCR5 target cells resulting from limited number of $CD4^+$ T cells hence the macrophage-tropic phenotype [138].

The fusion step begins after the co-receptor binding which leads to a second conformational change allowing the gp41 N-terminal fusion peptide to penetrate the cell membrane of the target cell [139]. On coreceptor binding, fusion of the HIV envelope and the CD4-containing cell membrane occurs therefore allowing the virus capsid core to enter the $CD4^+$ cell. This is followed by reverse transcription of the dimeric viral RNA into double stranded DNA using the viral reverse transcriptase enzyme. During the reverse transcription process, RNAseH portion of the reverse transcriptase is used to degrade the RNA template strand from the RNA-DNA hybrid. The high diversity of the variants in an HIV-infected individual is attributed to the error-prone nature of the reverse transcriptase. In addition, the process of switching templates between two virion RNA strands results in recombination which is apparent when two species of RNA from different proviruses invade the same cell [140]. The DNA is complexed with other viral and cellular proteins to form what is called the pre-integration complex (PIC) [141]. The PIC enters the cell nucleus using the nuclear pore thereby starting the integration stage of the lifecycle. The ability of HIV-1 to transport its genetic material via

an intact nuclear envelope makes it stand out from other retroviruses which rely on disintegration of the nuclear membrane during cell division [142].

In the nucleus, the double stranded viral DNA is integrated into the host genome using the viral integrase enzyme [106]. The viral genome that is integrated may either be transcriptionally active or remain dormant within the host cell . HIV Latent reservoirs are composed of transcriptionally inactive proviral genomes but they may be transcriptionally active once host factors become upregulated. The Trans-Activator of Transcription (Tat) protein recruits the transcriptional factors to the HIV LTR region to begin the transcription process. Full-length mRNA primary transcripts and several spliced mRNAs encoding the various viral envelope and accessory proteins are produced. The transcripts are then translated into gag, pol, vpr, and vif and together with the full length mRNA transcript are assembled at the plasma membrane where they are packaged in viral particles. The assembled virus particles then bud off cholesterol-rich islands on the plasma membrane [143]. The cholesterol enrichment is thought to assist in the fusion of the virus envelope with the plasma membrane of the host cells during entry [143].

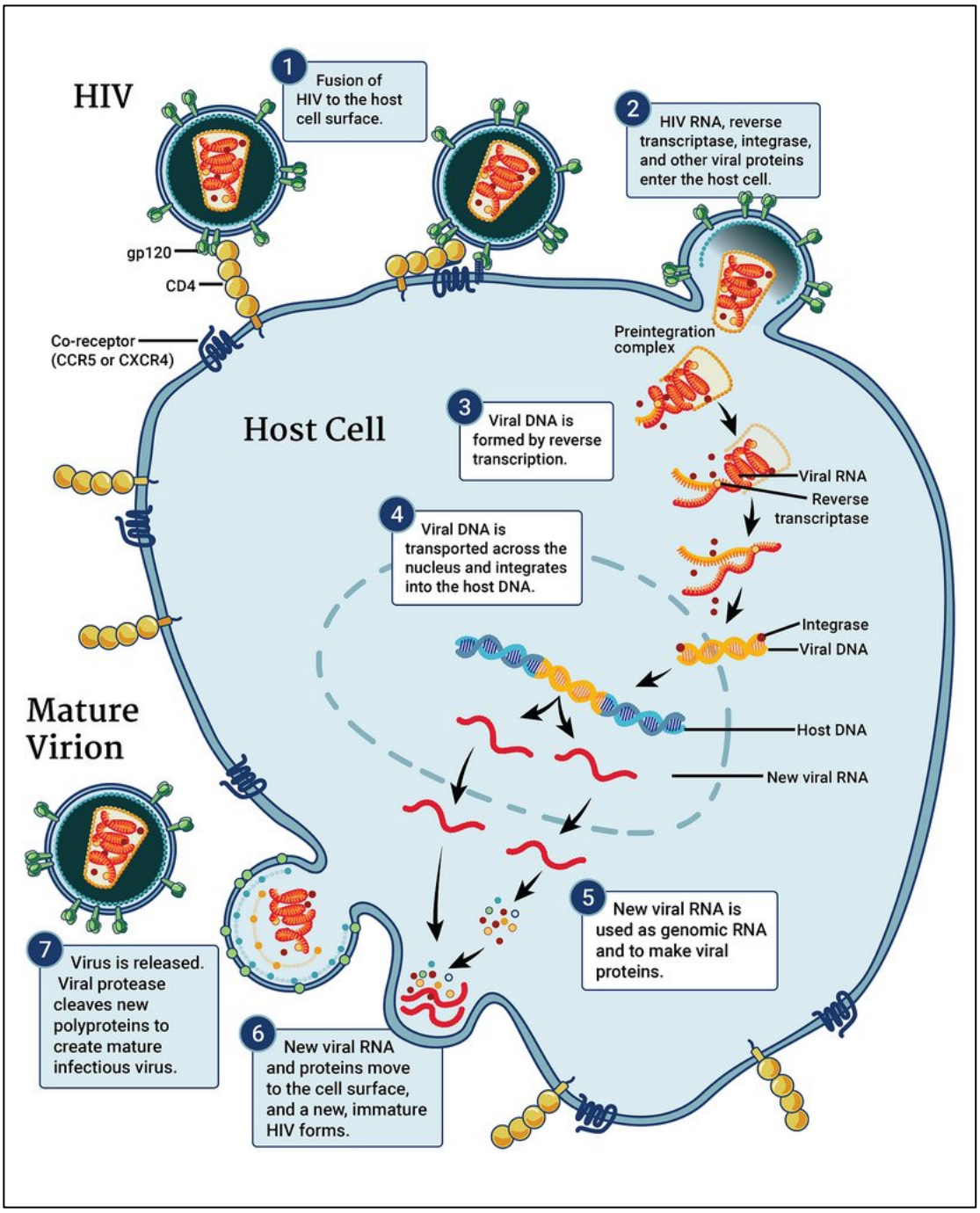


Figure 9: The lifecycle of HIV-1. Photo obtained from https://www.niaid.nih.gov/

1.1.9 The HIV-1 sexual transmission Biology

1.1.9.1 HIV-1 infection routes and risks

HIV-1 infection is acquired via several routes; sexual (male and female genital tracts), intestinal tract, placenta and bloodstream (contaminated blood and needles). Current

UNAIDS statistics show a reduction in HIV incidences although the decline is different among the risk groups [70]. Below are the brief summary of these routes.

a. Sexual transmission

Sexual transmission contributes to more than 80% of all the HIV-1 transmission globally [144]. Within this route, the risk of infections varies based on the directionality i.e. female to male transmission harbour a risk of exposure of 1 in 700-3000 contacts, male to female 1 in 200-2000 and male to male transmission 1 in 10-1600 while Fellatio has a risk of 1 in 16 [145]. The virus that gets transmitted comes from semen, blood, rectal and cervicovaginal secretions.

b. Bloodstream and contaminated equipment

Transmission of HIV-1 also occurs through blood to blood contact with contaminated blood and needles. This route may occur in different ways with varying probabilities of infection per exposure event: blood transfusion (95 in 100), needle sharing (1 in 150) or needle stick injuries (1 in 200) [146]. In some parts of the world, the HIV-1 epidemic has traditionally been driven via needlestick injury by the key population of people who inject drugs (PWID) [66]. Some of these regions include eastern Europe and central Asia (EECA) [71, 147]. Today, cases of acquisition of HIV-1 through receiving contaminated blood during transfusion are rare due to proper screening.

c. Placenta and intestinal tract

The HIV-1 is transmitted from an infected mother to her baby though the placenta (vertical transmission) or through the intestinal tract during breastfeeding . This is also referred to as vertical transmission or mother-to-child transmission (MTCT) [148]. It is possible for HIV-positive mothers to give birth to HIV-negative babies [149]. Without treatment of the mother, the risk of vertical transmission is at 1 in 4 births, which falls to <1 in 10 if the mother is on treatment [146]. This is achieved through treatment of the mother during gestation, labour and delivery and continuation of treatment after delivery [150]. In addition, administration of neonatal and infant prophylaxis after delivery must also be done [149]. Neonatal prophylaxis acts as a pre-exposure prophylaxis so that the baby is not infected during breastfeeding and post-exposure prophylaxis in case the baby got infected during delivery, labour or late gestation period [151]. This is however complicated in a resource poor settings

especially in sub Saharan Africa where sometimes the mothers don't know their HIV-status, the delivery is done at home and infant survival depends on long term breastfeeding [152, 153]. The WHO has a target of eliminating MTCT through its recommendation of adoption of Option B+ guideline of 2010/2012 which requires ART initiation for pregnant or breastfeeding mothers regardless of their $CD4^+$T cell counts or clinical stage of disease, and continuation of treatment for life [154].

The transmission of HIV-1 can also happen through the intestinal tract (rectum and upper gastrointestinal tract) in adults [155, 156]. The transmission medium here is usually semen or blood and involves anal intercourse which can occur between men or between men and women [157-159].

1.1.9.2 Female genital tract HIV-1 acquisition

The majority of the HIV acquisitions in females happen through vaginal intercourse, semen being the main carrier [160]. Vaginal intercourse carries lower probability of successful infection compared to anal intercourse or parenteral inoculation [146]. Various organs of the female genital tract can be infected by HIV-1 (Figure 10). Using macaque models it has been shown that the labia, vagina, ectocervix, endocervix, uterus and ovaries are all infected by SIV [161, 162]. The relative contribution of each of these tissues has not yet been fully elucidated. The virus from semen overcomes the huddles in the female genital tract related to presence of mucus, the female genital tract epithelial layers, defensive proteins and antibodies among others [163], collectively called the 'mucosal barrier' [164].

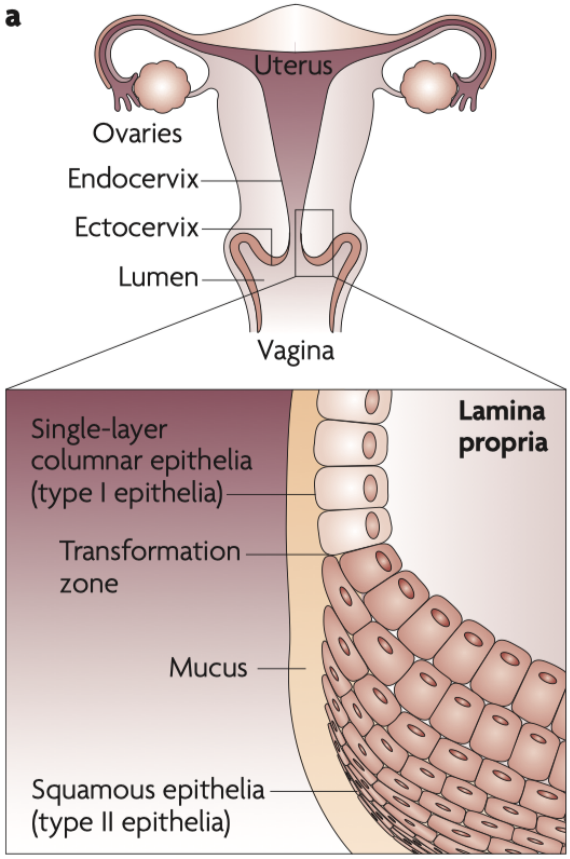


Figure 10: The HIV-1 invasion sites in the female genital tract. Photo adapted from [163]

The lower female genital tract is composed of the vagina and the ectocervix (Figure 10). The ectocervix and the vagina provide a large surface area which the virus can potentially exploit for entry [163]. The fluid/exudate produced by the lower female genital tract covers the stratified vaginal epithelium and therefore provides the first barrier of protection preventing dissemination of the virus [165].

Beneath the exudate is the multi-layered squamous epithelium of the vagina and ectocervix which provides a mechanical protective barrier against HIV-1 and other viruses (Figure 11). Even though this barrier is thought to be protective, the virus is still capable of bypassing it [166]. HIV-1 is thought to be captured here by Langerhans cells and translocated into deeper tissues [167, 168]. Other HIV-1 target cells available here are the $CD4^+$ T cells and dendritic cells [134, 167]. When there is trauma on these surfaces, the acquisition of HIV-1 is even more probable. Trauma during intercourse is more common in women [169, 170] although there is

no clear definition of what really constitutes traumatic sex. This is partly believed to be the reason why more and more women seroconvert more than men [170]. Vaginal atrophy especially in post-menopausal women and those that use some vaginal tightening remedies especially in some Africa cultures [171, 172] exposes women to vaginal trauma that HIV-1 can exploit for entry. HIV-1 can be also be acquired solely through the vagina as evidenced in women who have undergone hysterectomy [173, 174].

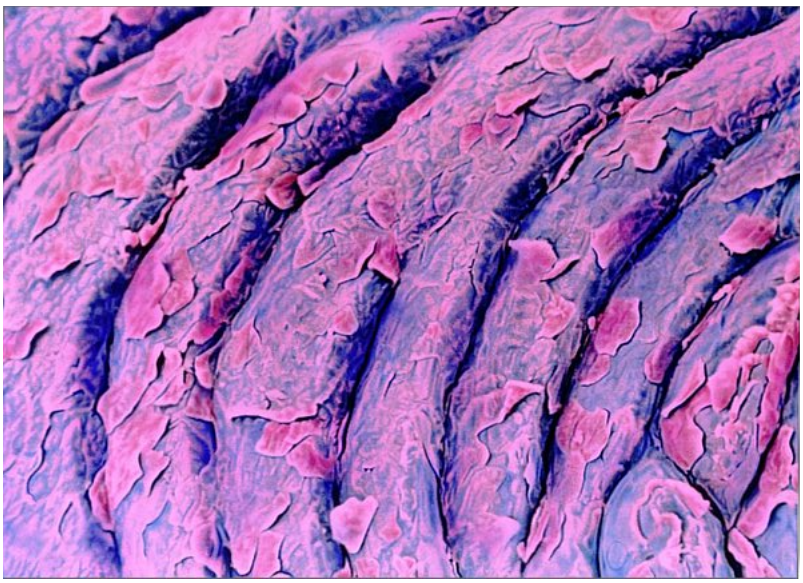


Figure 11: Scanning electron microscopy (SEM) photo of the vaginal epithelium showing the flattened and loosely attached epithelial cells. Photo adapted from https://fineartamerica.com

The higher part of the female genital tract is comprised of the endocervix, fallopian tubes and ovaries. The endocervix has a single-layer columnar epithelium which is more vulnerable to HIV-1 infection due to its relative fragility [175]. It is also covered by mucus as previously mentioned whose main composition is the mucin which comes from two genes MUC4 and MUC5B [165]. Mucin in mucus is known to inhibit HIV infection [165]. In addition, the mucus also traps the viruses and cells thereby slowing its diffusion and therefore minimizing its infectivity [176, 177].

The transformation zone between the endocervix and the ectocervix has been documented as the main portal of entry of HIV-1 [175]. This site is enriched with more HIV-1 target cells and therefore uniquely susceptible to viral entry [167, 175]. Besides, it is thought to be the region that is most susceptible to trauma during intercourse.

The endocervix mucosal surfaces have been thought to affect the R5 and X4 HIV-1 differently [178, 179]. It is thought to favour more the R5-tropic viruses than X4-tropic ones [180]. The cationic change on the V3 loop of X4 viruses is more than that of the R5 viruses [181]. This in turn would mean that X4 viruses would be more attracted to the polyanionic cervical mucin [182] and also to heparin sulphate proteoglycans that are found on the cervical mucosal surfaces [183]. This in principle would favour R5 viruses and act as a sink for the X4 viruses. Indeed, using flow cytometry techniques with dissociated cells, it has been shown that cervical tissue is rich in CCR5 expressing $CD4^+$ lymphocytes [184]. This corroborates with earlier findings that showed the endocervix stromal papillae has more CCR5 mRNA compared to CXCR4 [185]. During the menstral cycle, a window of vulnerability exists post-ovulation when the endocervix defence is low thereby increasing susceptibility to HIV-1 infection [186, 187].

1.1.9.3 Male genital tract and HIV-1 in semen

Semen contains fluid secretions from the male genital tract and the cellular fraction. The cellular fraction is composed of, spermatozoa, leukocytes, macrophages and some epithelial cells generally referred to as non-sperm cells [188]. Leukocytes include $CD4^+$, $CD8^+$, granulocytes and some B cells [189, 190]. The seminal plasma contains the fluid matrix produced by the epididymis, testes, prostrate, seminal vesicles, distal and proximal bulbourethral glands [191]. The largest fraction of the fluid however comes from the seminal vesicles and prostrate [191]. It also contains high amount of fructose and enzymes responsible for coagulation and dissolution of coagulation as well as citric acid, lipids and phosphates [192].

Majority of HIV-1 acquisition in men result from vaginal intercourse with an infected female. Semen is also the main medium that carries the virus in transmissions involving men who have sex with men and in male-to-female transmission; making semen the single most contributor to transmission of HIV-1 [166]. The product of probability of transmission, duration of semen infectiousness and the number of sexual partners determine the basic reproductive number (R_0) of male-associated forward transmission [193].

The HIV-1 target cells in the male genital tract are more abundant in the foreskin. The outer foreskin is more keratinized than the inner one (Figure 12) and more HIV-1 target cells are found in the inner foreskin than the outer foreskin [163]. These cells include Langerhans cells and $CD4^+$ T cells on the lining and macrophages, T cells and dendritic cells (DC) in the underlaying stroma [194, 195]. The dendritic cells contain dendritic cell - specific C-type lectin (DC-SIGN) which capture the HIV-1 envelope allowing efficient infection of $CD4^+$ cells in trans [196]. The vulnerability of the foreskin to HIV-1 infection is the basis for the protective effect after circumcision [197]. Circumcision however does not provide total protection against acquisition of the virus suggestion that other target sites are important too. The glans penis has squamous epithelium that is heavily keratinized like the outer foreskin making virus entry here impossible. On the other hand, the urethra is made up of columnar non-keratinized epithelium and has high concentration of $CD4^+$ $CD8^+$ T cells in the epithelium and underlaying lamina propria [198] which could be infected by HIV-1 and SIV [199].

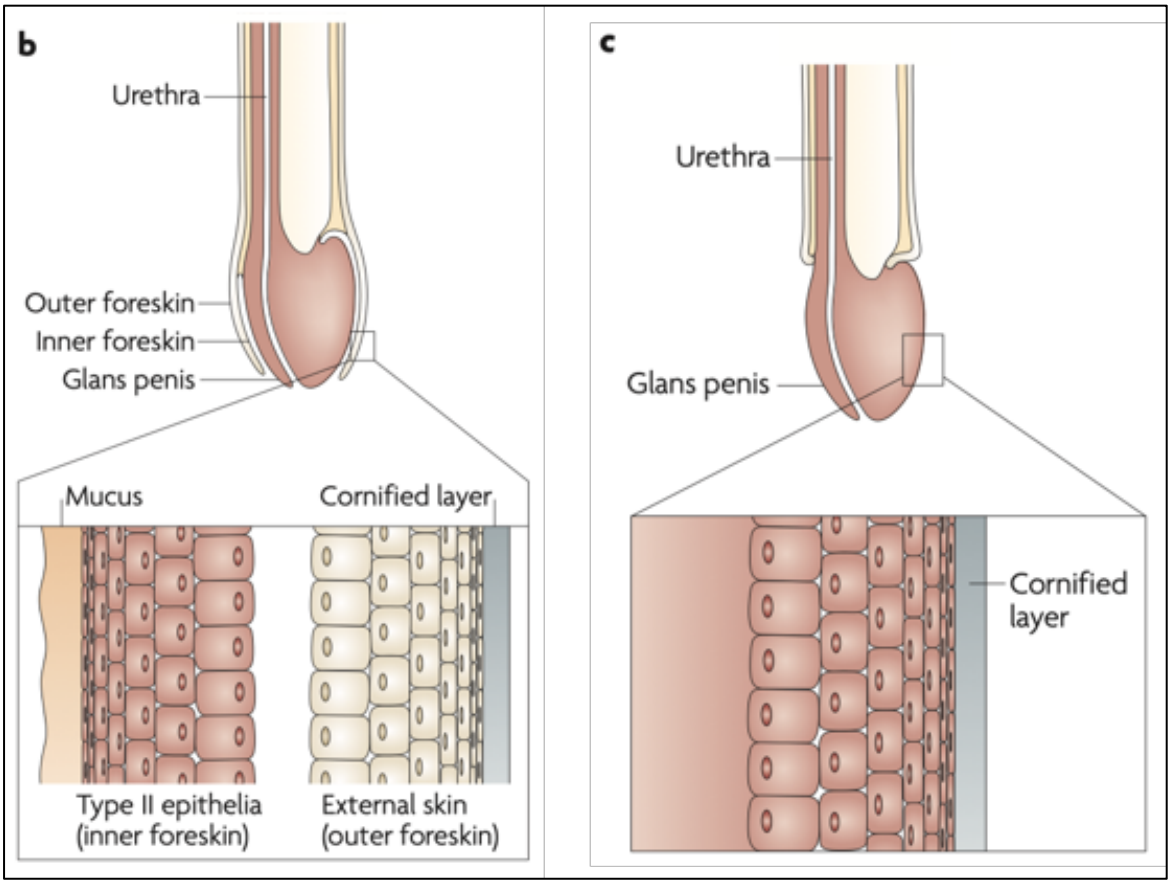


Figure 12: The HIV-1 invasion sites in the male genital tract. Photo adapted from [163]

1.1.10 Crossing selective barriers

The mucosal surfaces of the recipient's gastrointestinal, female and male genital tracts are natural barriers that deter the virus from establishing and therefore most exposures do not lead to a productive infection [200-203]. The openings of these areas which also are the point of contact to the external environment are composed of stratified epithelia which is a multilayer of epithelial cells which offer the first mechanical protection against invading pathogens [163]. Besides being the physical barriers, the mucosal surfaces also produce substances that actively antagonizes the virus. For example production of mucins [204], innate and adaptive immunity factors [200, 205, 206]. Further protection is provided by the lower numbers of HIV-1-infected $CD4^+$ T cells compared to the general circulation [207].

On the other hand the rest of the mucosal surfaces are covered by a simple epithelia made up of a single layer of polarized epithelial cells with tight junctions between them [163]. This type of barrier only provides limited protection against infection. HIV-1 therefore exploits the physiological process of the mucosal surfaces including transcytosis to gain entry into the body [208]. At the stratified epithelium, binding and internalization of the virus is facilitated by the DC-SIGN and the Langerhans cells available in these regions where the virus is then directed into the draining lymph nodes [195]. Langerhans cells are professional antigen presenting cells and are usually the first line of encounter with the virus due to their proximity to the site of delivery [209]. In most cases when the viral load is low, the virus is successfully antagonized but if the viral load is high, the protective effect of langerin [210] is diminished and successful infection of T cells happens either at the local site or within the draining lymph nodes [166, 211].

At the male genital tract, the foreskin has been implicated in facilitating the majority of the HIV-1 infections, and circumcision is known to protect men from acquiring infection [212, 213]. In an explant tissue experiment, one study showed that some sections within the outer male genital tract are important in HIV-infection which include the meatus, urethra and glans all of which were found to favour more the R5 virus and block X4 viruses [214]. The presence of HIV-1 target cells at the urethra, foreskin and the rest of the epithelium increase the risk of acquisition of HIV-1. In yet another study with explant tissues, the inner foreskin was shown

to be more involved in HIV-1 acquisition suggesting that keratinization of the outer foreskin provided protection [208].

On the other hand, at the simple epithelium of the endocervix, rectum and gastrointestinal surfaces, the main mechanism of internalization of the virus is through transcytosis [215]. This is where the infected and the epithelial cell come together in a polarized synapse enabling the budding virus to be quickly internalized without the epithelial cell being productively infected.

1.1.11 HIV-1 prevention

The risk of acquisition of HIV-1 from a single sexual exposure is very low [216] but staying with an infected person increases the risk of acquiring the infection due to multiple exposures [217]. In addition, other factors such as infectiousness of the sexual partner, their viral load [218], stage of disease [219] and susceptibility of the partner at risk including their sexually transmitted infections status [145] also increases the risk. Therefore prevention of transmission of HIV-1 involves both behavioural and biomedical strategies.

1.1.11.1 Behavioural strategies

Behavioural strategies include modification of sexual and drug use behaviours in individuals or among social units. Though difficult to evaluate [220] intensive campaigns on behaviour change to reduce transmission of HIV-1 have been credited with reduction of HIV-1 incidences [70]. The approaches involved include skill-building approaches, peer-led campaigns, motivational activities and educational approaches. Success of such intervention strategies were realized in reduction in HIV-1 incidences among men who have sex with men in the USA [221]. Many countries attribute the success in reduction in HIV incidences to effective behavioural changes. Some of the behavioural strategies include condom use, monogamy and reduction of casual sex partners, sexual abstinence and debut delay, HIV counselling and testing, elimination of substance use or harm reduction strategies [222]. Among the countries where change of behaviour has influenced HIV-1 prevention include Kenya, Zimbabwe, Uganda, Malawi, Burkina Faso, Namibia and Swaziland [223].

1.1.11.2 Biomedical strategies

Biomedical strategies include male circumcision, pre- and post- exposure prophylaxis, treating of STIs and vaccines.

a. Male circumcision

Male circumcision is an old tradition in many communities in the world and it is estimated that around 35% of all the adult males in the world are circumcised [224]. In societies where the mode of HIV-1 transmission was heterosexual, male circumcision was negatively correlated with HIV-1 transmission [225]. Overall male circumcision have been reported to have a protective efficacy of between 50% - 88 % [213, 226-230] which is corroborated by a meta-analysis that showed an adjusted relative risk of 0.42 (95% CI 0.34 – 0.54) in all the circumcised males [231]. There are several biological explanation to the HIV-1 protection observed after circumcision: The inner foreskin which is exposed during erection has more HIV-1 target cells than the outer foreskin [230] and nine times more than the cervical tissues [232]. The outer foreskin has proportionately higher keratinization compared to the inner foreskin [195]. The foreskin is also susceptible to ulcerative sexually transmitted infections [233, 234]. Finally the uncircumcised penis is more susceptible to trauma during intercourse than a circumcised one.

b. Treating sexually transmitted infections

Various sexually transmitted infections (STIs) are associated with increased relative risk of acquiring HIV-1 in longitudinal studies: genital herpes, syphilis, gonorrhoea, trichomonas and chlamydia [235-237]. The STIs cause urethritis, genital ulcers or cervicitis and may increase HIV transmission by creating entry points when the disease is ulcerative or increasing the HIV-1 target cells if the disease creates an inflammatory situation [238, 239]. On the other hand the results from randomized clinical trials for treatment of STI for prevention of HIV-1 have reported mixed results [240, 241]. The incomplete understanding of the precise mechanism in action has led to the term ‘epidemiological synergy’ in explaining the role of STIs in HIV-1 transmission [242]. In sum, STIs work to increase the infectiousness of the infected index case, increase the susceptibility of the recipient or both [242, 243].

The presence of an STI has also been associated with increased shedding of the HIV-1 in the genital fluid [244-248] and treating STIs has been shown to reduce the viral shedding [247, 249-251]. Therefore programs to diagnose and treat STIs are presumed to be important in prevention of HIV-1.

c. Pre- and post-exposure prophylaxis (PreP and PEP)

Pre-exposure (PreP) and post-exposure (PEP) are recommended for occupational and non-occupational exposures for people with a high risk of acquiring HIV-1 [252]. Though at present there is no randomized clinical trial for PEP, one case-control study showed 81% protection [253]. Other evidence for efficacy of PEP comes from animal models [254-256] and treatment of pregnant women [257, 258]. On the other hand there have been several randomized clinical trials for PreP done targeting men who have sex with men [259-261], heterosexual transmissions [262-265] and among intravenous drug users [266]. These studies showed between 44% - 86% reduction acquisition of HIV-1.

d. HIV-1 Vaccines

With more than 70 million people and 35 million deaths since the beginning of the epidemic, there is a serious need for a HIV-1 vaccine. The road towards an effective HIV-1 vaccine has been long and winding since the traditional vaccine approaches of whole-inactivated or live-attenuated virus did not work with HIV-1 [267]. Recombinant vaccines on the other hand provide some promise. Recently modest efficacy was observed with RV144 “Thai” vaccine trial in 2009 [268]. This vaccine is being modified in a current vaccine trial called HVTN 702 launched in 2016 among South African Adults. The HVTN 702 vaccine is based on two previous experimental vaccines: a canarypox vector-based vaccine ALVAC-HIV (Sanofi Pasteur) and AIDSVAX two-component gp120 protein subunit vaccine (GlaxoSmithKline) and MF59 adjuvant (GlaxoSmithKline).

Recently, the RV144 vaccine was adapted to subtype C virus which is the highly prevalent clade in South Africa under the HVTN 702 vaccine clinical trial. Fashioning an HIV-1 vaccine for every clade raises questions on licensure, i.e. each region seeking a licence for a HIV-1 vaccine that targets their most prevalent subtype/subtypes. Janssen Phamaceutica vaccines

and their partners are working in developing a HIV-1 vaccine composed of several copies of each gene that captures a significant amount of the diversity of the HIV-1 subtypes. The phase 1/2a trial called APPROACH with volunteers in Thailand, South Africa, Uganda, Rwanda and USA showed safety and tolerability as well as high antibody titres [269]. In addition, it showed 67% protection against SHIV-SF-162P3 in rhesus monkeys [269]. The phase 2b has already been launched (NCT03060629).

In addition to recombinant vector vaccines, other researchers are infusing broadly neutralizing antibodies to test whether they can provide protection. The VRC01 monoclonal antibody as a passive immunotherapy is being tested in sub-Saharan Africa as well as in the USA, Peru and Brazil. The concern here is the rapid decay of antibodies hence repeated infusions will be mandatory to keep antibodies at protective levels. This brings also the issue of cost-effectiveness since it is expensive to produce the antibodies therefore these studies are rather a proof of concept to show that antibodies can offer protection and therefore inform vaccine development.

e. Treatment as prevention of HIV-1 (TasP)

Initially, initiation of HIV-1 treatment was based on the level of CD4 counts [270]. Treatment as prevention (TasP) is a new concept which was derived from the observation that among discordant couples, the infected partner had undetectable or very low concentration of viral RNA [218, 271]. This led to randomized clinical control studies as well as mathematical modelling studies that showed near zero transmissibility if suppressed viral load was achieved [270, 272-274]. This was the basis for revision of the previous recommendations to treat based on $CD4^+$ T cell counts and stage of disease [270] to the current universal 'test and treat' whose guidelines were drafted by World Health Organization (WHO) in 2016 [275]. If successful it will help achieve the UNAIDS Fast-Track target of 90-90-90 by 2020, where 90% of the people who are infected know their status, 90% of them to the receive the treatment and 90% of those on treatment to achieve viral suppression [276].

1.2 HIV-1 compartmentalization

HIV-1 compartmentalization is the restriction of movement of the virus between anatomical sites leading to formation of related yet distinct viral variants within these anatomical sites [277]. In an infected individual HIV-1 exists as quasispecies; which are related yet distinct variants. The diverse population undergoes a founder effect when there is adaptation to a specific microenvironment that leads to the appearance of a compartmentalized population [278, 279]. The diversity results from the error prone-nature of the viral reverse transcriptase (RT) [280-282]. Additional causes of diversity comes from high recombination rates, the short viral replication cycle and the large population size [283]. The genetic diversity varies from one tissue to another thereby impacting on viral fitness and progression of disease [282, 284].

HIV-1 compartmentalization between blood and various anatomical sites have been documented with the most frequent compartmentalization occurring within the CNS (cerebral spinal fluid (CSF) or brain autopsy) compared to virus in the blood [277]. Other anatomical sites showing this phenomenon include the liver, breast, gut, lung [285], kidney and the male and female genital tracts [286-292]. Some of the proposed reasons for the existence of viral compartmentalization within these anatomical sites is the differential immune selective pressures, tissue and cell type differences in gene expression and replication, and other factors that alter the cellular or tissue environment [277]. Some of these alterations of cellular environment may include but not limited to antiretroviral concentrations for the subjects on therapy [293, 294]. These perturbations and selective pressures lead to some mutations that may impact phenotypic properties of the virus [277]. Some of these phenotypic attributes include cellular tropism, coreceptor utilizations, replicative fitness, transmission, neutralization properties, syncytium formation and cytopathic effects. Studying these compartmentalized viral populations enhances our knowledge of HIV-1 pathogenesis enabling designing of targeted remedies to prevent viral adaptation to these microenvironments.

1.2.1 Viral compartmentalization in leukocytes

HIV-1 compartmentalization between peripheral blood mononuclear cells (PBMC) and plasma has been studied mainly using the pol or env sequences. These studies have revealed the existence distinct variants within these two sites [295]. In addition, differential drug resistance mutations between plasma and infected lymphocytes have been described [296]. In one study [297], out of the patients who were failing antiretroviral therapy, a quarter of them were due to drug resistance mutations in PBMC that were lacking in plasma. Compartmentalization has also been reported to exist between specific populations of lymphocytes e.g. monocyte/macrophages and $CD8^+$ T cells [296, 298]. However other studies have also failed to demonstrate this compartmentalization [299, 300].

1.2.2 Viral compartmentalization in the brain and cerebral spinal fluid

The brain and cerebral spinal fluid (CSF) have been reported to show more severe compartmentalization relative to what has been observed in other anatomical sites [301-305]. This is probably due to the existence of a blood-brain barrier (BBB) [306-308] and the blood-CSF barrier (BCSFB) [304, 309]. These barriers are very selective, but seem to not only restrict virus trafficking but also other macromolecules circulating systemically [310]. The HIV-1 infection in the brain and CSF leads to HIV-1-associated neurocognitive disorders (HAND) that occur during the advanced stage of disease [311-313]. HAND refers to a group of neurocognitive impairments including asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND) and HIV-associated dementia (HAD) [314].

The central nervous system is colonized during the primary infection [315] leading to brain injury and inflammatory responses [316, 317]. Even during the era of antiretroviral therapy, a less severe cognitive impairment exists suggesting that some ARVs don't penetrate sufficiently into the BBB and BSCFB [318]. This means that the CNS could act as a latent virus reservoir which could present challenges in its elimination [82].

Compartmentalization of HIV-1 between blood and the central nervous system has been evaluated in previous studies mainly from brain autopsy [319, 320]. In addition, viral compartmentalization, equilibration and clonal amplification in living subjects have been

observed between CSF and blood [305, 321-323]. The compartmentalized viral populations in CSF and CNS were found to exhibit some characteristics that were distinct from the viruses circulating systemically [305, 324-326]. Schnell *et al* reported existence of both R5 T cell tropic and macrophage-tropic viruses circulating in the CSF of patients exhibiting signs of HIV-associated dementia (HAD) [327]. On therapy initiation, the R5 viruses were associated with rapid decay in the CSF while the macrophage-tropic ones were associated with slow decay indicating that the macrophage tropic phenotype arose as an adaptation to grow in cells expressing low levels of CD4 in the CNS [327]. Indeed the team showed that the macrophage phenotype was restricted to CNS/CSF and did not exist in blood plasma of the same subjects. Testing them in affinofile cells which have an inducible CD4 and CCR5 levels, it was confirmed that macrophage tropic viruses were better at infecting cells with low levels of CD4 than their T cell tropic counterparts [328]. Development of X4 and m-tropic variants expands the host target cells when the $CD4^+$, $CCR5^+$ T cells become limiting.

1.2.3 Viral compartmentalization in the gastro-intestinal tract

The Gut-Associated Lymphoid Tissue (GALT) contains the majority of the body's lymphocytes compared to the general circulation [78, 329]. The mucosal $CD4^+$, $CCR5^+$ T cells are the predominant lymphocytes, thereby facilitating the replication of the virus during the first few weeks of infection [83, 330, 331]. Extensive depletion of the $CD4^+$ T cells occurs in the gut, and evidence shows significant loss still persist even after years of therapy [85]. The proximity of the gut to the external environment, and access to gut microbial antigens and food add to the characteristics that make the gut hospitable to replication of HIV-1 [331]. These factors cause the $CD4^+$ T cells to be in a perpetual activated state and also some having differentiated into the memory state. In addition the access to microbial antigen leads to inflammation and injury resulting in HIV-1 stimulatory cytokines [332].

Within the various organs of the gut; stomach, duodenum, oesophagus and colon, differences in viral HIV-1 diversity have been observed [333]. Distinctive viral populations have also been observed in the rectum relative to those in blood [334]. This can be due to founder effect, but it can also indicate that the virus experiences differential selective pressure depending on local selective milieu of the organ is resides in. Further studies analysing the virus in tissues in

autopsies have also come to a similar conclusion of distinct viral populations within tissues in the gut [334, 335]. Interestingly, distinct viral populations have also been reported between HIV-1 in faecal samples and that in blood samples [336-338] further supporting viral compartmentalization between blood and the gut.

Compartmentalization has also been observed in other bodily anatomical sites. These include the breasts [339-342] where also mastitis was associated with increased viral replication [341]. Like with other anatomical sites, not all the studies have reported compartmentalization of HIV-1 between breast milk and blood [340, 342, 343]. HIV-1 genetic differences have been observed between the liver cells and blood in a subset of the donors [344, 345]. Coinfection with HIV have been reported to exacerbate other liver conditions such as hepatitis C, liver cirrhosis and fibrosis [346] hence affecting treatment response and causing premature death. Within the lung increased replication of HIV-1 has been reported in people coinfected with tuberculosis [347, 348], while other studies also showed independent evolution of HIV-1 within the lung compared to the general blood circulation [285, 349]. Other tissues with evidence of HIV-1 compartmentalization include kidneys [350-352], lymph nodes [353, 354] and spleen [355-358].

1.2.4 HIV-1 compartmentalization in the female genital tract

HIV-1 is able to replicate independently in the female genital tract resulting in viral compartmentalization [359]. Initial studies had observed discrepancies in viral loads between the systemic circulation and the female genital tracts [291, 360, 361]. These differences gave initial evidence that the two compartments were distinct and HIV-1 can replicate independently in each of them [292]. In the era of antiretroviral therapy some scientists were also quick to notice that some female patients on antiretroviral treatments still continued to shed the virus in the genital tract [362]. These observations lead to studies on the genotype and the phenotype of the viruses compartmentalized between blood and the genital tract.

Within the female genital tract, compartmentalization has been observed in about half of the donors studied [287, 288, 290, 291, 363-368]. Most of these studies were cross-sectional. Longitudinal studies revealed that viral compartmentalization between blood and the female

genital tract was transient, meaning with time blood and genital tract viruses mixed; it is in only a minority of the cases where viral compartmentalization between blood and the female genital tract was sustained over substantial periods of time. HIV-1 compartmentalization in the female genital tract was also observed to be a result of clonal amplifications of monotypic populations rather than unique genotypes [286, 369]. In addition, several factors have been shown to correlate with the extent of viral compartmentalization within the female genital tract; $CD4^+$ T cell counts [288, 290], availability of coreceptors on host cells [185], selective immune pressures [370], inflammation, antiretroviral selective pressure or presence of a sexually transmitted infection [145].

1.2.5 HIV-1 compartmentalization in the male genital tract

HIV-1 in semen has been reported to appear very early during infection [371, 372] and has a significant contribution to the global burden of disease [373]. The viral load in the genital fluid is a very important risk factor in the transmission of HIV-1 [374]. Acutely infected individuals are more likely to transmit the virus since the viral titres are higher during this time and mostly they are unaware of their HIV status. [375].

The male genital tract has previously been reported to have partially distinct viral compartments in around 50% of the cases and in different subtypes [189, 190, 205, 251, 289, 292, 303, 372, 374, 376-420]. This has been shown in several ways including discordance in viral load levels [392]. There is a weak association of viral load in blood and semen [373]. A study carried out in an African cohort established a spearman's correlation coefficient rho of 0.55 between the seminal and blood plasma viral RNA and the risk of virus transmission [359]. They were able to establish that for every log increase in seminal RNA copies, there was a corresponding 0.79 fold increase in risk of HIV-1 transmission [421]. However the quantities were higher in blood than in semen. It has been reported that the higher the viral load in semen, the higher the chances of the virus getting transmitted [373].

Sub-compartmentalization of HIV-1 between seminal plasma and seminal cells has been reported [380, 422-424] and provides insight as to the source of the virus found in semen [380]. Originally it was thought the HIV-1 in semen came from infected cells that got imported

from blood [425]. However, more recent studies have shown that the genital tract is a distinct viral compartment able to locally produce the virus and infected cells [292, 386]. Anderson and colleagues identified three sources of the viral populations in semen, direct import of the virus, clonal amplification and compartmentalization [381]. Before this study, previous ones had observed that HIV-1 is still detectable in the ejaculate of men who have undergone vasectomy [426] and also in the pre-ejaculatory fluid [427]. This gave insight of the source of HIV-1 in semen to be from more distal sources of the male genital tract.

Recently, a detailed tissue origin of the SIV in semen has been reported [191]. Simian immunodeficiency virus (SIV) is a relative of HIV that infects non-human primates and the results therefore can be extrapolated to explain the source of the latter as well. Using a phylogenetic approach the team was able to show that the virus in semen comes from various organs that make up the male genital tract [191]. The team also showed that virus from different organs of the male genital tract is phylogenetically compartmentalized [191]. In addition, viral populations from the epididymis, seminal vesicles and vas deferens were genetically similar to the virus found in semen [191] suggesting that they contribute more to the virus found in seminal plasma. This is at odds with older studies that had reported that the source of the virus in semen came from prostrate and urethra [428]. Because the male genital tract in humans is generally inaccessible to direct sampling, some of these earlier studies resulted in quantification of viral loads in semen [428-432].

Not all studies have observed differences in viral shedding in various organs of the male genital tract. One example is Fieni *et al* study that used a macaque model to study the viral shedding in various organs of the male genital system [433]. The team did not find compartmentalization of the viral populations between various organs of the genital tract [433]. The differences could possibly be explained by the different modes of inoculation of the virus, intravenous vs penile inoculation. Further, sampling was also not done in some organs like urethra and vas deferens. The persistent viral shedding of virus in semen long after symptoms were cleared in the recent attacks by Ebola [434] and Zika [435] viruses have revealed existing gaps in our understanding of the virus in semen [436, 437].

Compartmentalization of HIV-1 between blood and the male genital tract has also been shown to have some compartment-specific traits related to reduced positive selection, lower diversity, altered glycosylation, less dependence on CXCR4 [409]. These altered traits of genital tract virus relative to the one in blood is driven by the need for the virus to adapt to the prevailing environment in the genital tract. These differences also reflected on the phylogenetic differences between blood and male genital tract viruses [377, 380, 386, 396, 409].

HIV-1 is first shed into the genital compartments of donors before it is subsequently transmitted to the recipients during sexual transmission. One known predictor of forward transmission of HIV-1 during sexual transmission is the level of viral load in blood [218]. This does not suggest blood-to-blood transmission but the viral load in the genital compartment also correlate with the levels in blood although it is at least a log lower [376, 413]. The levels of virus shed into the genital compartments of donors varies from individual to individual [413]. It is not yet clear why this is so, but it appears there is a critical blood viral load threshold (above 10000 cp/ml) when shedding into semen happens at increased probability [205].

1.2.5.1 HIV-1 compartmentalization: Antiretroviral drugs and HIV shedding into semen

The advent of highly active antiretroviral drugs has reduced the number of HIV-associated deaths globally [66]. The drugs are able to control the virus in blood to undetectable levels thereby drastically decreasing transmission [270, 438]. There is a corresponding decrease in viral loads in semen as well [439, 440]. However it has been observed that even in patients who have suppressed virus in blood, between 2% - 48% of them still shed the virus in semen [375, 393, 402, 414, 441-443]. This is possibly due to the drugs not penetrating efficiently to critical parts of the genital tract [444, 445]. In addition it could also be a consequence of stimulation of viral replication in the genital tract by a sexually transmitted infections, inflammation [412, 414, 442, 446] or seminal cytomegalovirus replication [401]. HIV-1 compartmentalization can also be due to clonal amplification where an infected cell produces some virus that is homogenous but not replicating.

1.2.5.2 HIV-1 in semen and sexually transmitted infections

Sexually transmitted infections promote the acquisition and transmission of HIV-1 in a number of ways. These include interfering with the physical or mechanical barrier, causing inflammation or attracting more HIV-1 target cells [447].

Sexually transmitted infections (STIs) that breach the physical barrier cause ulceration and also non-ulcerative conditions and have been implicated with an increased risk of contracting HIV-1 [447]. Human herpes simplex virus type 2 (HSV-2) is the most common ulcerative genital coinfection and has been reported to increase the HIV-1 transmission efficiency [448] and shedding in semen [449, 450]. HSV-2 interacts with the host proteins causing inflammation thereby making the environment conducive for HIV-1 replication [451]. In addition, the genital ulcer is made into a portal from which HIV-1 is actively shed. Further, cytokine release due to HSV-2 has been shown to stimulate HIV-1 expression [452]. Other ulcerative STI that result in an increase in viral shedding in semen include syphilis [453].

Non-ulcerative urethritis has also been linked with up to 8 times more virus shedding in semen when groups of HIV-infected people were compared to their uninfected controls [251]. Among the non-ulcerative urethritis investigated in this study, Gonorrhea was observed to result in the highest shedding of HIV-1 in semen [251]. The viral burden was significantly reduced in semen within a week of treatment of the sexually transmitted infection but remained unchanged in blood further supporting the role of STI in HIV shedding in semen. [251]. Non-ulcerative sexually transmitted infections result in inflammation. Mucosal tissues are enriched with HIV-1 target cells when an inflammation-inducing STI like gonorrhea is involved. These infected cells are recruited by the inflammatory cytokines such as interleukins and tumour necrosis factor which intern lead to more viral shedding in the semen [454, 455].

1.3 Bottlenecks in HIV-1 transmission

Transmission of HIV-1 is known to involve a transmission bottleneck where a single variant is transmitted in around 80% of the cases [456-458]. Earlier evidence came from cross-sectional studies involving discordant couples that showed that early transmission of HIV correlates

with viral population bottleneck [459-462]. Studying the viral transmission bottlenecks and the pressures involved during selection in the donor, genital fluids and recipients helps to understanding the nature of the transmitted founder viruses, information that can be used for the design of vaccines and other intervention strategies.

1.3.1 Transmission bottleneck based on mode of HIV-1 transmission

It has been appreciated for a number of years that a stringent bottleneck occurs during HIV-1 transmissin, where the HIV-1 infection in the recipient is established by a single variant or by a limited number of founder viruses [463-466]. More recently, it has become more and more clear that the observed narrowing is very sharp even in parenteral HIV-1 transmission [467, 468]. This phenomenon has been termed “transmission bottleneck” and is the subject of our current study (Figure 13). The biological mechanisms behind this observed population bottleneck are poorly understood but physical conditions at the mucosal surfaces of the recipient [469], level of immune activation [145, 470] and availability of HIV-1 target cells [471] are some host factors seen to influence it.

Traditionally, the number of the transmitted founder (T/f) had been shown to vary depending on the mode of HIV-1 transmission i.e. sexual [457], mother-to-child transmission [464] or parenteral [466]. During the heterosexual transmission of HIV-1, 80% of the transmitted virus is due to a single virus [457, 458, 469, 472-474]. The risk of transmission of multiple viruses increases in virus transmissions involving men-who-have sex with men and in intravenous drug users (IVDU) [475]. However more recently, it has become apparent that the transmission of the virus in men-who-have sex with men, involves single variant transmission in 75% of the recipients [469, 476]. Transmission of single variants in mother-to-child transmission is approximately 70% [477] and 40-80% in injecting drug users [72, 467, 478]. These levels are identical to those observed in heterosexual transmissions. This indicates that even though the narrowing seems to be stochastic mediated by the mucosal surfaces of the recipient or low titre inoculum encountering a rare target cell, some narrowing steps are heavily selective. This shows that the mucosal surfaces of the recipient are not the only source of selection but more selection happens systemically as well and therefore the T/F virus may

have some features different from chronic virus that provide necessary advantage during transmission.

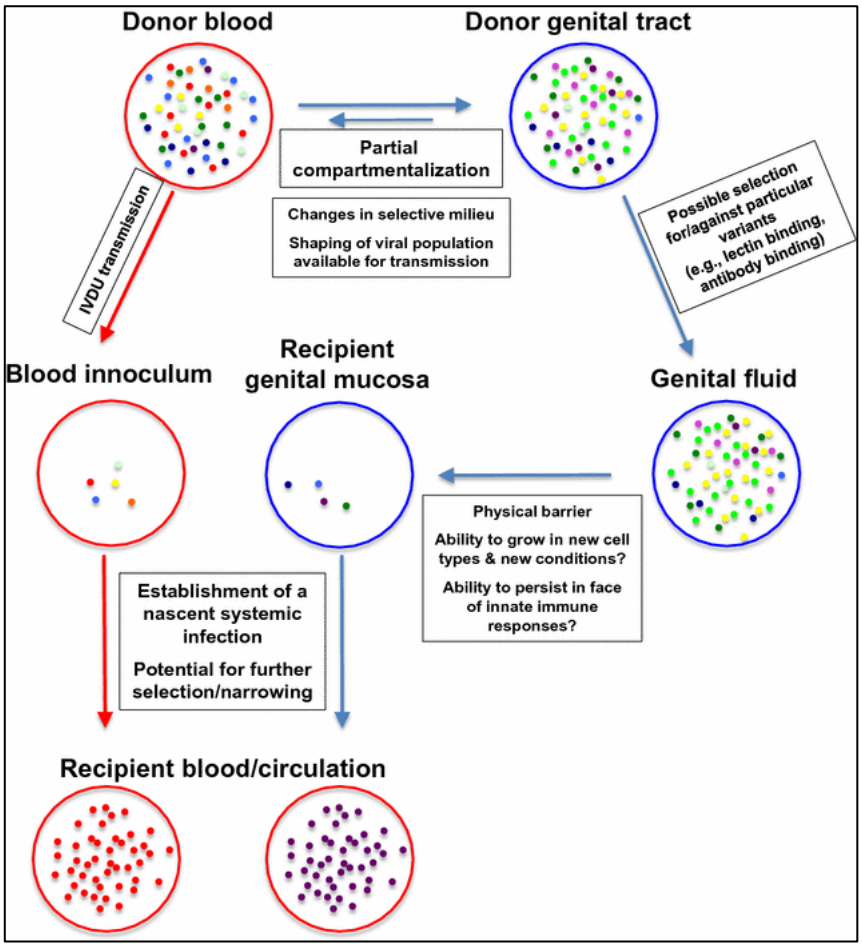


Figure 13: Depiction of the HIV-1 transmission bottlenecks from an infected donor to the recipient. (adapted from [468])

1.3.2 Transmission bottleneck among intravenous drug users

Drug use is a complex, chronic behaviour usually difficult to uproot due to development of dependence. Unfortunately, several disease agents including HIV-1 are transmitted through this route resulting from sharing injection apparatus and risky sexual behaviour while under influence. Intravenous drug use involves introduction of a drug systemically via exposed blood veins on the skin through a needle and syringe. In some parts of the world, like Eastern Europe and Central Asia, the HIV-1 epidemic was driven by intravenous drug users and their sexual partners [147, 479]. In Russia, the clone circulating among some intravenous drug users has

sequence divergence constraints meaning that it was a clonal introduction into the population in more recent times followed by spread from this single introduction [480].

HIV-1 has been shown to involve a substantial bottleneck irrespective of the mode of transmission. The bottleneck had been reported to be more in heterosexual transmission [472], intermediate in MSM [481] and least in blood-to-blood transmission. However, recent studies have shown that there is substantial bottleneck that exist among intravenous drug users [72, 475]. The virus that established the infection in these studies was derived from single variants in 40-75% of the donors. The two cohorts studied were independent of one another and came from different settings, one from North American [475] and another from Russia [72]. In the north American cohort, 6/10 (60%) of the transmissions involving IVDU were traced to multiple variants while 40% were traced from single variants [475]. In the Russian IVDU cohort, 69% (9/13) involved transmission of single variants [72]. Combining this with unpublished data from the same cohort involving 6 out of 7 single variant transmissions [482], then 15/20 (75%) of IVDU HIV-1 transmissions involved single variants in this cohort. This data indicates that substantial bottleneck exists in blood-to-blood transmissions where no mucosal barriers are involved. Unexpectedly, this was very similar to the proportion that is observed in sexual transmissions [467].

A potential limitation to interpretation of these studies however, is in sizing the inoculum involved. The donor viral load and amount of blood left in the needle are variable. During the acute phase of infection, the viral loads are usually very high and since IVDUs share needles in a network and more frequently, the virus can spread fast. In some simulations [483], without syringe rinsing, approximately 84µl and 2µl of donor blood is left in the high and low dead volume syringes respectively. These residue donor blood volumes could be used to approximate the size of the inoculum. However, this can be confounded by practices such as booting; where blood is redrawn into the syringe a second time to rinse residual drug left in the dead space [484]. With syringe rinsing, approximately 1µl and <0.001µl is left in high and low dead volume syringes respectively [485-487]. Other practices that could complicate sizing of the inoculum include heating the drug before injection, rinsing of the needles after and before use and time between needle sharing.

Another limitation could be that these transmissions that are thought to occur through intravenous drug use could have been transmitted sexually. However, for the Russian cohort, transmission is more likely to be through intravenous drug use since the IVDU-linked transmissions are mostly subtype A while sexual transmissions are subtype B [480]. Overall, the rate of single variant transmissions involving IVDU seems to be higher than it would be expected. These observations are corroborated by a study showing reduction in viral diversity after controlled intravenous inoculation of macaques with SIVmac251 or SIVmac660 [488]. These studies suggest that the mucosal surfaces of the recipient are not absolutely responsible for the observed transmission bottleneck but rather it involves multiple selective barriers within the donor genital tract, donor and recipient genital fluid as well as within the recipient blood (Figure 13).

1.3.3 A fitness bottleneck during HIV-1 transmission

Transmission of HIV-1 selects for fitter viruses at the transmission bottleneck where establishment could easily be extinguished [489]. The fitness bottleneck is observed during the establishment of HIV-1 at the portal of entry during the infection of the first cell and more cells thereafter, usually measured by the number of virus particles produced by an infected cell or the infectiousness of these particles. Despite the complexity and diversity of HIV-1 across the epidemic and within an infected individual, heterosexual transmission mostly selects a single founder outgrowth virus which comes from the minor variant in the donor [490]. Carlson *et al* observed a selection towards transmission of amino acids that correlate with increased fitness of the virus [489]. Thus the probability of transmission was influenced by the viral fitness.

The concept describing the virus establishing infection is based on the reproductive number (R_0) which is the average number of cells infected from an infecting single host cell. The infection spreads only when this number is greater than 1 and dies out if the number is less than 1 [193]. Studies done during the early events before expansion show that the R_0 is initially low but it increases with time as more cells get infected [193, 491]. Carlson *et al* also reported more polymorphism being transmitted in male-to-female transmission than female-

to-male transmission meaning less fit virus transmission were more common in male-to-female transmission than vice versa [489]. This severe bottleneck during female-to-male transmission can be mitigated by a genital urinary infection in the recipient or if the viral load was high in the donor [489].

1.3.4 Multivariant HIV-transmissions and sexually transmitted infections (STIs)

Recent observations that HIV-1 transmission among the intravenous drug users involves a single variant raises questions about the role of sexually transmitted infections (STIs) in transmission of multiple variants. Acquisition of multiple variants has previously been linked to faster diseases progression evidenced by higher viral loads and lower CD4 counts among these individuals [492]. Some STIs such as syphilis [243, 493], chlamydia [494] gonorrhea [495] lead to increased risk of HIV-1 infection [446] as well as multiple variant transmission [469, 496]. These results are usually attributed to the breaks in the mucosal surfaces [497], recruitment of target cells or production of inflammatory cytokines [498]. In some studies, however, the presence of an STI in the donor or recipient was neither associated with risk of HIV-1 infection [499, 500] nor multivariant transmission [469]. It is possible that these studies were underpowered to detect the association between STI and increased multivariant transmission. Even for the two studies linking HIV with risk of multivariant transmission [469, 496], association with any particular STI was not shown. Instead vaginal/urethral discharges [469] or grouping all STIs together [496] were associated with multivariant transmission, possibly because there were too few events in the study of any one STI to be able to easily detect a specific association. Therefore, perhaps the effect of an STI in promoting HIV-1 infection is only to provide a modifying effect without necessarily compromising the integrity of the mucosal barriers. These effects provide the synergy necessary for increased HIV-1 infection and risk of multivariant transmission [242, 447, 501]. Also, the observations of an elevated risk of infection establishment by a single variant among intravenous drug users [72, 467, 475, 482], disfavour the idea of the mucosal surfaces or defences being solely responsible for the transmission bottleneck.

1.3.5 Features of the transmitted/founder virus

Observation that the transmitted/founder virus involves a transmission bottleneck from a highly diverse swarm in chronically infected individuals has led to a holy grail of characteristics associated with this outgrowth virus. Several attributes both genotypic and phenotypic seem to correlate with the transmitted founder virus. Most of these attributes seem to be inconsistent among the studies. Utilization of CCR5 co-receptor by the T/F virus [502-506] is consistent among many studies but other attributes are observed in one study while absent in another. While these features support the idea of the transmitted/founder virus being selected for during transmission [468, 489, 490, 507, 508], other processes involve stochastic narrowing of the viral population at the transmission bottleneck.

Since the features of transmitted/founder virus selection might extend to the donor as well, this thesis will examine the differences in phenotype of the variants found systemically and those found in the male genital tract since they are the viruses that are potentially transmitted to the recipients. In the next sections both the genotypic and phenotypic features relevant to our current study that have been shown to characterise the transmitted/founder virus are discussed.

1.3.5.1 Genotype of the transmitted/founder virus

Several genotypic features have been observed in transmitted/founder viruses compared with chronic viruses found in the transmitting partner. These features have been gathered from studies comparing genotypes between transmitting pairs or general comparisons of sequences collected at chronic and acute phases of HIV-1 infection. Some of these features include variable loop characteristics, i.e. number of potential N-linked glycosylation sites found on the HIV-1 envelope and length of the envelope variable loops. These features are individually discussed below.

a. Transmitted/founder virus has shorter hypervariable loops of the HIV-1 env

It is thought that the long hypervariable loops of the HIV-1 envelope could be disadvantageous to the virus that is initiating transmission. One explanation is that the long loops which result from insertions may limit the virus access to CD4 receptor and CCR5

coreceptor. On the other hand, longer variable loops are advantageous to the chronic virus as they can mask neutralizing antibody epitopes facilitating escape. This is necessary in chronic infections where immune pressure increases. Even with the said benefits, the finding of shorter envelope loops as a property of the transmitted/founder viruses has been inconsistent among studies.

Studying HIV-1 subtype B for example, Gnanakaran and colleagues confirmed that early infecting viruses had shorter V4 loops compared to chronic infections [509]. This however was not seen in other studies involving HIV-1 subtype B during early infection [510-512]. However transmission of viruses with shorter hypervariable loops has been consistent among subtype A [510], D and C [459]. This therefore suggests that transmission in these subtypes favours compact envelope glycoproteins that interact better with target cells at the mucosal surfaces of the recipients [513]. During disease progression, the variable loops grow in length towards chronic infection. This is however sustained up to a particular limit where they shorten again in late stage of disease [514]. This might be due to release of immune pressure as the host is severely immune compromised at the later stage of infection. Indeed, this is supported by a study showing slower evolution of the virus in patients with low $CD4^+$ T cell counts [515].

b. Transmitted/founder virus carry fewer potential N-linked glycosylation sites (PNGS)

HIV-1 envelope protomer is comprised of approximately 30 N-linked glycoproteins sites (Asparagine) whose role is to anchor the carbohydrates to the HIV-1 envelope. Consequently, the carbohydrate protects the underlying protein from recognition by circulating antibodies [516]. Viruses in chronic infections are selected partly on increase in number or varying the position of the carbohydrates on them. This has led to development of the term 'glycan shield' although some neutralizing antibodies can recognize these carbohydrates [517]. With the disease progression also, the HIV-1 loop lengths become longer as they adapt to immune pressure [514]. But as previously indicated, the T/F virus favours variants that have shorter loops. It is therefore also possible to expect fewer glycosylation in the T/F virus as well.

Traditionally, the role of these glycans was thought to only influence immune escape. But more recent studies are showing that glycans have more roles than that. One of these is their involvement in HIV-1 transmission as will be detailed in this section. In addition, they are involved in providing some structural characteristics as they are found as part of the epitopes and are targeted by some broadly neutralizing antibodies [518].

Among the original studies, Deydeyn and colleagues observed that among the 8 transmitting pairs that they studied, five of the pairs involved transmission of variants that had smaller number of PNGS compared to the median number in the donor [459]. However, taken together the number of either donor or recipient sequences did not differ significantly from HIV-1 subtype C viruses available in the Los Alamos database which has sequences from all infection stages. Similar findings were also observed among HIV-1 subtype A viruses [510] and in another subtype C virus study [506]. Interestingly, one study showed that under-glycosylation was observed more when comparing sequences from acutely infected men with chronically infected women [506].

Though the advantages of fewer PNGS for the T/F virus has not been fully elucidated, one study has shown that having fewer glycosylation sites is important in enhancing the interaction of the virus with α4β7 integrin [519]. This is a homing marker available on CD4+ T cells that directs them to the gut-associated lymphoid tissues (GALT). This is the reason the GALT is among the first to amplify the virus following successful establishment. Despite this, fewer PNLGS as a feature of T/F virus has not been consistent among subtypes suggesting that the phenotype resulting from it might be beneficial for one subtype and not the other. For example, it is absent in HIV-1 subtype B [510, 520]. Even with the inconsistency, it appears that the differences might be small or even unnoticeable in some subtypes, but they might affect key structural characteristics of the HIV-1 envelope necessary to provide the needed property to establish. These properties might include among others enhanced binding to α4β7 as previously mentioned.

1.3.5.2 Phenotype of the transmitted/founder virus

Among the phenotypic features that have been observed to be favoured during HIV-1 transmission include difference in utilization of the coreceptors and neutralization sensitivities to donor antibodies. These are individually discussed in the below sections.

a. Transmitted/founder virus preferentially utilize CCR5 coreceptors

Among all the features of the transmitted/founder viruses studied, the utilization of the CCR5 coreceptor is the only characteristic that has been a consistent feature [503, 521]. There is a selective transmission of CCR5-utilizing viruses and less of the X4 tropic viruses, possibly due to the fact that X4 tropic viruses develop later in the infection. Dependency on CCR5 coreceptor in subtype B is higher during early infection and coreceptor switch to CXCR4 or utilization of both is seen with disease progression. Recently it has been reported that some chronic viruses are able to utilize a maraviroc-bound conformation of CCR5 [504] or a conformation that is not usually bound by maraviroc when CCR5 expression levels were high. This trait of promiscuous utilization of CCR5 coreceptors by chronic viruses was selected against by transmitted/founder (T/F) viruses in both subtype B and C viruses [504].

b. Transmitted/founder virus is sensitive to the transmitter's antibodies

Transmitted/founder virus has been seen to favour epitopes that are lost during chronic infection as the virus tries to escape the immune pressure from the infected host. This is possibly due to absence of autologous antibodies in the recipient until later [522, 523] or the associated cost is detrimental to establishment and therefore selected against. The latter is a more plausible explanation since it has been established that some low titre neutralizing antibodies are available as early as two weeks of infection [524]. At the population level, broadly neutralizing antibodies do not provide protection against HIV-1 but they do provide substantial selective pressure on the viral populations. This selective pressure makes viruses in later stages of infection more resistant to neutralization by antibodies.

Transmission of autologous neutralizing antibody-sensitive viruses has been reported to be a feature of HIV-1 T/F viruses [459, 525, 526]. The study by Derdeyn *et al* showed that recipient pseudoviruses were 10 times more sensitive to donor plasma compared to their donor counterparts [459]. These results were also corroborated by similar ones by Deymier and

colleagues using infectious molecular clones [526]. The explanation comes from another feature of a compact V1-V4 region of the T/F virus [459]. This tendency of compact V1-V4 region would lead to enhanced exposure of the CD4 binding site. This increased exposure has often been associated with sensitivity to neutralization [527, 528]. In addition, sensitivity of recipient viruses to donor antibodies was also shown in a subset of cases in HIV-1 subtype B transmission in a cohort involving men who have sex with men [525]. It appears that the envelope conformations favoured at chronic infections are selected against at the transmission bottleneck. This means that the HIV-1 transmission favours envelope conformations that are accessible to receptors on the target cells.

1.4 HIV phylogenetics

HIV- popolations expand from an initial small populations during transmission and diversity increases towards chronic infection [457]. Therefore viral phylogenetic becomes a good tool to study HIV populations. Traditonally, phylogenetic approach has been used in study HIV-1 compartmentalization [377, 380, 394, 400, 529] and phylogenetic compartmentalization has for long been the gold standard. HIV Bayesian phylogenetics have also been used in HIV-studies especially in epidemiology. This is because is combines several factors such as viral movement dynamics, demgraphics and epidemiological factors that can be encoded into the molecular sequence data [48]. Bayesian phylodynamcis have also been used in studying HIV-1 movement between anatotomical sites [303, 325, 530]. These analyses provide information that is beneficial in evidence-based solutions for prevention, management and control of disease.

1.5 Motivation and rationale for the study

Sexual transmission of HIV-1 remains the most predominant mode of spread contributing to 70-80% of all the transmissions [81]. Despite the high diversity that characterizes the donor population and transmission fluids, it has been observed that only one variant gets transmitted in 80% of the cases in heterosexual [457, 472], 75% in homosexual [467, 481] and 60-80% in intravenous drug users [467, 468]. This suggests that transmission of HIV-1 involves

one or more selective events within the path of transmission from the donor to the recipient. The biological mechanisms of this bottleneck remains largely unknown.

The majority of this genetic bottleneck is thought to be mediated by the mucosal surfaces of the genital tract [469]. It is not clear to what extent the intact mucosa of the recipients selects for the transmitted virus over and above the physical obstruction of the virus. More studies showing that a substantial bottleneck exists during HIV-1 transmission among intravenous drug users, show that the mucosal surfaces are not solely responsible for observed bottleneck [72, 467]. Availability of HIV-1 target cells [471], genital inflammation [145] and level of immune activation [470] are other host factors suggested to play a role. Previous studies that worked to understand the transmission bottleneck compared isolates from chronic and acute HIV-1 infections [459, 489, 525, 526]. By doing this, they cannot address at what stage selection happens, anywhere from the path from donor blood to disseminated infection of the recipient, with selection at multiple steps very possible.

One potentially very important step in this bottleneck may be associated with the compartmentalization of HIV-1 in the genital tract which has been shown to be a distinct compartment separate from systemic circulation [289, 377, 386, 396]. In addition to being compartmentalized, the virus in the genital compartment has also been shown to sometimes be clonally amplified [381]. Techniques that have been used previously to study HIV-1 compartmentalization and clonal amplification cannot detect minor variants robustly. Deep sequencing on the other hand gives the advantage of sampling the HIV-1 populations represented in minor proportions. The current deep sequencing technique has been improved to include primerID [531-537] thereby mimicking the viral composition in vivo as opposed to collapsing the sequences into haplotypes.

Even though today antiretroviral drugs are available that have substantially reduced the morbidity and mortality resulting from HIV-1 related infections, understanding the molecular mechanism of sexual transmission is needed in order to design some precise, targeted, more specific and potent prevention strategies.

Several studies have studied the variations that exist between the seminal and blood virus variants [367, 380, 409, 433, 490] but their findings have not extended to neutralization studies, nor has deep sequencing been employed. Deep sequencing would afford us the opportunity to get a fuller picture of the selective pressure on the viruses and to employ more sophisticated statistical methods to analyse our data.

1.6 Study objectives

Since compartmentalization of minor variants cannot be tracked using techniques used previously to describe compartmentalization between blood and the male genital tract, deep sequencing was used in this study. In addition, the changes in function for clones derived from blood and semen were evaluated by testing them against HIV-1 entry inhibitors, We further hypothesized that the change of selective milieu as HIV-1 moves from blood into the genital tract may lead to viral variants in semen that are sensitive to autologous neutralization. Hence the virus in semen comes from a site that is relatively protected from antibodies and may be permitted to evolve differently relative to absence of antibody pressure. This evolution of the virus within the genital tract may be a significant part of the change the virus undergoes on its way to establishment in the new recipient. We considered this possibility because even some small molecules like those of some antiretroviral drugs do not penetrate the genital tract effectively. The objectives were;

4. To evaluate HIV-1 compartmentalization in blood and the male genital tract using next generation sequencing to understand the nature of viral populations in these anatomical sites in greater detail.
5. To identify the differences in sensitivity of blood and semen variants to entry inhibitors to obtain information about differences in function between HIV-1 populations in blood vs the male genital tract.
6. To compare neutralization sensitivities of viral variants compartmentalized in blood and semen by testing their sensitivity to neutralization by autologous antibodies. As a control, we measured sensitivity to a pool of clade-matched heterologous sera to determine if any observed difference was due to global changes in neutralization sensitivity.

2 CHAPTER 2

Compartmentalization and clonal amplification of HIV-1 between blood and the male genital tract shown using next generation sequencing

2.1 Abstract

Compartmentalization of HIV-1 between blood and the male genital tract may have substantial impact on which viruses gets transmitted during sexual transmission. Knowledge and characterization of these variants is important for the design of vaccine and other intervention strategies. Here the power of next generation sequencing was utilized to study compartmentalization and clonal amplification in a greater detail. A total of 44 study participants were recruited and in the end, a subset of 11 with sufficient viral load in semen to allow deep sequencing. Illumina Miseq was used in combination with PrimerID approach which groups together all the PCR products derived from each cDNA molecule during the reverse transcription process. This combination of approaches allowed us to achieve a high sequence depth including minor populations and the ability to account for PCR artefacts of PCR-based recombination and base substitutions.

Compartmentalization was analysed using three standard techniques, Slatkin-Maddison test, nearest neighbour statistic and Wright's measure of population subdivision. Clonal amplification was evaluated by collapsing the sequences at 99% similarity using Vsearch.

Viral sequence compartmentalization between blood and the male genital tract was observed in all the 11 study participants, unlike in previous studies where compartmentalization was reported in about half of the donors. The extent of compartmentalization however varied among the study participants. Simulation of lower depth of sampling (i.e. like used in single genome analysis) suggested that compartmentalization could have been detected on only about half of the study individuals. Thus, the difference in detecting HIV-1 compartmentalization in this thesis vs previous studies can be explained by the depth of sequencing achieved here and that there is no evidence that the dynamics of the viral populations studied in this thesis were different from those previously studied.

Importantly, the location of the most recent common ancestor (MRCA) of semen sequences was located in blood suggesting that the genital reservoir is seeded from the incoming variants from blood. Clonal amplification was observed in all the study participants and it was a characteristic of both blood and semen viral populations which support the idea of

independent replication of variants within blood and the male genital tract. In addition, the proportion of clonally amplified variants that traffic poorly between compartments varied among the donors.

These findings will help in better understanding the dynamics that affect the composition of viral populations that are available for transmission in the HIV-1 infected male donors.

2.2 Background

Compartmentalization of HIV-1 i.e. restriction of viral movement between the genital tract and the general blood circulation is important in understanding sexual transmission of HIV-1. An estimated 80% of the HIV-1 transmission occur through the sexual route especially via contaminated genital tract secretions [160]. Indeed, during male-to-female transmission and to some extent in men-who-have-sex-with-men (MSM), the main carrier of the virus is semen. In addition, the compartmentalized viral variants in the genital tract could become a reservoir from where the blood compartment is reseeded after a successful but interrupted antiretroviral treatment (ART) [444] or be a site of generation of drug resistance mutations [538]. Compartmentalization of HIV-1 can be promoted by physical barriers, differential selective pressure, host cell types and immune pressure.

The diversity of HIV-1 in an infected person is immense and has previously been compared to the entire flu diversity in the world [539]. This therefore means that innovative tools are needed to study it and to design potential vaccines and other intervention strategies. Previous studies working on HIV-1 compartmentalization between blood and the male genital tract had relied upon either single genome amplification (SGA) technique [289, 303, 381, 394, 490, 529, 540] or cloning from bulk [377, 541]. Each of these methods have their advantages and disadvantages. For Example, SGA is advantageous over bulk PCR by allowing less PCR-induced recombination and resampling which can result in underestimation of the sequence diversity of the target population [542]. The technique is however expensive and labour-intensive, thereby limiting the number of isolates that can be studied. With the above constraints also, the few isolates possible, mostly come from the dominant variant in the HIV-1 population of an infected individual. On the other hand, cloning form bulk PCR is subject to cloning bias where some variants are preferentially cloned than others [543]. In addition, it is

prone to artificial, PCR-based recombination between different templates. Older studies had relied upon Heteroduplex Tracking assay (HTA) [367, 461, 544, 545] but that technique does not provide the sequence information; rather it provides qualitative information only.

Next generation sequencing minimizes these artefacts and more recent studies have employed it to study HIV-1 populations in infected individuals [546-549]. A more recent improvement of deep sequencing is the introduction of PrimerID approach [532, 534, 537] which involves introduction of a random 9-15 nucleotide sequence at the 5' end of the PCR products. The primerID therefore, enables grouping of sequences according to the single cDNA molecule from which they were amplified. This makes it possible to account for resampling, recombination and minimise the risk of reading point mutations into the final sequence [550]. Thus, primerID approach is a critical improvement to minimize the error rate of the existing next generation sequencing techniques [534].

In this second chapter of the thesis, the increased power afforded by next generation sequencing is utilized to study viral compartmentalization between blood and the genital tract in HIV-1 infected male donors in more detail than studied previously. In addition, utilization of primerID allowed analysis of clonal amplification in the two compartments in greater detail than previously possible.

2.3 Methods

2.3.1 Study participants and samples

A total of 44 HIV-1 positive therapy-naïve men were recruited from a clinic that gives care to predominantly HIV-1 infected men at Woodstock, Cape Town, South Africa. Individual written consent was obtained from every study participant before blood and semen samples were collected. The inclusion criteria included: male, HIV—infected, adults of over 8 years and not on antiretroviral treatment (ART). The enclusion criteria included: past or current history of antiretroviral treatment, those with poor venous access, transgender individuals or any other individual who, in the opinion of the clinician giving care or recruiting staff, is unsuitable for participation in the study due to poor health. Demographic information relating to age,

nationality and racial group was collected. In addition, history of infection relating to current or previous sexually transmitted Infection (STI) was also obtained through an interview with the clinician and from the participants' medical records. To estimate the duration of infection, the last HIV-1 negative test and earliest HIV-1 positive test dates were recorded, if they were available. The details of the study participants are described in Table 2.

Blood was obtained via venous puncture and collected in 4 tubes; 5ml PPT tube for viral load assay, 4ml EDTA tubes for $CD4^+$ cell counts, 9ml EDTA tubes for plasma and in serum separator tubes without anticoagulant. The 9ml EDTA plasma and serum separator tubes were then centrifuged at 2000rpm for 15 minutes. Blood plasma and serum were aliquoted in cryovials vials and Eppendorf tubes respectively and immediately stored at -80^{o}C.

Semen was collected through self-masturbation at the clinic, immediately transported to the lab and processed within four hours of collection. Briefly, following a 30-60 minutes wait (for semen liquefaction) at room temperature in a laminar flow hood, the semen specimen was transferred to a 15ml sterile polypropylene conical tube with a wide-mouth plastic transfer pipette and then diluted with an equal volume of PBS containing penicillin (100U/ml) and streptomycin (100µg/ml). The diluted semen specimen was underlaid with 2ml of 19% cushion of Nycodenz (Axis-Shield PoC AS, Oslo, Norway) reconstituted in PBS with penicillin and streptomycin added. The tube was then capped, placed on a centrifuge bucket with a cap, and centrifuged at 1000xg for 20 minutes. The non-seminal-cell (NSC) layer was atop the cushion, the sperm pellet to the bottom of the tube (Figure 14). Supernatant seminal plasma was removed to within approximately 1cm of the NSC layer and filter sterilized through a 0.8µm filter. The seminal plasma virus was then concentrated by centrifugation at 100,000xg for 1 hour at 4^{o}C. After ultracentrifugation, the supernatant was discarded, and the pellet resuspend in 200µl of PBS. From the 200µl volume, 40µl was diluted 12.5 times using PBS and taken for viral load assay. The rest was aliquoted in two volumes of 70µl and one of 20µl and stored at -80^{o}C.

The NSC layer was removed with a wide bore transfer pipette, transferred to a fresh 15ml conical tube, and diluted with five times volume of PBS with penicillin, streptomycin and nystatin (100 to 200 U/ml). The NSC was then recovered by centrifugation at 500g for 10

minutes, the supernatant removed and discarded, and the process repeated twice. The Washed NSC pellet was then preserved at -80°C. These samples were not analyzed as part of this thesis.

A subset of 11 study participants with sufficient viral loads (above 4000 copies per sample) in semen were selected for deep sequencing; all had sufficient viral loads in both blood and semen to be suitable for analysis: SVB008, SVB012, SVB021, SVB025, SVB026, SVB029, SVB030, SVB039, SVB041, SVB043 and SVB049.

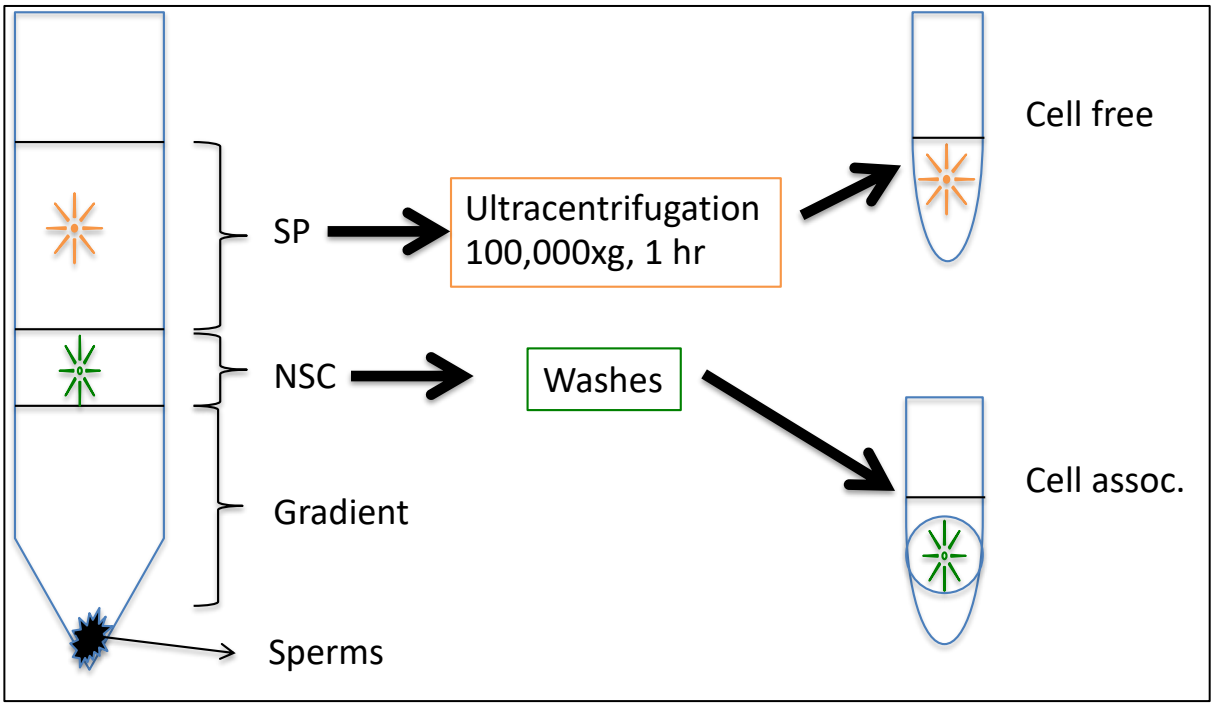


Figure 14: Schematic representation of semen processing. After liquefaction, semen was underlaid in 19% Nycodenz gradient and centrifuged at 1000xg for 20 minutes. After centrifugation, the sperm pellet was at the base of the tube, the non-seminal cells (NSC) on top of the gradient cushion followed by semen plasma (SP). Semen plasma was then concentrated through ultracentrifugation while the NSC was washed several times and both specimen stored.

2.3.2 Primers and the sequencing region

The sequenced region was either the C3-V5 region encompassing 423 bp (HXB2 7179-7602) or the V3 region encompassing 416 bp (HXB2 6909-7325). The primers used for deep sequencing of the viral RNA from blood and semen are summarized in Table 1. The cDNA gene-specific long primers covering either the C3-V5 or V3 were designed such that the gene-

specific part that primed the production of cDNA was at the 3' end (Figure 15). It was preceded by the PrimerID which comprised a 15-nucleotide region. The PrimerID was then preceded by a region where the universal reverse primer for the first round PCR bound. This design allowed the use of one universal PCR primer irrespective of the region sequenced. The sequencing region was designed in such a way that the PCR amplified region allowed the use of paired end sequencing with no information lost in the intertwining region between the forward (R1) and reverse (R2) after sequencing. The PrimerIDs were designed in such a way that they are always sequenced at the reverse end (R2).

Figure 15: Design of the cDNA primer used for generation of cDNA from either the V3 or C3-V5 of the RNA extracted from blood and semen samples. The cDNA primer contained a gene specific primer at the 3' end preceded by the 15-nucleotide PrimerID and then the binding site of the reverse primer during the PCR reaction.

Table 1: The list of primers used for the cDNA synthesis and amplification of either the V3 or the C3-V5 region of the blood and semen viruses. The 'N' in the cDNA primer represent the PrimerID section that was used to combine the sequences that were generated from the same cDNA.

Primer name	Sequence (5'-3')	Purpose	Region
Env_C5_cDNA	GAGATGTGTATAAGAGACAGNNNNNNNNNNNNNNNGTCCYTCATATYTCCTCCTYCAGG	cDNA primer	C3-V-5
Env_V3_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNGGACCAGGGAGAGCATTTGTTAC	Forward primer	C3-V5
Env_C3_cDNA	GAGATGTGTATAAAGAGACAGNNNNNNNNNNNNNNTGTGTTGTAAYTTCTAGRTC	cDNA primer	V3
Env_C2_F	TGGACGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNGCTGGTTATGCGATTGTAAACTG	Forward primer	V3
Univ_i7_Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	Reverse primer	Universal reverse

2.3.3 RNA extraction and cDNA synthesis

The viral RNAs from blood and semen samples were extracted using QIAmp Viral RNA mini Kit (Qiagen) following the manufacturer's instructions. The extracted RNA was eluted in a final

volume of 50µl. The RNA was then reverse transcribed into cDNA using a gene-specific cDNA primer (Table 1) using SuperScript III First Strand System for RT-PCR (Invitrogen) and following the manufacturer's instructions. Briefly, the extracted RNA was heated to 65°C for 5 minutes to denature the RNA secondary structures. This was followed by addition of 35µl of a mixture made of 1x of 5x buffer, 5mM Dithiothreitol (DTT), 2 U/ul RNAse inhibitor and 10 U/ul Superscript III reverse transcriptase. The reaction mixture was then incubated in a thermocycler with two step temperatures of 45°C and then at 50°C for an hour each. After reverse transcription, the enzyme was denatured by incubating the reaction mixture at 70^{o}C for 15 minutes before holding at 4^{o}C. The generated cDNA was removed from the DNA-RNA hybrid by digestion using RNaseH (Invitrogen) by incubation at 37^{o}C for 20 minutes.

2.3.4 PCR amplification of the region of interest

The cDNA generated from the previous step was divided into 5µl aliquots and first round PCR performed using Platinum HIFI Taq polymerase (Invitrogen). First round PCR for the C3-V5 or V3 regions was done using forward primers env_V3_subC_F and env_C2_subC_F (Table 1). A universal reverse primer (Univ_i7_REV, Table 1) which binds to the primer binding site on the cDNA long primer (Figure 15) was used. The PCR reaction mixture was comprised of 2mM $MgSO_4$, 0.5mM dNTPs, 0.2µM each of reverse and forward primers and HIFI buffer. The PCR conditions involved an initial incubation of 2 minutes at 94^{o}C followed by 25 cycles of 94^{o}C (30 sec), 58^{o}C (15 sec), and 68^{o}C for 2 minutes. A final extension at 68^{o}C for 10 minutes was done followed by holding at 4^{o}C.

2.3.5 Addition of Illumina indices

Addition of Illumina indices was done using PCR, which was performed from 5µl of the first round PCR product using the Platinum HIFI Taq (Invitrogen). The primers comprised of a 6-nucleotide index region complementary to the illumina sequencing adapters fixed on the sequencing chip. Multiple combinations of indexing primers were used for each sample to allow multiplexing of the samples within the same sequencing chip. The sequenced region was designed to exploit paired end sequencing without information being lost at the intertwining region between first (R1) and second (R2) reads. The primer ID region was

designed to always be sequenced at the second end (R2). The PCR reaction mixture contained 1X of the 10X Buffer, 2mM $MgSO_4$, 0.5mM DNTPs and 0.2μM of each of the indexing primers. The PCR cycling conditions involved an initial temperature of 94°C for 2 minutes followed by 25 cycles at 94°C (30 sec) denaturation, 58°C (15 sec) annealing and extension at 68°C for 2 minutes. A final extension at 68°C for 10 minutes was also performed before holding at 4°C.

2.3.6 Bead purification

Before performing the first round PCR, the cDNA was purified using the Agencourt RNAClean XP beads (Beckman Coulter).The first round PCR and indexing PCR products were also purified using SPRISelect beads (Beckman Coulter) following the manufacturer's instructions. For both types of beads, a bead to PCR-product ratio of 0.6 was maintained and washing done using 70% ethanol. The purified products were eluted in 50μl of nuclease-free water.

2.3.7 Illumina sequencing

The size of the C3-V5 env amplicon was approximately 423bp, covering HXB2 7179-7602 [551] on the HIV-1 genome while that of V3 region was approximately 416bp covering 6909-7325 of HXB2. Paired-end (2x300bp) multiplex Illumina Miseq was used to sequence the constructed libraries.

2.3.8 Sequence processing pipeline

An in-house R-based pipeline (MotifBinner2.0), (accessible at GitHub: https://github.com/HIVDiversity/MotifBinner2) was used to process the fastq format files into primerID-consensus sequences as described by Jabara *et al*, 2011 [537]. The pipeline comprised a combination of tools for assessing the quality of both the reverse and forward reads as well as ambiguous nucleotides and primer dimers. Following PCR and sequencing, each sequence should be from the same input template if they contain same primerID. The cleaning of the sequences was done by scanning each sequence for primerID. The sequences bearing the same primerID were binned together. The bins made up of sequences whose primerIDs possibly contains sequencing errors were removed by use of a consensus cutoff

approach as described by Zhou *et al*, 2015 [533]. To remove these offspring primerIDs, a computation was performed to determine the likelihood that more than a certain percentage of the contents of a small bin was the result of sequencing errors in the bigger bin. If it is more likely that the contents resulted from sequencing errors in a larger bin's primerIDs than it is for the bin to naturally exist (i.e. its primerIDs to have been sampled from the pool of available primerIDs), then the bin was discarded. If it is unlikely that the prespecified percentage of the bin is the result of sequencing error in the primerIDs of the larger bin, then the bin was kept provided that it was larger than a certain minimum size. For the purposes of flagging bins as potential offspring bins, the cuttoff was doubled. Bins smaller than double the consensus cutoff were referred to as small bins while bins larger than double the consensus were referred to as large bins.

The pipeline utilized tools for assessing the quality of the read data from both R1 and R2 reads supplied. The reads that didn't match the required read lengths, and sequences that contains more than 2% of ambiguous nucleotides or obvious primer dimers were discarded. The sequences were then trimmed to high quality reads. The second end reads (R2) were then searched for the primerID and the number of reads with these primerIDs tabulated. The forward R1 and reverse R2 reads were consequently matched, paired together and another quality trim performed followed by another binning together of the sequences with the same primerID. The total number of sequences obtained defined the depth of sampling. Finally, gaps were removed from the primerID-consensus sequences. The number of consensus PID (PrimerID) sequences obtained after employing the pipeline are summarized in Table 3.

2.3.9 Phylogenetic analysis

The processed PID-consensus sequences or haplotypes (variants) from blood and semen sequences were combined into one fasta file. The alignment was then performed using MAFFT [552] and visually inspected and edited in Aliview [553]. Maximum likelihood trees were constructed from the PID-consensus sequences using FastTree [554] employing the GTR model. Fasttree was chosen because it can handle large alignments with reasonable computational memory and time. In addition, Bayesian phylogenetics were performed from equal number of haplotypes (variants) from blood and semen constructed using a Vsearch [555] haplotyper tool that allowed 99.75% similarity.

To study the HIV-1 movement between blood and semen, Bayesian Evolutionary Analysis by Sampling Trees version v1 (BEASTv1.10.4) was used [556]. Blood-semen alignments were run using a subsitition model suggested by jmodeltest [557], employed the uncorrelated relaxed molecular clock and assumed a Bayesian Skyline population model for the tree priors of the virus population change over time. To infer the acenstral state reconstruction, Bayesian stochastic search variable selection (BSSVS) approach was used [558]. Based on the compartment (blood or semen) of origin, symmetric diffusion model was used which allow equal probability of movement of the virus between blood and semen compartments. These parameters were specified within the XML file generated by Bayesian Evolutionary Analysis Utility (BEAUTI). Five hundred million generations were run, and trees sampled at every 100,000th tree and convergence inspected in Tracer [559] for Estimated sample size (ESS) of above 200. The Maximum Clade Credibility trees were then constructed from the sampled trees using TreeAnnotator. If the ESS were below 200 i.e. convergence was not reached, the analysis was repeated again using the same XML file. The logs from the replicates were then combined using LogCombiner and resampled at a lower frequency. The convergence was then inspected in Tracer and if achieved, the trees were sampled using the same frequency and MCC tree constructed as earlier explained. The MCC trees were edited and visualized in FigTree [560]. Location (compartment) state randomization was done by introducting a tip-swap operator in the XML file and repeating the Bayesian analysis and indicated above. The inferred location state of the most recent common ancestor should be purely as a result of sampling bias when tip-swap is performed. The root state posterior probabilities of each compartment was then plotted against those of the tip-swap analysis.

2.3.10 Compartmentalization analysis

The PID-consensus sequences were used in the evaluation of compartmentalization. To prevent sampling bias, equal number of sequences were first randomly selected from blood and semen. To do this, the sequences from the compartment (blood or semen) with the smaller number of sequences were retained. The same number was then randomly sampled from the other compartment with high number of PID-consensus sequences using a python library script. Compartmentalization was assessed by using three statistical-based tools. Two distance-based methods, nearest neighbour statistic (S_{nn}) [561], Wright's Measure of

Population Subdivision (F_{ST}) [562, 563] and one tree-based method Slatkin Maddison Test (SMT) [564]. These methods are all implemented in a software package called HYPHY [565]. The F_{ST} computes mean genetic distances between and within the compartments using the Tamura Nei-93 algorithm and computes the F-statistic. The closer to 1 the F-statistic is, the more compartmentalized the viral populations and vice versa. The S_{nn} statistic compares the observed mean genetic distances with a hypothetical randomly distributed neighbour. On the other hand, SMT counts the number of inferred migration events between the compartments and compares this count with what would be expected in a fully mixed (panmictic) population.

2.3.11 Distribution of identical sequences in the phylogenetic trees

To study the distribution of variants in blood and genital compartment, phylogenetic trees were first constructed from the PID-consensus sequences using RaxML [566]. An R-based package (Analyses of Phylogenetics and Evolution, "APE") [567] was used to extract the terminal node leaves of the phylogenetic trees. The number of blood and semen sequences from nodes with 10 or more sequences were then counted and tabulated in bubble plots.

2.3.12 Clonal amplification

Clonal amplification analysis was performed from the haplotypes constructed after collapsing the PID-consensus sequences at 99.75% similarity using Vsearch [555]. The frequency was tagged to the name of the haplotype (variant) which allowed to determine the minor and major variants. This way clonal expansion of a single cell was accounted for. The 99% frequency was also utilized to draw bubble plot trees with the size of the bubble proportional to the number of sequences used to make that particular variant. RaxML [566] was used for constructing trees while python library ete3 [568] was used to plot the bubble trees.

2.3.13 Diversity analysis

From the unique sequences made by collapsing identical sequences, HIV-1 diversity between blood and semen sequences was analysed. The median pairwise genetic distances were then determined using Molecular Evolutionary Genetics Analysis (MEGA) software version 7 [569].

2.3.14 Inference on migration events between blood and semen

We estimated the minimum number of viruses moving between the blood and semen compartments after normalizing for equal numbers of sequences using the Slatkin-Maddison test which allows the counting of the number of inferred migration events from blood to semen and vice versa. The difference between the migration events from blood to semen and semen to blood were analysed using non-parametric Wilcoxon signed rank test. The inferred migration events from SMT have previously been shown to correlate with Bayesian Markov Jump counts [303].

2.3.15 Statistical Analyses

Data was analysed using non-parametric methods as implemented in GraphPad prism 8 (La Jolla California USA, www.graphpad.com) or Stata statistical software v15 (College Station, Tx: StataCorp LLC)

Results

2.3.16 Study participants

A total of 44 HIV-1 infected therapy naïve men were recruited for the study who donated a paired blood and semen samples. Their characteristics are summarized in Table 2 . The median $\log_{10}$ viral load in the blood was 4.10 cps/ml (IQR 2.69-4.56) while that in semen was 2.72 cps/sample (IQR 1.3-3.71). Viral load in blood and semen were weakly associated (adjusted $R^2 = 0.28$, $p < 0.0005$) and semen viral loads were generally detected when the blood viral load was above 10,000 copies/ml (Figure 16 A). Of the 21/44 study participants (48%) who had detectable viral loads in both semen and blood, 11 were picked for deep sequencing (Figure 16 B). There was no difference in median age for the 11 donors (median = 28yrs, IQR 26-33yrs) versus those that were not deep-sequenced (median = 29yrs, IQR 25-35yrs) (Mann-Whitney test p = 0.792). The $CD4^+$ T cell counts for the 11 selected study subjects (median =

390 cells/µl (IQR 233-467 cells/µl)) was slightly lower than that of the entire population (median = 514 cells/µl (IQR 371-626 cells/µl) (Mann-Whitney test p = 0.0416). This is probably due to the choice of study participants who had higher viral loads in semen who by association also had higher viral loads in blood (Figure 16 B). Blood viral load is negatively associated with $CD4^+$ T cell count in HIV-1-infected adults [570, 571].

Table 2: Study participants details.

SampleID	Age (years)	CD4 count (cells/ml)	Blood viral load (cps/ml)	Semen viral load (cps/sample)	Reported time since known to be HIV positive (yrs)	Recent STI history
SVB001	43	494	21993	4082.5	6	none
SVB002	37	743	28001	<50	15	none
SVB003	34	1123	20	482.5	6	none
SVB005	29	622	20289	890	5	none
SVB006	55	727	1648	<50	5	none
SVB007	37	200	42072	915	7	Syphilis treated >1yr ago, Genital warts treated >1 yr ago
SVB008	42	257	309388	38640	5	none
SVB009	29	855	20	<50	5	Syphilis treated >1 yr ago
SVB010	46	639	13993	<50	4	Gonorrhoea treated >1 yr ago
SVB012	27	586	145745	6365	2	Gonorrhoea and Chlamydia both treated more than 1 yr ago
SVB013	25	603	31838	2152.5	2	Gonorrhoea and Chlamydia both treated >1 year ago
SVB014	52	615	3637	<50	3	none
SVB015	34	668	12500	<50	2	Syphilis
SVB016	32	960	81	<50	1	Genital Herpes treat<1yr
SVB017	20	491	4070	<50	1	none
SVB018	27	669	692	<50	2	Gonorrhoea treated <1yr ago
SVB019	26	519	114	<50	1	Unknown
SVB020	25	783	30261	79247.5	4	Genital warts
SVB021	34	148	183339	4435	1	Syphilis-not treated
SVB022	57	264	9369	420	1	Gonorrhoea - on treatment
SVB023	20	460	20889	<50	0	Unknown
SVB024	25	425	448	<50	4	Unknown
SVB025	25	368	440969	23357.5	0.25	Unknown
SVB026	31	209	46202	4842.5	1	Gonorrhoea-treated >yr ago
SVB027	21	543	41616	1985	1	Unknown
SVB029	28	149	20947	273305	2	Unknown
SVB030	27	390	25781	235710	0.08	Unknown
SVB032	29	505	11624	450	10	Unknown
SVB033	0	372	81702	7810	0	Genital warts
SVB034	24	589	333	1157.5	0	Genital Herpes HSV
SVB036	34	147	57052	1152.5	7	Syphilis - on treatment
SVB037	32	789	115	<50	3	Unknown
SVB038	25	1104	11073	3232.5	1	Genital warts

SVB039	18	425	43191	132605	1	none
SVB040	24	542	1809	<50	1	syphilis, gonorrhoea, genital warts
SVB041	25	567	89090	27540	0.08	none
SVB042	50	403	20	530	14	none
SVB043	59	509	122297	5556597.5	2	none
SVB044	32	152	2455	<50	15	none
SVB045	29	300	20691	5470	1	none
SVB046	23	529	529	<50	1	none
SVB047	20	325	6871	<50	0	Genital warts
SVB048	47	570	348	<50	1	Gonorrhoea
SVB049	30	423	85711	7915	1	none

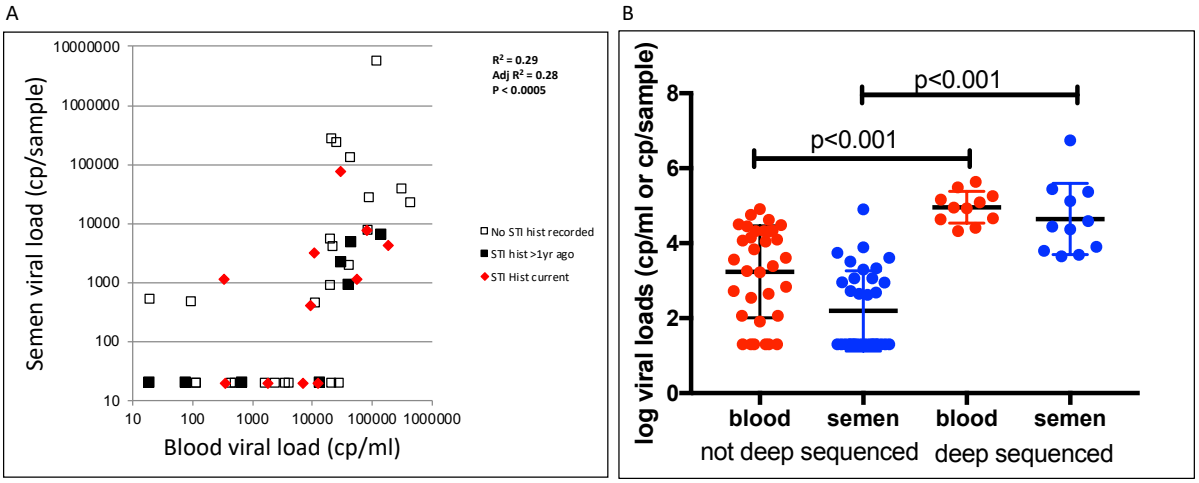


Figure 16: The viral loads from the 44 study participants (A). Blood viral loads were weakly associated with the semen viral loads (Adjusted R^2 = 0.28, p < 0.0005). In addition, viral load in blood was generally only detectable when the viral loads in semen were above 10,000 copies/sample with three donors having detectable viral loads in semen but undetectable or low in blood (A). Eleven participants who had sufficient viral load in semen were selected for deep sequencing (B).

2.3.17 Relationship between blood and semen viral loads and sexually transmitted infections (STIs)

There was no difference in viral loads between semen and blood for participants who recorded to have a concomitant or recent sexually transmitted infection (STI < 1 yr) or had an STI treated more than a year before (STI > 1 yr) (Figure 17). The $CD4^+$ T cell count for the subjects with concomitant or recent STI did not differ from those without a concomitant STI as well (Mann-Whitney test p = 0.6435).

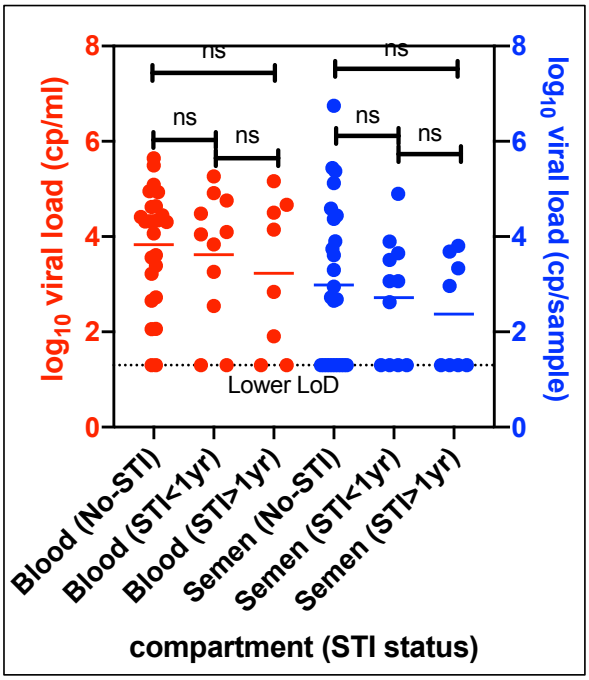


Figure 17: Relationship between viral loads in blood and sexually transmitted infections (STIs).

2.3.18 Deep sequencing output and selection of sequences for compartmentalization evaluation

A total of 22352 PID-consensus sequences were obtained after processing and clean-up of the sequences (median = 1130 (IQR = 481-1257 per compartment and donor) (Table 3). There were more PID-consensus sequences from blood than those from semen (median blood = 1259 (IQR 1130 – 1699 sequences), median semen = 448, (IQR 206 – 1039 sequences, p = 0.004). Consequently there were more unique variants in blood than in semen (median blood = 319 (IQR 259 – 599 sequences) vs median semen = 127 (IQR 63 – 183 sequences) (Table 3).

Table 3: Summary table with the PID-consensus sequences obtained after sequencing and sequence processing. The table also indicates the number of unique sequences obtained from each compartment as well as the sequences used to analyse compartmentalization.

SampleID	PID-consensus sequences	Number used for compartmentalization analysis	Unique sequences
SVB008_BP	1643	448	848
SVB008_SP	448	448	160
SVB012_BP	1139	81	286
SVB012_SP	81	81	26
SVB021_BP	1645	165	667
SVB021_SP	165	165	45
SVB025_BP	1753	246	164
SVB025_SP	246	246	52
SVB026_BP	2209	394	919
SVB026_SP	394	394	96
SVB029_BP	580	580	256
SVB029_SP	892	580	255
SVB030_BP	1259	1201	378
SVB030_SP	1201	1201	247
SVB039_BP	2234	1193	351
SVB039_SP	1193	1193	191
SVB041_BP	1121	912	267
SVB041_SP	912	912	140
SVB043_BP	707	707	185
SVB043_SP	1165	707	113
SVB049_BP	1251	114	258
SVB049_SP	114	114	19
Total	**22352**	**12082**	**5923**
Median (IQR)	**1130 (481 – 1257)**	**448 (185 – 861)**	**219 (133 – 302)**
Median blood (IQR)	**1259 (1130 – 1699)**		**319 (259 – 599)**
Median Semen (IQR)	**448 (206 - 1039)**		**127 (63 – 183)**

2.3.19 Compartmentalization was found in all the study subjects

Compartmentalization analysis was done using three statistical techniques with same number of PID-consensus sequences from each compartment (average = 448 sequences). Compartmentalization was said to occur when all the three standard tests reached significance ($p < 0.05$). Compartmentalization was observed in all the 11 study participants (Figure 18). However, the severity of the compartmentalization varied among the 11 donors. Donor SVB043 had the most severe blood and genital tract viral compartmentalization while SVB041 had the least, as measured by Wright's F_{ST} and S_{nn} (Figure 18 A,B). There was no apparent association between the severity of compartmentalisation and either viral load or $CD4^+$ T cell count (Figure 18 D).

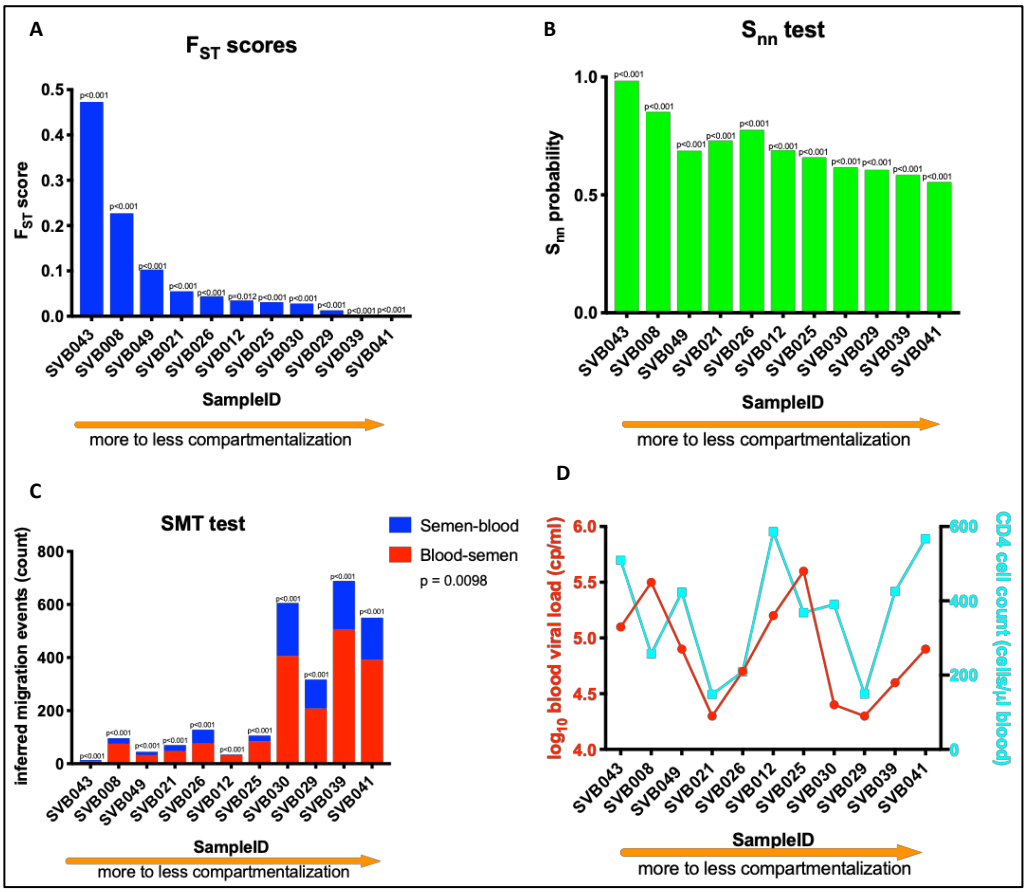
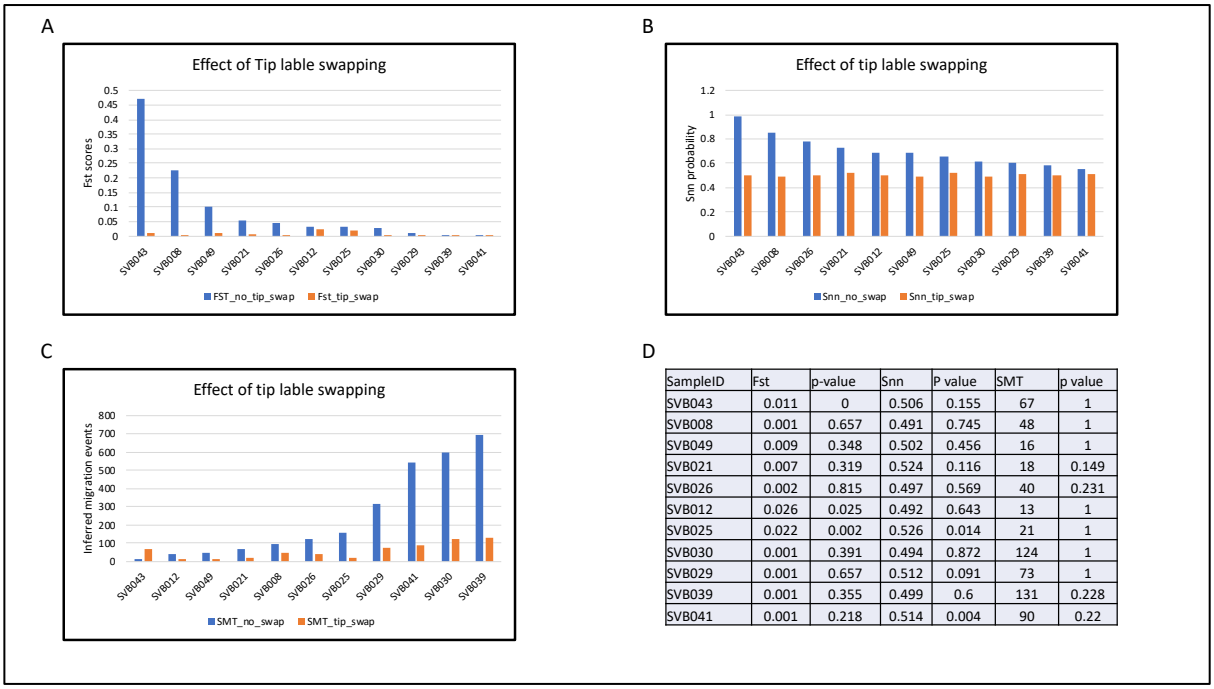


Figure 18: Compartmentalization of viral populations between blood and the genital compartment was observed in all the study participants using three statistical techniques; Wright's F_{ST} (A), Nearest neighbour statistic (S_{nn}) (B) and Slatkin Maddison test (SMT) (C). The extent of compartmentalization varied among the donor. Figure C also shows the number of inferred migration events from blood to semen and vice versa. There were more inferred migration events from blood to

semen than from semen to blood (Mann-Whitney p = 0.0098). There was no relationship between extent of viral compartmentalization between blood and the male genital tract with the viral loads or CD4⁺ T cell counts.

To investigate whether sample bias could be the cause of the observed compartmentalization in all the 11 donors, tip-label randomization analysis was performed [572, 573]. This was done using a Python library script that shuffled the sequences names 100 times and then randomly reallocated them to two groups in the same proportion as there were blood vs semen sequences in that sequence set. The resulting alignment was then used to test compartmentalization using the three statistical tests. This test is intended to reveal sampling bias in the structure of the overall sequence set or its tree that might erroneously score as compartmentalisation. None of the tip swapped sequence sets of any of the donors achieved compartmentalization (i.e. no significant p values in all the three tests) (Figure 19). This shows that sampling bias was highly unlikely to explain the compartmentalization signal that was observed among the donors studied.



SampleID	Fst	p-value	Snn	P value	SMT	p value
SVB043	0.011	0	0.506	0.155	67	1
SVB008	0.001	0.657	0.491	0.745	48	1
SVB049	0.009	0.348	0.502	0.456	16	1
SVB021	0.007	0.319	0.524	0.116	18	0.149
SVB026	0.002	0.815	0.497	0.569	40	0.231
SVB012	0.026	0.025	0.492	0.643	13	1
SVB025	0.022	0.002	0.526	0.014	21	1
SVB030	0.001	0.391	0.494	0.872	124	1
SVB029	0.001	0.657	0.512	0.091	73	1
SVB039	0.001	0.355	0.499	0.6	131	0.228
SVB041	0.001	0.218	0.514	0.004	90	0.22

Figure 19: Tip-label randomization test. Swapping the tip label designations to investigate if sampling bias was responsible for the observed viral compartmentalization in all donors. None of the donors achieved compartmentalization. Overall this shows that sampling bias was not the cause of the compartmentalization that was observed among these donors.

We noticed that Slatkin Maddison (SM) method was particularly sensitive to the number of sequences analyzed. To control potential bias due to the number of sequences analysed, a

single compartment with the highest number of sequences was tested for statistical compartmentalization using the SM method. First, sequences from the largest compartment were split randomly into two sets. One hundred of such subsamples were done and each time compartmentalization was assessed using SM method. Since the sequences comes from the same compartment, theoretically, compartmentalization is not supposed to be observed. Therefore, the rate of false positive was determined from the 100 subsamples. None of the 100 subsamples showed compartmentalization for all the 11 donors Table 4.

Table 4: Number of migration events and p-values from Slatkin-Method test of compartmentalization using 100 subsamples from the same compartment

	SVB008		SVB012		SVB021		SVB025		SVB026		SVB029		SVB030		SVB039		SVB041		SVB043		SVB049	
subsample	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value	migration count	p-value
1	711	1	517	1	519	1	837	0.998001998801998	977	1	398	1	572	1	1035	1	510	1	557	0.998001998801998	579	1
2	708	1	509	0.997002997002997	510	1	835	0.993006993006993	978	1	390	1	562	1	1045	1	522	1	557	1	576	1
3	721	1	519	1	503	1	837	1	968	1	400	1	561	1	1051	1	522	1	543	0.877122877122772	586	1
4	701	1	517	0.999000999000999	510	1	840	0.999000999000999	961	1	395	1	561	1	1044	1	509	1	561	1	578	1
5	706	1	521	1	513	1	843	1	965	1	389	0.994005994005994	549	0.975024975024975	1042	1	521	1	547	0.800199800199800	591	1
6	716	1	524	1	501	1	843	1	970	1	391	0.995004995004995	560	1	1053	1	509	0.999000999000999	546	0.865134865134865	571	0.995004995004995
7	699	1	515	0.999000999000999	520	1	840	1	976	1	393	1	549	0.995004995004995	1058	1	515	1	557	0.993006993006993	572	0.995004995004995
8	711	1	528	1	513	1	837	1	964	1	393	1	560	1	1025	0.938061938061938	510	0.995004995004995	544	0.953046953046953	582	1
9	706	1	526	1	513	1	825	0.965034965034965	970	1	396	0.999000999000999	561	1	1045	0.999000999000999	517	1	553	0.997002997002997	580	1
10	700	1	518	1	512	1	846	1	973	1	383	0.980019980019980	572	1	1058	1	518	1	558	0.994005994005994	576	1
11	716	1	516	1	513	1	835	0.998001998001998	971	1	402	1	576	1	1028	0.988011988011988	519	1	538	0.746253746253746	570	0.999000999000999
12	705	1	524	1	509	1	834	0.997002997002997	970	1	396	1	565	1	1047	1	513	1	553	0.996003996003996	578	0.999000999000999
13	710	1	517	0.999000999000999	508	1	842	1	974	1	401	1	578	1	1030	0.994005994005994	516	0.998001998001998	554	0.998001998001998	575	0.998001998001998
14	719	1	519	0.999000999000999	519	1	840	1	969	1	394	1	572	1	1051	1	508	0.991008991008991	552	0.975024975024975	567	0.991008991008991
15	712	1	514	1	511	1	837	1	960	1	402	1	567	1	1048	1	516	1	551	0.985014985014985	576	1
16	715	1	519	1	515	1	843	1	960	1	400	1	563	1	1030	0.975024975024975	515	1	544	0.798201798201798	585	1
17	700	1	523	1	505	0.999000999000999	833	0.997002997002997	969	1	396	1	569	1	1036	0.999000999000999	518	1	552	0.978021978021978	564	0.981018981018981
18	709	1	506	0.984015984015984	509	1	834	0.995004995004995	952	1	403	1	568	1	1043	1	498	0.958041958041958	553	0.991008991008991	569	0.996003996003996
19	692	1	517	1	493	0.997002997002997	813	0.458541458541458	959	1	405	1	559	1	1059	1	517	1	558	0.999000999000999	574	0.997002997002997
20	706	1	511	0.995004995004995	514	1	828	0.975024975024975	951	1	405	1	569	1	1050	1	517	1	564	1	579	1
21	714	1	500	0.974025974025974	502	0.999000999000999	846	1	977	1	397	1	570	1	1061	1	517	1	553	0.992007992007992	584	1
22	714	1	517	1	520	1	836	0.999000999000999	960	1	400	1	564	1	1058	1	517	1	555	0.993006993006993	573	1
23	715	1	514	1	499	0.996003996003996	833	0.999000999000999	974	1	404	1	559	0.999000999000999	1048	1	518	1	537	0.454545454545545	565	0.999000999000999
24	714	1	521	1	494	0.992007992007992	836	0.998001998001998	961	1	389	1	572	1	1046	1	520	1	544	0.895104895104895	573	0.999000999000999
25	716	1	502	0.968031968031968	511	1	848	1	958	1	398	1	557	0.999000999000999	1028	0.987012987012987	508	0.997002997002997	558	0.999000999000999	580	1
26	707	1	518	0.997002997002997	510	1	830	0.987012987012987	971	1	395	1	560	0.999000999000999	1039	0.996003996003996	516	1	547	0.984015984015984	578	1
27	718	1	502	0.988011988011988	497	0.998001998001998	830	0.974025974025974	954	1	395	0.999000999000999	575	1	1043	1	518	1	553	0.986013986013986	555	0.803198031968032
28	715	1	518	1	516	1	823	0.880119880119881	967	1	393	0.999000999000999	564	1	1047	1	513	1	555	0.998001998001998	572	1
29	720	1	512	0.999000999000999	488	0.962037962037962	837	1	956	1	397	0.998001998001998	560	0.999000999000999	1037	0.982017982017982	505	0.985014985014985	548	0.967032967032967	569	0.991008991008991
30	702	1	510	0.986013986013986	508	1	836	0.999000999000999	971	1	395	1	567	1	1053	1	516	1	561	1	572	0.999000999000999
31	707	1	522	1	506	1	836	0.999000999000999	949	1	390	0.999000999000999	558	1	1051	1	517	1	559	1	585	1
32	709	1	513	0.997002997002997	512	1	820	0.734265734265743	972	1	397	1	575	1	1050	1	508	0.995004995004995	556	0.999000999000999	575	1
33	695	1	514	0.999000999000999	509	1	833	0.999000999000999	968	1	400	1	556	1	1054	1	519	1	558	0.999000999000999	564	0.992007992007992
34	710	1	519	1	520	1	843	1	968	1	396	1	569	1	1046	1	505	0.985014985014985	554	0.998001998001998	579	1
35	698	1	510	0.998001998001998	506	1	831	0.989010989010989	975	1	392	0.997002997002997	577	1	1044	0.999000999000999	507	0.998001998001998	545	0.940059940059401	583	1
36	712	1	506	0.984015984015984	506	1	825	0.914085914085914	983	1	393	1	566	1	1037	0.999000999000999	518	0.999000999000999	548	0.885114885114885	574	0.995004995004995
37	708	1	524	1	512	0.999000999000999	836	0.999000999000999	963	1	408	1	558	1	1059	1	520	1	540	0.748251748251748	579	1
38	712	1	518	1	510	1	838	0.997002997002997	963	1	396	1	566	1	1058	1	512	0.999000999000999	552	0.995004995004995	575	0.998001998001998
39	710	1	514	1	507	1	843	1	970	1	400	1	568	1	1046	1	518	1	548	0.921078921078921	589	1
40	721	1	522	1	509	1	830	0.989010989010989	973	1	392	0.999000999000999	573	1	1054	1	509	1	541	0.849150849150849	579	1
41	725	1	520	1	520	1	833	1	968	1	394	0.999000999000999	569	1	1047	1	519	1	548	0.959040959040959	577	1
42	718	1	497	0.955044955044955	513	1	832	0.994005994005994	965	1	383	0.994005994005994	572	1	1041	0.999000999000999	521	1	546	0.876123876123876	571	1
43	725	1	520	1	501	0.999000999000999	832	0.994005994005994	968	1	408	1	574	1	1051	1	510	0.991008991008991	562	1	570	0.998011988011988
44	713	1	520	1	501	1	838	0.999000999000999	958	1	397	1	561	0.999000999000999	1047	1	509	0.999000999000999	554	0.99000999000999	582	1
45	719	1	515	1	500	0.997002997002997	825	0.875124875124875	968	1	391	1	568	1	1052	1	513	0.999000999000999	545	0.838161838161838	567	0.992007992007992
46	717	1	513	0.996003996003996	502	1	841	1	968	1	399	1	554	0.999000999000999	1039	1	511	1	543	0.952047952047952	572	1
47	697	1	518	0.999000999000999	505	1	822	0.923076923076923	965	1	399	1	553	0.996003996003996	1045	1	487	0.742257742257422	563	0.999000999000999	572	0.998001998001998
48	712	1	518	1	508	1	828	0.943056943056943	969	1	400	0.998001998001998	565	1	1040	0.998001998001998	516	1	557	0.999000999000999	569	0.997002997002997
49	708	1	524	1	494	0.987012987012987	837	0.997002997002997	972	1	404	1	559	1	1046	1	521	1	547	0.823176823176823	583	1
50	704	1	518	1	505	1	833	0.999000999000999	964	1	399	1	567	1	1041	1	521	1	561	0.999000999000999	578	1
51	709	1	526	1	518	1	846	1	982	1	403	1	563	0.999000999000999	1047	1	524	1	547	0.935064935064935	578	1
52	703	1	505	0.993006993006993	507	0.998001998001998	825	0.935064935064935	975	1	395	0.999000999000999	564	1	1045	1	517	1	548	0.918081918081918	570	0.997002997002997
53	712	1	508	1	508	0.999000999000999	838	1	969	1	401	1	548	0.995004995004995	1060	1	519	1	549	0.943056943056943	568	0.998001998001998
54	707	1	525	1	509	1	833	0.999000999000999	976	1	398	1	582	1	1045	1	511	0.999000999000999	545	0.849150849150849	581	1
55	706	1	521	1	515	1	837	1	957	1	387	0.999000999000999	568	1	1063	1	519	1	547	0.951048951048951	579	1
56	705	1	525	1	506	1	843	1	953	1	399	1	556	0.999000999000999	1038	1	516	1	551	0.973026973026973	554	0.817182817182172
57	722	1	514	1	518	1	841	1	966	1	395	0.999000999000999	561	1	1049	1	511	1	551	0.976023976023976	571	1
58	728	1	521	1	493	1	838	1	966	1	389	0.993006993006993	573	1	1026	0.956043956043956	522	1	557	0.997002997002997	575	0.998001998001998
59	707	1	510	0.999000999000999	517	1	829	0.996003996003996	972	1	398	1	562	1	1049	1	512	0.997002997002997	562	1	579	1
60	711	1	510	0.996003996003996	504	1	837	1	958	1	382	0.995004995004995	570	1	1020	0.845154845154852	519	1	558	0.999000999000999	564	0.999000999000999
61	711	1	522	1	495	0.999000999000999	838	1	975	1	395	1	569	1	1036	0.994005994005994	519	1	538	0.628371628371628	573	0.999000999000999
62	712	1	511	0.997002997002997	515	1	846	1	965	1	389	0.991008991008991	552	0.988011988011988	1058	1	521	1	559	1	582	1
63	713	1	515	0.998001998001998	507	1	844	1	954	1	405	1	559	0.998001998001998	1046	1	513	0.994005994005994	552	0.958041958041958	579	1
64	694	1	514	0.999000999000999	511	1	818	0.770229770229702	968	1	384	0.994005994005994	559	1	1043	1	514	1	544	0.806193806193806	585	1
65	700	1	522	1	512	1	841	0.999000999000999	956	1	389	0.999000999000999	563	1	1038	0.998001998001998	512	0.999000999000999	553	0.976023976023976	567	0.984015984015984
66	722	1	514	0.997002997002997	510	1	833	0.981018981018981	973	1	393	1	572	1	1044	1	517	1	555	1	572	0.996003996003996
67	704	1	532	1	500	0.998001998001998	836	0.999000999000999	981	1	399	1	571	1	1045	1	491	0.821178211788211	558	1	578	1
68	699	1	509	1	510	1	827	0.950049950049951	985	1	400	1	555	0.997002997002997	1010	0.886113886113886	522	0.999000999000999	553	0.999000999000999	575	0.998001998001998
69	712	1	517	0.999000999000999	504	1	844	1	960	1	392	1	550	0.998001998001998	1036	1	505	0.999000999000999	546	0.929070929070929	577	1
70	712	1	523	1	501	0.998001998001998	833	1	958	1	401	1	554	0.998001998001998	1037	0.999000999000999	510	0.998001998001998	550	0.961038961038961	576	0.999000999000999
71	705	1	515	1	513	1	844	1	983	1	405	1	552	0.998001998001998	1057	1	502	0.960039960039601	559	0.999000999000999	579	1
72	697	1	512	1	510	1	822	0.951048951048951	965	1	398	1	572	1	1051	1	517	1	552	0.996003996003996	565	0.975024975024975
73	718	1	514	1	511	1	817	0.607392607392674	965	1	394	1	561	1	1060	1	515	1	547	0.974025974025974	577	1
74	723	1	511	0.998001998001998	520	1	828	0.980019980001998	974	1	402	1	562	1	1047	0.999000999000999	498	0.955044955044955	557	0.996003996003996	590	1
75	716	1	522	1	516	1	827	0.940059940059401	970	1	388	1	578	1	1050	0.999000999000999	512	0.997002997002997	553	0.985014985014985	584	1
76	717	1	517	0.998001998001998	512	1	840	1	974	1	404	1	569	1	1035	0.999000999000999	526	1	552	0.991008991008991	576	1
77	714	1	511	0.999000999000999	504	1	844	1	957	1	392	1	549	0.996003996003996	1057	1	508	0.997002997002997	562	1	577	1
78	720	1	522	1	513	1	823	0.939060939060939	941	1	397	0.999000999000999	566	1	1018	0.948051948051948	522	1	552	0.959040959040959	582	1
79	708	1	504	0.919080919080919	510	1	833	0.992007992007992	978	1	395	1	560	1	1053	1	500	0.976023976023976	557	0.998001998001998	572	0.997002997002997
80	712	1	507	0.992007992007992	502	1	822	0.951048951048951	965	1	398	1	572	1	1051	1	517	1	552	0.996003996003996	565	0.975024975024975
81	717	1	517	0.999000999000999	498	0.998001998001998	817	0.607392607392674	965	1	394	1	561	1	1060	1	515	1	547	0.974025974025974	577	1
82	712	1	507	0.994005994005994	502	0.999000999000999	832	0.991008991008991	976	1	398	1	560	0.999000999000999	1053	1	506	0.989010989010989	562	1	572	1
83	687	1	511	0.996003996003996	505	1	837	1	979	1	394	1	557	1	1056	1	518	0.999000999000999	545	0.936063936063936	579	1
84	720	1	510	0.997002997002997	512	1	836	1	967	1	386	0.996003996003996	556	1	1039	1	507	0.999000999000999	563	1	580	1
85	693	1	502	0.994005994005994	512	1	829	0.979020979020791	949	1	394	1	546	0.999000999000999	1061	1	510	0.999000999000999	538	0.426573426573426	557	0.885114885114885
86	725	1	511	0.994005994005994	498	0.960303969309691	835	1	969	1	390	0.995004995004995	563	1	1049	1	512	1	553	0.987012987012987	581	1
87	706	1	528	1	519	1	827	0.975024975024975	966	1	398	1	575	1	1062	1	520	1	553	0.982017982017982	579	1
88	726	1	520	1	512	1	834	0.993006993006993	973	1	393	0.998001998001998	565	1	1063	1	502	0.993006993006993	553	0.992007992007992	571	1
89	729	1	518	1	489	0.995004995004995	837	1	948	1	400	1	562	1	1049	1	504	0.983016983016983	545	0.974025974025974	576	1
90	713	1	511	1	520	1	830	0.972027972027972	976	1	396	1	567	1	1053	1	514	1	555	0.996003996003996	565	1
91	702	1	519	0.999000999000999	507	1	842	1	961	1	396	1	565	0.999000999000999	1060	1	516	1	554	0.996003996003996	562	0.994005994005994
92	701	1	513	1	511	1	829	0.973026973026973	965	1	390	0.998001998001998	571	1	1057	1	515	0.999000999000999	557	0.997002997002997	584	1
93	718	1	522	1	512	1	833	0.997002997002997	986	1	400	1	556	1	1037	0.993006993006993	527	1	544	0.724275724275724	580	0.999000999000999
94	707	1	512	0.999000999000999	504	0.999000999000999	840	1	966	1	393	0.998001998001998	557	0.997002997002997	1046	1	513	0.999000999000999	557	0.999000999000999	564	0.972022797022797
95	688	1	530	1	500	1	824	0.841158841158412	971	1	399	1	563	1	1059	1	508	1	553	0.988011988011988	579	1
96	697	1	518	1	507	1	834	0.994005994005994	973	1	399	1	555	0.999000999000999	1059	1	507	0.995004995004995	552	0.976023976023976	579	1
97	702	1	509	0.996003996003996	508	1	821	0.728271728271728	969	1	399	1	572	1	1049	1	502	0.976023976023976	549	0.983016983016983	555	0.776223776223763
98	717	1	517	1	505	0.997002997002997	821	0.845154845154852	957	1	392	0.998001998001998	559	0.999000999000999	1055	1	508	1	556	0.999000999000999	568	0.991008991008991
99	715	1	521	1	514	1	836	0.999000999000999	976	1	397	1	556	0.999000999000999	1045	0.999000999000999	523	1	547	0.899101089911089	578	1
100	707	1	519	1	505	1	837	1	978	1	405	1	578	1	1028	0.962037962037962	508	0.999000999000999	545	0.861138861138612	579	1
Percentage false postive		0		0		0		0		0		0		0		0		0		0		0

2.3.20 Some clusters were heavily compartmentalized even in minimally compartmentalized donors

Some previous studies had relied on visual inspection of phylogenetic trees to evaluate viral compartmentalization between blood and other anatomical sites [377, 380, 394, 529, 541]. However mere inspection of trees may not be entirely sensitive or accurate in detecting compartmentalization; especially when only some of the variants in anatomical tissues are compartmentalized [191, 277]; and when a very large number of sequences are analysed. In this report, phylogenetic trees revealed large clusters within blood and semen that had identical sequences within them (Figure 20). Due to the huge amout of high thoughput data, other tree annotations were not possible. But the clustering of blood and semen sequences is clearly visible with SVB043 showing severe phylogenetic viral compartmentalization between blood and semen sequences.

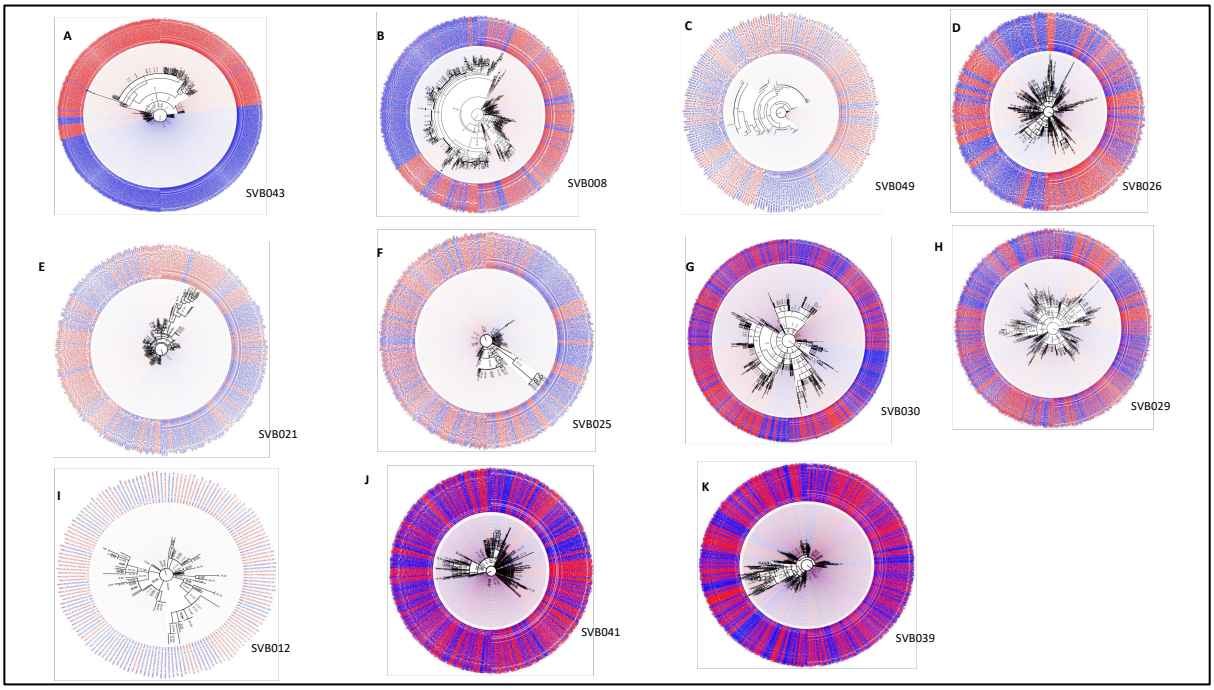


Figure 20: phylogenetic trees showing viral compartmentalization between blood and the male genital compartment. clusters of either purely blood and semen are visible from the trees. Red represents PID-consensus sequences from blood while blue represent those from semen.

To further understand the nature of blood and semen sequences within these clusters, the distribution of blood and semen sequences within the clusters of the phylogenetic trees was analysed by counting the proportion of blood and semen sequences within them. To have a

meaningful result, only clusters carrying more than 10 sequences were considered and bubble plots were drawn (Figure 21). Donors SVB021 and SVB026 did not have such clusters, so are not shown. Analysis revealed that some viral variants do not traffic well even among the minimally compartmentalized donors (Figure 21). For donor SVB043, who exhibited severe compartmentalization, the clusters were made of purely or almost purely blood or semen sequences. In donors with lesser extent of compartmentalization, the majority of the clusters contained roughly equal number of blood and semen sequences. However, there were some clusters that had only blood or semen sequences (Figure 21). This suggest that some variants do not traffic well between the compartments and rather remain in the compartment of their production. These localized variants likely contributed to the compartmentalization signal observed within the minimally compartmentalized donors.

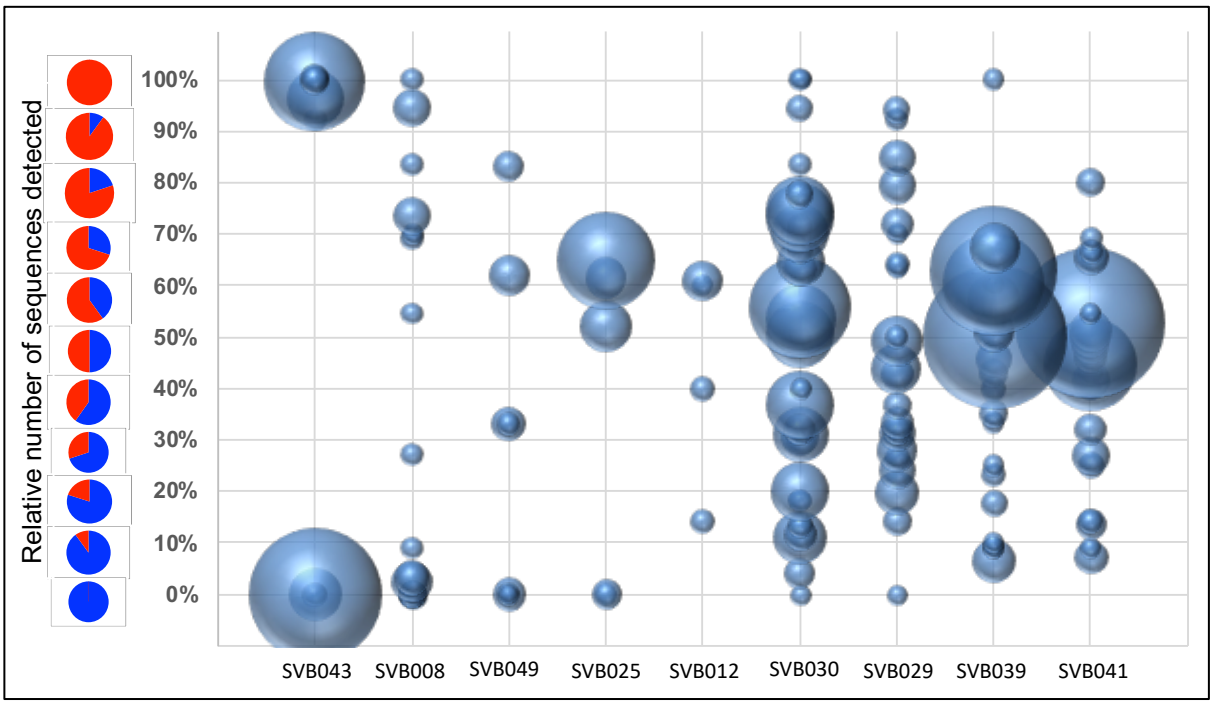


Figure 21: Analysis of clusters with identical sequences from the phylogenetic trees. The analysis revealed that some variants do not traffic well between compartments. This was so even in donors showing lesser extent of compartmentalization where some clusters contained either purely blood or semen sequences. The y-axis on the graph represent the proportion of either blood or semen within the terminal nodes of the combined tree with blood and semen sequences from each the donors on the x-axis. Only donors whose tip nodes had clusters of 10 or more sequences were picked for this analysis (Donors SVB021 and SVB026 did not have such clusters)

To address the contribution of clonal amplification to compartmenatalization, sequences from all the donors were analyzed for viral compartmentalization in three ways; (1) using equal number of PID-consensus sequences, (2) collapsing the PID-consensus sequences into

haplotypes at 100% similarity and (3) taking equal number of haplotypes. Compartmentalization was observed in all the three situations for donor SVB043 who had the highest extent of clonal amplification Table 5. This observation was also consistent for donors SVB008 and not in others Table 5.

Table 5: Role of clonal amplification in determining viral compartmentalization between blood and the male genital tract.

	primerID sequences			Haplotypes		equal number haplotypes	
SampleID	FST	Snn	SMT	FST	Snn	FST	Snn
SVB043	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SVB008	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SVB049	<0.001	<0.001	<0.001	0,429	0,999	0,245	0,554
SVB026	<0.001	<0.001	<0.001	0,864	0,994	<0.001	<0.001
SVB021	<0.001	<0.001	<0.001	0,995	0,99	0,005	<0.001
SVB012	0,012	<0.001	<0.001	0,999	0,999	0,435	0,737
SVB030	<0.001	<0.001	<0.001	0,108	0,999	<0.001	0,994
SVB025	<0.001	<0.001	<0.001	0,021	0,159	0,339	0,531
SVB029	<0.001	<0.001	<0.001	0,027	0,994	0,014	0,997
SVB039	<0.001	<0.001	<0.001	0,747	0,998	0,169	0,979
SVB041	<0.001	<0.001	<0.001	0,195	0,999	0,426	0,999

2.3.21 Model of lower depth of sampling from Illumina Miseq data

Previous studies reported viral compartmentalization between blood and the male genital tract in only about half of the donors [289, 381, 490, 529, 574]. This is a sharp contrast to the findings in this thesis where compartmentalization was found in every donor studied. These previous studies relied upon single genome analysis (SGA), for which the cost and labour-intensiveness nature of the technique limit the depth of sampling to a few sequences. It is possible that the greater depth of sequencing was responsible for the apparent discrepancy between the proportion of study subjects with compartmentalization found in this thesis vs those found in previous studies. Alternatively, it is possible that the donors analysed were fundamentally different from the majority of donors previously analysed for blood/male genital tract compartmentalization in previous studies.

To address this, a small number of sequences were randomly sampled from each compartment to simulate single genome analysis (SGA) and predict if compartmentalization

would have been detected if only 35 sequences from each compartment were sampled. This number corresponds approximately to the maximum number of amplicons per compartment used in previous studies (Table 8). This random set of sequences was resampled from each compartment 1000 times and compartmentalization analysed using Wright's F_{ST} after each time. The proportion of iterations that achieved compartmentalization was noted (Table 6). This proportion should approximate the likelihood that compartmentalization would have been observed if SGA was used instead of deep sequencing.

Only 5 out of the 11 donors exhibited a >50% likelihood that viral compartmentalization between blood and semen would have been detected had 35 sequences from each compartment used (Table 6). This proportion agrees with previous studies. Thus, there is no evidence that supports the idea that the donors studied here were fundamentally different from those studied by other researchers. The high depth of sequencing used in this thesis therefore appears sufficient to explain why compartmentalization was observed in all the 11 donors.

Table 6: Simulation of lower depth of sequencing from Illumina Miseq data. A fraction of 50% and above was considered to have a high likelihood of being scored compartmentalized when using a random set of 35 sequences.

SampleID	Compartmentalized iterations (%)
SVB043	99.1
SVB008	99.5
SVB049	82.3
SVB021	50.0
SVB026	72.6
SVB012	15.8
SVB025	6.9
SVB030	27.1
SVB029	17.6
SVB039	8.1
SVB041	0.1

2.3.22 Highly compartmentalized donors had viral populations with higher diversity in blood than in semen

The median pairwise genetic distances varied between viral populations in blood and semen, but overall there was no significant differences between blood vs genital compartment sequences (p = 0.125) (Figure 22). In contrast, donors exhibiting the highest extent of compartmentalization, SVB043, SVB008 and SVB021 had high median pairwise genetic distances in blood compared to semen (Figure 22).

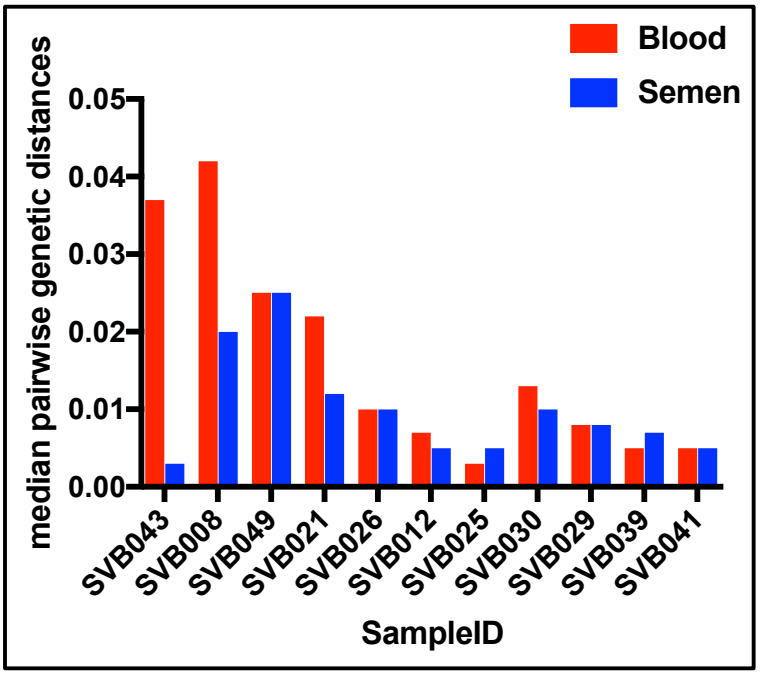


Figure 22: Comparison of median pairwise genetic distances between variants from blood vs those from the male genital compartment. Donors with highest extent of compartmentalization, SVB043, SVB008 and SVB021 had blood variants with overall higher median pairwise genetic distances compared to their semen counterparts.

2.3.23 Evidence suggesting purifying or diversifying selection

Because selective pressures experienced by the virus in different anatomical sites may differ, codon sites under purifying and diversifying selection between blood and the genital compartment of each donor were evaluated using a maximum likelihood approach; Fixed-Effect Likelihood (FEL) [575]. Though only a small section of the HIV-1 envelope (414bp, coding for 138 amino acids) was analysed, sites under selective pressures varied between the individuals (Table 7). There were more codon sites in blood than semen under both positive

(p=0.0195) and negative (p=0.002) selection suggesting higher selective pressures in blood as compared to semen by paired Wilcoxon test.

Table 7: Sites under positive and negative selective pressures in blood and seminal compartments. There were significantly mores sites under positive (p=0.0195) and negative (p=0.002) selective pressures in the blood compartment than in the semen compartment. Sites under selective pressures were assessed using Fixed Effects Likelihood (FEL) as implemented in HYPHY.

SampleID	Sites under Negative / purifying selection		Sites under Positive / diversifying selection	
	Blood plasma	Seminal plasma	Blood plasma	Seminal plasma
SVB008	51	27	13	6
SVB012	36	4	1	0
SVB021	53	16	14	0
SVB025	24	16	1	0
SVB026	70	25	3	0
SVB029	30	20	3	5
SVB030	44	29	3	0
SVB039	34	16	9	4
SVB041	36	22	0	0
SVB043	36	17	8	3

2.3.24 Movement of virus between blood and semen

Previous studies showed that there were more virus movements from blood to the genital compartment [303, 530]. This observation was confirmed in this thesis using Slatkin-Maddison test [564], where more inferred migration events were calculated from blood to semen than from semen to blood (Figure 18 C). To support the SM test method, Bayesian phylogenetics were performed and the Root State Posterior Probabilitities of the resulting trees with blood and semen sequences were determined. High RSSP indicates the origin of a particular sequence was originally found in that compartment. Donors SVB043, SVB021 and SVB021 indicated the ancestral state of semen sequences were in blood Figure 23.

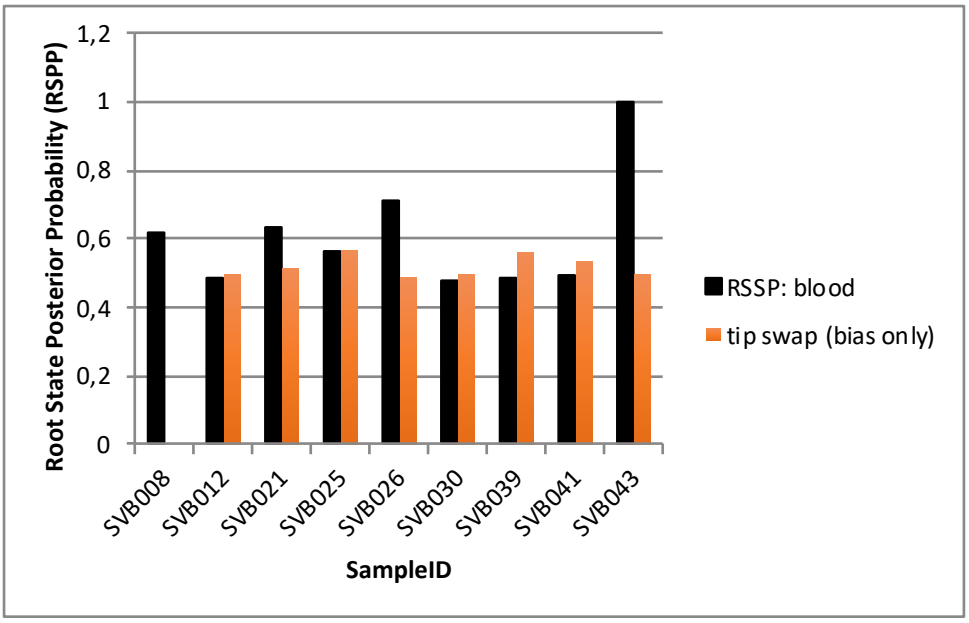


Figure 23: Root State Posterior Probabilities (RSPP) for movement of viruses from blood to semen. Donors SVB043, SVB026, SVB021 had viral sequences the moved from blood into the genital tract. Donor SVB029 and donor SVB008 tip swap trees did not converge and hence the data is not available.

2.3.25 Clonal amplification was found in all of the study participants

Clonal amplification is the expansion of low diversity variants resulting in high proportional representation of these variants. These clonally amplified variants are localized and can substantially contribute to compartmentalization. Clonal amplification was evident in all the study participants and was a characteristic of both blood and semen populations. The clonally amplified variants were represented as bubbles in phylogenetic trees (Figure 24).

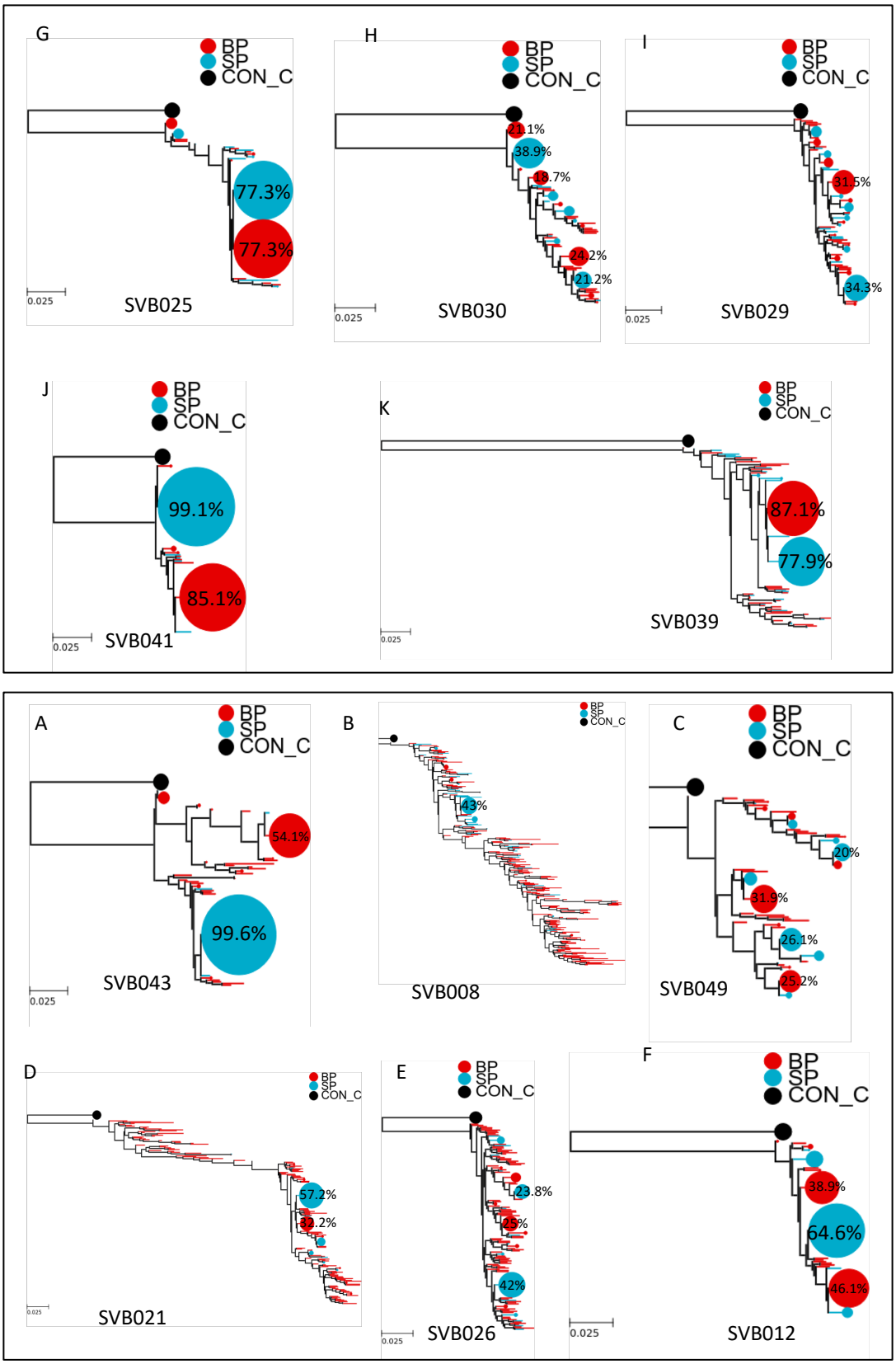


Figure 24: Clonal amplification was identified in both blood and the male genital compartment. The clonally amplified variants are represented here as bubbles, the size of the bubble being proportional to the number of sequences used to make that particular haplotype. Red bubbles represents variants from blood while green bubbles represent variants from semen. The trees are rooted by HIV-1 subtype C consensus of 2004 from Los Alamos.

2.4 Discussion

In this thesis and for the first time, viral compartmentalization between blood and the male genital tract was studied in HIV-1 chronic infections using next generation sequencing. The use of primerID and high fidelity Taq polymerase allowed us to recover the sequences as close as possible to the way they appeared prior to the PCR amplifications. PrimerID technique has previously been used to study HIV-1 populations [532-537, 576]. The utilization of PrimerID not only helped to eliminate PCR-based artefacts, but in this study it gave two extra advantages. Firstly' it allowed the use of individual PID-consensus sequences in analysis of compartmentalization as opposed to constructing haplotypes. Thus, the sequences were representative of the in vivo situation. Secondly, PID-consensus sequences allowed the study of clonal amplifications in more detail and with higher confidence.

In evaluation of compartmentalization, three statistical techniques were used, two distance-based methods, Wrights, measure of population subdivision (FST) [562, 563], and nearest neighbour statistic (Snn) [561] and one tree-based method Slatkin-Maddison Test (SMT) [564]. It is generally recommended that three techniques be used in evaluating compartmentalization [577]. Compartmentalization is said to exist when all the three techniques arrive at significant p values. When two techniques achieve significant p values, the viral populations are said to have partial compartmentalization and if only one or none achieve significant p values, the viral populations are said to be equilibrated. In this thesis, HIV-1 compartmentalization between blood and the male genital tract was found in all the study subjects.

Previous studies working on compartmentalization of HIV-1 between blood and the male genital tract found compartmentalization in only about half of the study participants (summarized in Table 8). The results in this thesis, where viral compartmentalization between blood and semen was found in all the 11 subjects, are in sharp contrast to these earlier studies. The use of next generation sequencing in this thesis as opposed to techniques used in previous studies gave the advantage of increased depth of sequencing thereby explaining the discrepancy. Only one study had previously used next generation sequencing [530]. This particular study however worked on HIV-1 compartmentalization in blood and the male

genital compartment at acute phase of infection while participants involved in this thesis came from chronic infections. This study [530] found compartmentalization in 2 of the 6 study participants at base line [530]. In addition, the study used 454 technology, which was among the first next generation technique. With 454, sequencing had to be preceded by a bulk PCR step which is known to alter the composition of viral populations through PCR recombination and amplification bias [550, 578, 579]. In this thesis, this is circumvented by using the PrimerID approach allowing use of viral sequences the way they existed in vivo.

Table 8: Summary of the studies that have previously worked on compartmentalization of viral populations between blood and the male genital compartment.

Article	Proportion of donors with compartmentalization, number (%)	Technique used	Mean number of amplicons per compartment	Region sequenced	Compartmentalization analysis
Chaillon A. (2017)	7/14 time points (50)	454	22 haplotypes	C2-V3	SMT, FST
Anderson J. A. (2010)	5/12 (42)	SGA	27	env	SMT
Chaillon A. (2014)	20/46 time points (43)	SGA, cloned	29	C2-V3	SMT, Snn, FST
Diem K. (2008)	1/2 (50)	SGA	13	C2-C5	SMT
Gupta Phalguni (2000)	3/5 (60)	bulk, cloned	10	C2-V5	phylogeny
Butler DM (2010)	2/6 (33)	SGA, cloned	20	C2-V3	FST
Pillai SK (2005)	6/12 (50)	Database	27	C2-V3	SMT
Curran R. (2002)	7/10 (70)	SGA	15	V1-V3	phylogeny
Boeras DI (2011)	1/1 (100)	SGA	14	V1-V4	SMT, Snn
Paranjpe S. (2002)	3/3 (100)	diluted, cloned	19	C2-V5	phylogeny, SMT, FST
Ghosn J. (2004)	1/3 (33)	bulk	6	protease	phylogeny
Brown RJ (2011)	3/4 (75)	SGA	7	env	phylogeny
Gianella S (2012)	7/14 (50)	Bulk, cloned	25	C2-V3	SMT, FST
This thesis	11/11 (100)	Illumina Miseq	549	V3 or C3-V5	SMT, FST,Snn

Other studies either used sequences derived from single genome amplification (SGAs) [289, 381, 490, 529, 574], cloned from bulk PCR [377, 400, 401] or tracked sequence variation using Heteroduplex Tracking Assay (HTA) [292]. Each of these techniques have their own advantages and disadvantages. The main demerit of techniques that utilize SGAs data as far

as evaluation of compartmentalization is concerned, is limited sampling of the viral populations. This results from the high cost and the labour-intensiveness nature of the technique which limits the depth of sequencing to a few sequences. In addition, low prevelance sequences in SGAs get missed due to the small sequence sample. Indeed, when a random set of 35 sequences was used in this thesis to mimic the result that would have been obtained if SGA was the technique of choice, only 5 out of the 11 had more than 50% likelihood of being reported to be compartmentalized. This identifies with what these previous studies found and underscores the importance of depth of sampling in studying HIV-1 compartmentalization. In addition, this analysis shows no obvious fundamental discrepancies from the populations studied elsewhere for the purpose of blood vs male genital tract viral compartmentalization; since the results were identifiable. Heterodruplex tracking assays (HTA) lack the sequence information and cannot be compared easily to the study in this thesis.

One potential limitation to our study is the choice of samples that had sufficient viral load in semen for deep sequencing (above 2500 copies). Coincidently, these also had high viral load in blood because the viral load in the two compartments were weakly associated [205]. However, two of the study participants (SVB012 and SVB026) who had the lowest viral loads in semen did not show obvious differences in compartmentalization of viral populations between blood and semen. Also on performing the tip swapping [572, 580], none of the donors including these two achieved compartmentalization (i.e. the three statistical tests achieving significant p values). This means that these two were not fundamentally different from the rest of the donors studied.

Clonal amplification was observed in all of the study participants and in blood and in semen. Clonally amplified variants have a potential of being transmitted as evidenced in one study working on transmission of HIV-1 from infected mothers to infants [581]. This is possibly due to relative fitness of these variants in the anatomical sites where they achieve clonal amplification. Findings in this thesis of high clonal amplification in the male genital tract, are in line with the observations of Boeras *et al* , who showed that up to 86% of the genital tract variants were found in clusters of (near) identical sequences while the rest were intermingled with blood in the phylogenetic trees [490]. In addition, data in this thesis add to the work of

Anderson *et al* [381] who observed clonal amplification in semen in 8 out of the 12 study participants. However, unlike Anderson *et al* where only 1 out of the 12 study subjects had clonal amplification in blood, all of the donors studied as part of this thesis showed clonal amplification in blood as well. This is possibly due to higher depth of sequencing afforded in this report. These clonally amplified variant could contribute to compartmentalization. Indeed, deconvoluting the phylogenetic trees and analysing the cluster composition at the terminal nodes, revealed that some clusters in minimally compartmentalized donors had purely blood or semen sequences. This suggest that some viral quasispecies don't move even in donors with little extent of compartmentalization and might contribute to compartmentalization signal.

Another limitation of this study was the use of cross-sectional data; which limits the study of the movement of viruses between blood and semen and change in viral compartmentalization over time. This was hampered by the fact that the study subjects were introduced to antiretroviral therapy as soon as the first sample was collected. Nonetheless, using Slatkin-Maddison test, bidirectional movement of the virus between compartments was evident in all the study subjects. In addition, there were more inferred migration events from blood to semen than from semen to blood (Wilcoxon $p = 0.0098$) which is consistent with a previous studies that had used Bayesian-based Markov Jump Counts [303, 530]. These two studies also had previously confirmed that Inferred migration events from Slatkin-Maddison test correlated with Markov Jump Counts.

Use of short reads (~400bp) is a limitation in interpretating phylogenetic data especially relating to clonaly amplification. The short read may have over estimated clonal amplification. It is difficult to know the contribution to viral diversity in the region outsided the 400bp that were sequenced.

Unlike previous studies that suggest that presence of sexually transmitted infections increases the level of virus in semen [145, 251, 446, 447, 582], there was no relationship between sexually transmitted infections and semen or blood viral loads.

In summary, this is the first study to probe HIV-1 compartmentalization between the systemic circulation and the male genital tract in chronically infected individuals utilizing deep sequencing. In contrast to previous studies, we show compartmentalization was present in all individuals studied. In addition, deep sequencing revealed clonal amplification in blood in a higher proportion of donors than had previously been reported. Both apparent discrepancies are likely due to higher sensitivity to detect these phenomena using deep sequencing. The extent of compartmentalisation was not obviously related to $CD4^+$ T cell count, viral load or history of sexually transmitted infections suggesting that it was independent of the overall state of immunodeficiency. This information furthers our understanding of how HIV-1 populations in the male genital tract are shaped prior to sexual transmission.

3 CHAPTER 3

Genotypic and phenotypic characterization of HIV-1 variants compartmentalized in blood and the male genital compartment

3.1 Abstract

Human Immunodeficiency Virus type 1 (HIV-1) transmission involves selection for particular genotypic and phenotypic characteristics whose understanding is critical for design of effective preventive interventions. Study of the characteristics of the virus establishing infection (Transmission /Founder virus) has produced mixed results. Majority of these studies had no information regarding the virus in the genital tract of the donors. Here we compared the variable loops characteristics and entry properties of HIV-1 envelope glycoproteins from blood and semen of three donors with compartmentalized viral populations in these two anatomical sites. The objective being to investigate whether there is change in function between HIV-1 populations in blood vs male genital tract.

Full-length env sequences were amplified at limiting dilution of cDNA (single genome amplification) in order to avoid PCR-based recombination artefacts. The resulting sequences were examined for loop length and glycosylation differences. In addition, sensitivity of blood- and semen-derived envelopes in constructed psedoviruses was tested against CCR5 inhibitors Maraviroc and PSC-RANTES, CXCR4 inhibitor JM2897 and fusion inhibitor enfirvirtide (T-20).

Differences in length and number of glycosylation sites were found between variants from blood and those from the genital tract: but the direction of the differences were inconsistent. Donor SVB043 (most severely compartmentalized) had semen variants that had longer and more glycosylated envelopes; while donor SVB008 (second most severely compartmentalized) had no difference in variable loop length, but semen variants were underglycosylated. The fourth variable loop (V4 loop) for donor SVB049 were longer and more glycosylated than those from donor SVB043 and SVB008. Differences in sensitivity to entry inhibitors between blood and semen variants were also observed and the direction of the differences varied as well. Semen variants for donor SVB043 were more insensitive to Maraviroc (CCR5 inhibitor) than the blood counter parts. The opposite was observed for donor SVB008 where the semen variants were more sensitive to inhibition by Maraviroc than the isolates from blood. There was no difference in sensitivity to maraviroc between blood and semen variants for donor SVB049. Surprisingly, sensitivity of donor SVB043 variants to PSC-RANTES (another CCR5 inhibitor) did not correspond with the observation for Maraviroc.

These observations support that there are distinct blood and genital tract HIV-1 phenotypes, and that the selective milieu of the two compartments shapes HIV-1 populations differently. Thus, some of the features of the transmitted/founder (T/F) virus may already be selected in the transmitting donor before they are transmitted to the recipients. The differences between variants from blood and genital tract is important in understanding HIV-1 transmission and in design of effective intervention strategies.

3.2 Background

The transmission of HIV-1 involves a genetic bottleneck irrespective of the mode of transmission [467] (reviewed in [508] and [468]). Some studies have suggested that the transmitted/founder (T/F) virus that establishes infection in the recipient is selected for leading to transmission of only one variant in 80% of heterosexual transmissions [456-458]. The biological determinants of the transmission bottleneck have not yet been fully determined. Most of the previous studies have characterized the nature of the T/F or acute vs chronic blood viruses [504, 506, 512, 583] as a source of information about the selective forces in play during transmission. Here we characterized the genetic and phenotypic features of the viruses compartmentalized in the male genital tract and compared these features to the virus found in the systemic circulation of the same donors. This information will dovetail the observations from previous studies that looked at the initially-infecting and/or acute viruses. Understanding these determinants will be important source of information for designing effective vaccine against HIV-1 and other interventions.

The virus found in semen is phylogenetically distinct from that in blood due to tissue-specific characteristics and difference in immunologic and other selective pressures [205, 396, 409]. In addition, seminal viruses exist as cell-free or cell-associated which can either be produced locally or shipped from blood [381] and have been shown to sub-compartmentalize [380]. The unique selective milieu of the male genital tract shapes the virus differently than what happens in blood. Indeed, compartmentalized viral variants in the male genital tract have been observed to have a strong genetic signature at C2-V3 region with the most prominent signature at position 464 [409]. This therefore means that the features of the

transmitted/founder virus may begin much earlier in the donor during the transition to the recipient. Understanding these features of the potentially transmitted virus in the donor is important since this is the viral population that must be antagonized by a successful intervention strategy in the recipients.

Previously, semen-derived variants were observed to have significant differences in the number of potential N-linked glycosylation sites (PNGS) at the C2-V3 region in subtype B-infected subjects [409] compared to blood. The difference was not in any particular direction with 3/7 of the pairs having more PNGS in semen than plasma and 3/7 having the opposite, 1/7 had no difference. This difference did not correspond well with observations in another study involving subtype B-infected individuals, where newly transmitted viruses were found to be underglycosylated in the V1-V4 region for only 1/6 of the transmitting pairs [525]. This however supports that there is likely to be local viral production, and that the HIV-1 populations are shaped by the prevailing environment within its vicinity. In the context of subtype C infections, underglycosylation has been reported as a feature of T/F virus both in virus initiating infection (T/F) [459] or acute vs chronic viruses [506]. However, similar information regarding potential glycosylations or length of variable loops is not available for the HIV-1 populations found in the genital tract. In yet another study, the PNGS at position 332 of the HIV-1 envelope was underrepresented in sequences of recently transmitted subtype C variants compared to variants existing one year post infection in the same donors [584]. This was not observed in subtype B viruses [509] suggesting a possibility that the selective conditions in the genital tract may impact the virus in a subtype-specific manner. This leaves it unresolved whether the selection for changes observed in PNGS happens within the donor's genital tract or during transmission into a new recipient.

Several studies have shown that HIV-1 entry inhibitors are efficient in preventing acquisitions of HIV-1 when used as microbicides or administered orally [585-589]. These compounds antagonize the virus entry in the initial stages of the lifecycle; i.e. prevent viral gp120 from interacting with the CD4 receptors or CCR5/CXCR4 coreceptors or prevent fusion of the virus with the cell membrane of host cells. When these compounds are used as topical microbicides, the virus they are required to antagonize during sexual transmission is the one that comes from genital tract of the infected donor. Additionally, the activity level of these

inhibitors against genital tract and blood HIV-1 may allow us to draw conclusions about the changes in function of the HIV-1 Env and systematic changes in viral properties.

This chapter of the thesis therefore characterizes and compares the utilization of entry inhibitors between the variants from blood versus those from the male genital tract using a pseudovirion-based TZM-bl assay for three of the most compartmentalized donors. Since the early infecting virus are R5-tropic, semen derived viruses would be expected to be more sensitive to CCR5 inhibitors compared to isolates from blood. In addition, genotypic features relating to variable loop characteristics were assessed from the single genome amplification of env sequences of the same participants.

3.3 Methods

3.3.1 Study participants

Blood and semen samples were collected from 44 HIV-infected men attending a health facility at ANOVA Health's Health4Men clinic at Woodstock or Green Point, Cape Town, South Africa as previously described, or who were invited to the joint study at ANOVA Health's mobile clinic in Khayelitsha, South Africa (Chapter 2). All samples were collected at one of the ANOVA Health facilities when the clinic was closed for normal business. Of eleven subjects who were analysed for level of compartmentalization by deep sequencing in chapter 2, three subjects who had the highest extent of compartmentalization were selected for this part of the experiment i.e. SVB043, SVB008 and SVB049 (Table 10 and chapter 2 figure 3).

3.3.2 RNA extraction and cDNA synthesis

From each blood or semen sample, RNA was extracted from the equivalent of approximately 10,000 copies based on the viral load of the sample, except for semen samples with low viral loads, in which case the entire sample was used. RNA was extracted using QIAamp Viral RNA Mini Kit (Qiagen). The extracted RNA was eluted in a final volume of 50µl. Extracted RNA was reverse transcribed into cDNA using the SuperScript First-Strand Synthesis System Kit

(Invitrogen) and following the recommendations from the manufacturer. Briefly, the reaction contained 1 x Reverse transcriptase (RT) buffer, 5mM DTT, 2U/μl RNaseout which is an RNase inhibitor, 0.5mM each of the deoxynucleotide triphosphate (dNTP), 0.25μM Oligo$(dT)_{20}$ and 10 U/μl of the SuperScript III RT enzyme. Reverse transcription was performed at two temperatures, first at 45°C for an hour followed by another hour at 50°C. The enzyme was then deactivated by heating at 70°C for 15 minutes before holding at 4°C. The RNA was digested from RNA-DNA hybrids using RNAaseH by incubating at 37°C for 20 minutes. the cDNA was either used immediately or stored at -20°C for later use.

3.3.3 Single genome amplification (SGA) of the HIV-1 envelope gene

The single genome amplification (SGA) technique was used to amplify single genomes from cDNA generated from blood and semen. This was done by serial diluting the cDNA templates and distributing the dilutions in a 96-well plate followed by first and second stages of the nested PCR. The dilution of cDNA that resulted in less than 30% positive wells from the total reactions after gel electrophoresis was targeted. Assuming a poison distribution of cDNA templates and no binding between templates, when 30% of wells are positive, the likelihood that the PCR product in any one positive well was amplified from only one amplified template is approximately 84% [456-458]. Use of a single template was confirmed to the extent possible by sequencing of the amplicons and checking for mixed bases from the chromatograms. Clones were obtained from several independent PCRs for each sample to minimize the sampling bias. The PCR amplifications were performed using the Platinum HIFI Taq (Invitrogen) to keep the likelihood of PCR mismatch errors as low as was practical. The PCR conditions in a 20μl reaction contained 1x HIFI Buffer, 2mM $MgSO_4$, 0.2mM each of dNTPs, 0.2μM of each primer and 0.025U/μl of Platinum HIFI Taq. The first-round primers were, forward VIF-1 5'– GGGTTTATTACAGGGACAGCAGAG-3' and reverse OFM19 5'– GCACTCAAGGCAAGCTTTATTGAGGCTTA-3' representing HXB2 positions 4903-4923 and 9604-9632 respectively. The second-round primers were, forward ENV A 5'– GGCTTAGGCATCTCCTATGGCAGGAAGAA-3' and reverse ENV N 5'– CTGCCAATCAGGGAAGTAGCCTTGTGT-3' representing HXB2 positions 5954-5982 and 9145-9171 respectively. This generated an amplicon of approximately 3kb. The PCR cycling conditions for first round PCR involved initial denaturation at 94°C for 2 mins, followed by 35

cycles at 94°C for 15 s denaturation, 55°C for 30 s primer annealing step and extension at 68°C for 4 mins. This was followed by a final extension at 68°C for 10 min. Two microliters of first round PCR were then added to second round PCR reaction and the same cycling conditions were used only that a total of 45 cycles were run this time. All the PCR reagent preparations and making of the master mixes were done in the PCR clean room. In addition, one sample was processed at a time to prevent cross-sample contaminations.

3.3.4 Cloning and generation of pseudovirus plasmids

PCR was first performed from 2μl of first stage PCR products in a reaction volume of 50μl. The primers used were, forward ENVARx 5'-CACCGGCTTAGGCATCTCCTATAGCAGGAAGAA-3' and reverse ENV N 5'– CTGCCAATCAGGGAAGTAGCCTTGTGT-3' representing HXB2 positions 5954-5982 and 9145-9171 respectively. The special primer ENVARx improves expression of HIV-1 subtype C pseudotyped clones in expression cell lines by introducing a stop codon between the rev-vpu and env open reading frames [590]. In addition, the 'CACC' tag [457] allows for the directional cloning of the amplicon relative to CMV promoter into pcDNA3.1D/V5.His.TOPO vector which comes with a CTGG overhang (Figure 25). The cycling conditions were like the ones applied for first stage second-round PCR as detailed previously. The PCR products were then purified using QIAquick PCR purification kit (Qiagen), cloned into pcDNA3.1D/V5.His.TOPO vector (Invitrogen), and transformed into TOP 10 competent bacteria (Invitrogen). Bacterial clones containing inserts were identified by finding a PCR product using T7 and BGH primers that were supplied together with the kit (Figure 25). A band of around 3kb was indicative of a successful cloning. The positive colonies were then grown in Luria Broth (LB) supplemented with 100mg/ml of ampicillin overnight at 32°C in a shaking incubator at 200 revolutions per minute (rpm). The plasmids were then harvested from bacterial cultures using QIAprep Spin Miniprep Kit (Qiagen). After quantification of the DNA using the NanoDrop 2000 spectrophotometer (Thermo Scientific), the plasmids were used fresh or stored for later use at -20°C.

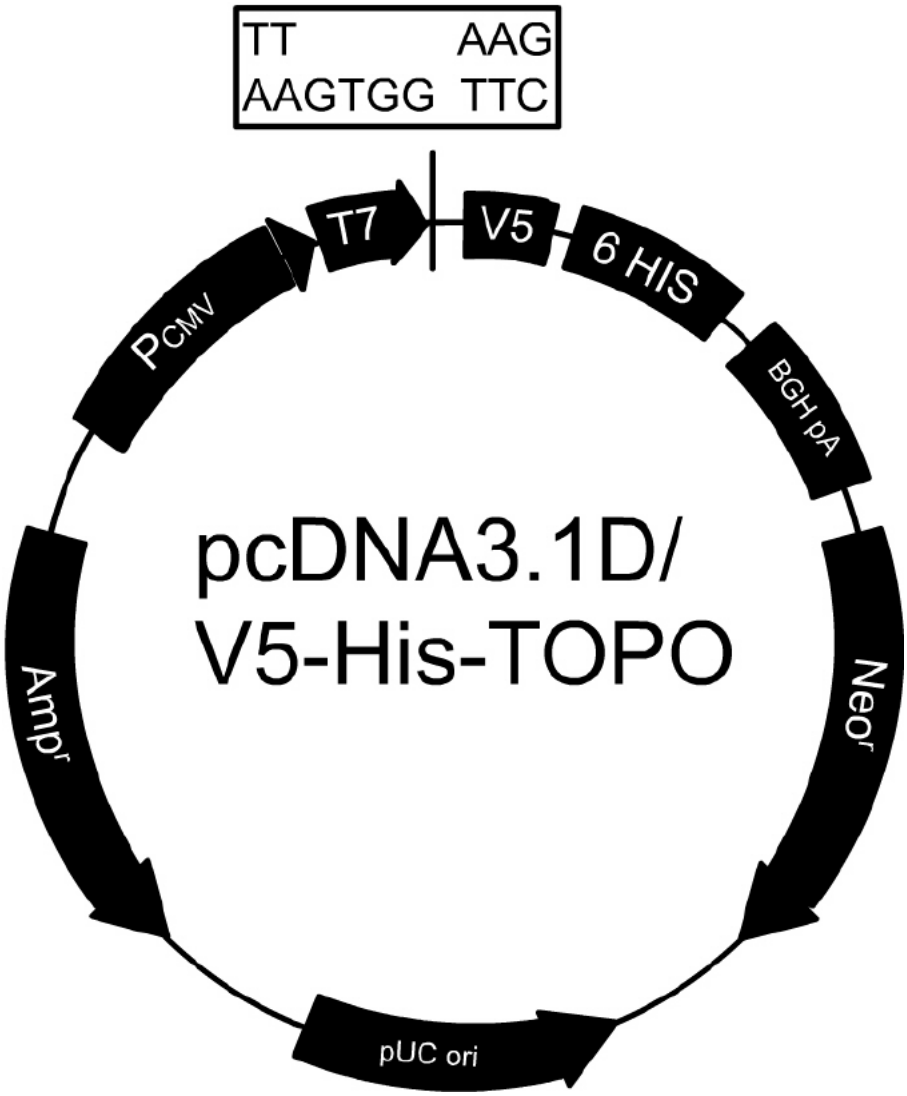


Figure 25: Schematic representation showing the map of pcDNA3.1D/V5-His-TOPO cloning vector. Figure obtained from Tarr et all [591]

Pseudoviruses were constructed using established methods [592]. Briefly, the env-cassette plasmids and a HIV-1 subtype B backbone deficient of the envelope (pSG3Δenv) were mixed in a mass ratio of 1:2, and co-transfected into HEK293T (packaging cell line) (Figure 26) cells using X-tremeGENE (Roche Diagnostics). The plasmids carrying the envelopes of interest (blood and semen strains envelopes) provided the envelope spikes in trans for the new psedoviruses while the rest of the HIV-1 proteins were provided by the pSG3Δenv genome (Figure 26). The later lacked a functional envelope gene due to a four base pair insertion in the *env* open reading frame. After 48 hours incubation at 37°C, 5% CO_2, the pseudoviruses were harvested from the supernatants, filtered through 0.45μM syringe filters to remove aggregates, and stored in single-use aliquots. The titre of the pseudoviruses was determined

in a TZM-bl assay. Titration of the pseudoviruses that resulted in approximately 50,000 relative light units (RLU) was used for neutralization assays, except that if this was not achieved using undiluted supernatant, then the undiluted supernatant was used.

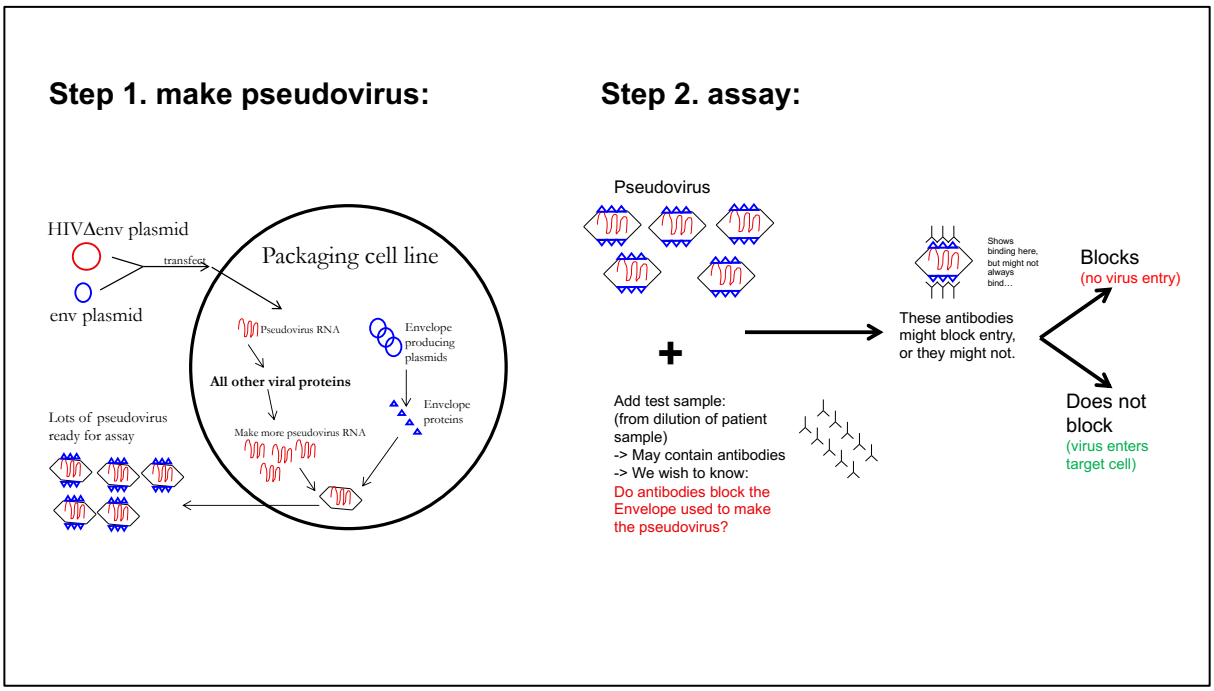


Figure 26: Schematic representation of the pseudovirus construction process and the neutralization assay. The plasmid carrying the envelope of interest and the envelope-deficient plasmid (HIVΔenv) were co-transfected into the packaging cell. The HIVΔenv provides the accessory genes while the envelope spikes of the pseudoviruses comes from the env plasmid. During the neutralization assay, the antibodies may or may not bind to the envelope spike. If they bind, the pseudovirus is blocked from entering the detector cell line and if not, entry to detector cell line happens. (The diagram used with permission from Jeffrey R. Dorfman).

3.3.5 Cell lines

TZM-bl cells were used as the target cells for HIV-1 pseudoviruses. These cells are modified from a clone of a Hela cell that expresses high levels of DC4, CCR5 and CXCR4 and support single cycle virus entry [592]. In addition, they are engineered with a luciferase reporter gene whose transcription is driven by an HIV-1 LTR. HIV-1 Tat, when produced by an invading HIV-1, activates transcription of luciferase, which is therefore indicative of infection of the cell. These cells were obtained through the National Institute of Health (NIH) AIDS Reagent Program from Dr John C. Kappes and Dr. Xiaoyun Wu. Transfection experiments to make the pseudoviruses were done using HEK 293T cells. Both types of cells were maintained at 1.0 x 10^6 cells in DMEM supplemented with 5% heat-inactivated Fetal Bovine Serum (FBS), 25mM

HEPES (Lonza) and 1% non-essential amino amino acids (Sigma-Aldrich). The cultures were split every 3-4 days after they obtained more >80% confluence. Medium used for the assay contained DMEM, 10% FBS, 25mM HEPES, 1% non-essential amino acids, 100μg/ml streptomycin, 100U/ml penicillin and 100μg/ml gentamicin.

3.3.6 Entry inhibitors

Sensitivity to entry inhibitors was done in a TZM-bl based assay in the absence or presence of varying concentrations of the entry inhibitors, Maraviroc, PSC-RANTES, JM2987 and enfuvirtide (T20). Maraviroc is an allosteric CCR5 antagonist and is already in use as a therapeutic drug against HIV-1 [585]. PSC-RANTES is a chemokine analogue which is a competitive inhibitor of CCR5 [589]. Enfuvirtide (T-20) is a fusion inhibitor made of a synthetic peptide mimicking the carboxyterminal heptad repeat sequence (HR2) of gp41 preventing the interaction with the aminoterminal heptad repeat sequence (HR1). This prevents the formation of the hairpin structure necessary for the fusion of the virus with the cell membrane of the target cell [593]. The JM2987 is a CXCR4 inhibitor. Each entry inhibitor was first serial diluted 5-fold in DMEM containing 10% FBS. TZM-bl cells were then added and a 1-hour pre-incubation done. This was followed by addition of the virus. Each entry inhibitor was tested in duplicate and luciferase activity was measured after 48 hours of culture. The entry inhibitor antiviral activity was determined as a percentage of viral inhibition in the presence of the inhibitor compared to the controls with no inhibitor. Plotting of the concentrations and the percentage inhibition in GraphPad Prism (GraphPad Prism software, La Jolla, USA) allowed the determination of EC_{50}. Tests were repeated at least 2 times, and the final EC_{50} was calculated by taking the geometric mean of the EC_{50} values for each test. The geometric mean EC_{50} of values of blood-derived isolates were then compared to those from semen-derived isolates.

3.3.7 Sequencing of the SGA clones

The full-length envelope gene was sequenced using a standard set of 12 primers (Table 9) in such a way that the contigs partially overlapped to provide one contiguous sequence after assembly. Sequencing was performed by the Stellenbosch University Central Analytical

Facility. The sequenced fragments were processed and assembled and edited using Sequencher DNA analysis software version 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). Sequences of all SGA clones will be submitted to GenBank. The timing of release of the sequences will be in accord with the requirements of the journal in which this work is accepted for publication.

Table 9: Primers used for Single Genome Amplification (SGA) PCRs, cloning and sequencing of the PCR amplicons

NAME	SEQUENCE	HXB2 POSITION	FORWARD/ REVERSE
	Primers used for PCR amplification		
VIF1(VIF)	5'– GGGTTTATTACAGGGACAGCAGAG-3'	4903-4923	Forward
OFM19	5'– GCACTCAAGGCAAGCTTTATTGAGGCTTA-3'	9604-9632	Reverse
ENV A	5'– GGCTTAGGCATCTCCTATGGCAGGAAGAA-3'	5954-5982	Forward
ENV N	5'– CTGCCAATCAGGGAAGTAGCCTTGTGT-3'	9145-9171	Reverse
	Primers used for sequencing of SGAs		
EF00	5'- GGGAAAGAGCAGAAGACAGTGGCAATGA-3'	6204-6228	Forward
REV15	5'-CTGCCATTTAACAGCAGTTGAGTTGA-3'	7015-6990	Reverse
FOR14	5'-TATGGGACCAAAGCCTAAAGCCATGTG-3'	6556-6582	Forward
E175	5'-TTTAGCATCTGATGCACAGAATAG-3'	6378-6398	Reverse
EF115	5'-AGAAAAATTCTCCTCTACAATTAA-3'	7351-7371	Reverse
EF55	5'-GCC CCA GAC CGT GAG TTG CAA CAT ATG-3'	7914-7937	Reverse
EF15	5'-CTTGCTCTCCACCTTCTTCTTC-3'	8424-8442	Reverse
FOR16	5'-TTTAATTGTGGAGGAGAATTTTTCTA-3'	7350-7375	Forward
REV19	5'-ACTTTTTGACCACTTGCCACCCAT-3'	8820-8797	Reverse
EF170	5'-AGC AGG AAG CAC TAT GGG-3'	7802-7816	Reverse
EF200	5'-GGG ATA ACA TGA CCT GGA TGC AGT GGG-3'	8095-8118	Forward
EF260	5'-TTCAGCTACCACCGATTGAGAGACT-3'	8523-8544	Forward
	Primers used for cloning		
ENVARX	5'-CACCGGCTTAGGCATCTCCTATAGCAGGAAGAA-3'	5954-5982	Forward
ENV N	5'– CTGCCAATCAGGGAAGTAGCCTTGTGT-3'	9145-9171	Reverse

3.3.8 Processing of the sequence data and phylogenetic analysis

The sequenced contiqs were assembled into full HIV-1 envelope sequences using Sequencher Sequence Analysis software version 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). Alignments were done using MAFFT [594] and edited in Aliview [553]. Maximum likelihood phylogenetic trees were constructed under GTR model using FastTree [554] and visualized in FigTree [560].

3.3.9 Analysis of the length of HIV-1 variable loops

The variable loop lengths were determined using an online tool found in Los Alamos HIV sequence database site (https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/index.html). The online tool takes the alignment as input.

3.3.10 Analysis of the number of potential N-glycosylation sites (PNGS)

The predicted N-glycosylation sites were counted using N-glycosite tool available as a web service at Los Alamos HIV sequence database site (https://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html). The site also takes an alignment as the input data and uses the HXB2 sequence to number the positions.

3.3.11 Statistical Analysis

The statistical analysis was done in STATA (STATACorp) and GraphPad Prism. The graphs were generated in GraphPad Prism. Significance testing were done using non-parametric methods with significant result cut off p-value set at 0.05.

3.4 Results

3.4.1 Study participants

This part of the thesis was done using three study participants who had the highest extent of compartmentalization as detailed in the previous chapter. The characteristics of these three study participants is available in Table 10 below.

3.4.2 Compartmentalization analysis

As indicated previously compartmentalization was assessed from deep sequencing data using three standard statistical tests as described in chapter 2. Using participants with the highest extent of compartmentalization enabled us to see the phenotypic differences more clearly.

Table 10: Study participants with CD4 counts, viral loads in blood and semen, number of Single genomes amplifications (SGAs) from each compartment and status of compartmentalization. From Chapter 2

SampleID	CD4 count (cells/μl)	Blood viral load (cp/ml)	Semen viral load (cp/sample)[a]	Number of SGAs from Blood	Number of SGAs from Semen	F_{ST}[b] pvalue	S_{nn}[c] pvalue	SMT[d] pvalue
SVB043	509	122297	5556598	12	10	<0.001	<0.001	<0.001
SVB008	257	309388	38640	13	10	<0.001	<0.001	<0.001
SVB049	423	85711	7915	8	10	<0.001	<0.001	<0.001
SVB021	148	183339	4435	0	0	<0.001	<0.001	<0.001
SVB026	209	46202	4843	0	0	<0.001	<0.001	<0.001
SVB012	586	145745	6365	0	0	0.012	<0.001	<0.001
SVB025	368	440969	23356	0	0	<0.001	<0.001	<0.001
SVB030	390	25781	235710	0	0	<0.001	<0.001	<0.001
SVB029	149	20947	273305	0	0	<0.001	<0.001	<0.001
SVB039	425	43191	132605	0	0	<0.001	<0.001	<0.001
SVB041	567	89090	27540	0	0	<0.001	<0.001	<0.001

[a]Viral load assumes 100% yield upon filtration and ultracentrifugation
[b]F_{ST} = Wright's F-statistic
[c]S_{nn} = Nearest Neighbour statistic
[d]SMT = Slatkin-Maddison Test

3.4.3 Genotypic characteristics

Each study participant's blood and semen sample full-length envelopes were amplified using the SGA technique [457, 458]. A median of 10 (range 8 - 13) sequences were generated per compartment from the three donors who showed the highest extent of compartmentalization; SVB043, SVB008 and SVB049 (Table 10). Genotypic comparisons for variants from blood and semen were done regarding the variable loop lengths (V1V2, V3, V4, V5) and predicted N-glycosylation sites (PNGS). In the individual phylogenetic trees, full-envelope each donors' sequences clustered together with high bootstraps supports (above 70) suggesting that the individuals were infected by one variant or a number of related variants (Figure 27). In addition, it shows that there was no cross-sample contamination between the three donors.

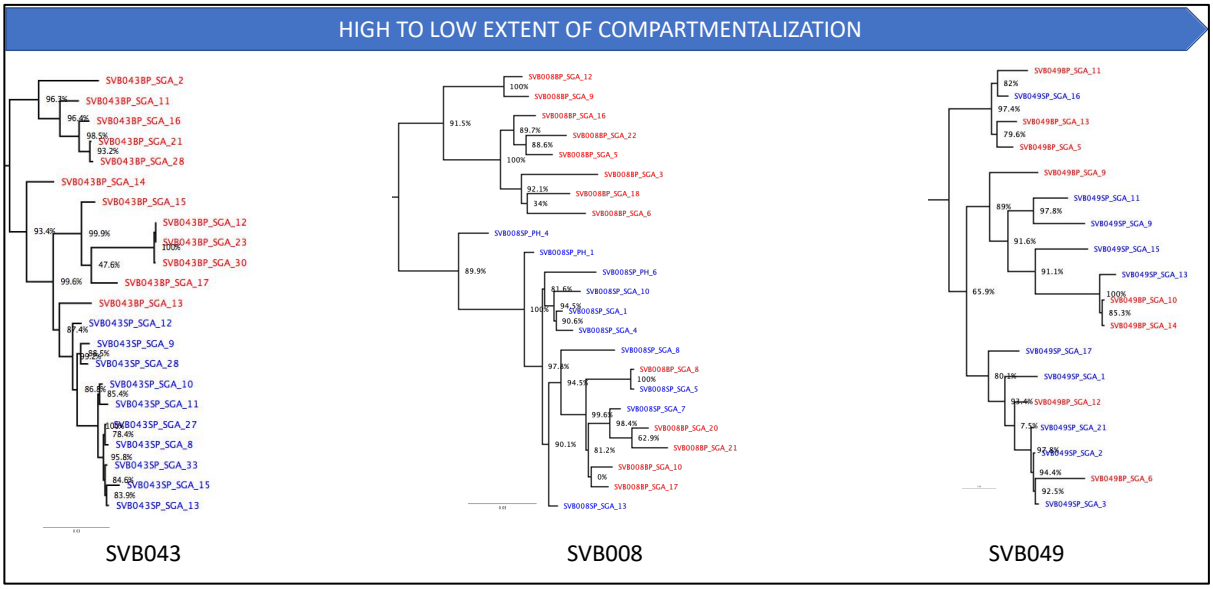


Figure 27: Maximum likelihood trees for donors SVB043, SVB008, SVB049 arranged by the extend of compartmentalization. Red color represent sequences from blood and blue color represent sequences from semen. Donor SVB043 had the highest extend of compartmentalization of viral variants between blood and semen while donor SVB049 had the least.

3.4.3.1 Variation of lengths of variable loops between plasma- and semen-derived populations

It has previously been reported that among the HIV-1 subtype C viruses, the transmitted founder viruses have short V1-V4 regions [459]. Here the degree of variable loop

polymorphisms between blood and genital tract sequences were evaluated from subjects infected with HIV-1 subtype C. The variable loop lengths between blood and semen sequences within each donor were compared using Mann-Whitney test while inter-group comparisons were done using Dunn's multiple comparison test.

Significance differences in length of V1V2 region were observed between blood and semen variants of donor SVB043 who had the most compartmentalization viral populations between blood and semen (Mann-Whitney $P < 0.001$) (Figure 28 A). This however was not the case for V1V2 region for donors SVB008 and SVB049 (Figure 28 A). Unexpectedly, donor SVB043 difference was in the opposite direction i.e. semen variants had longer V1V2 variable loops than their blood counterparts (Figure 28). There were no significant differences in length of V3, V4 and V5 between blood and semen clones for all the three donors (Figure 28 B-D). The V3 region lengths among all the three donors and compartments remained constant at 37 amino acid with one clone from SVB049 blood variants having 38 amino acids (Figure 28 B).

Between donor variations also occurred in all the variable regions except the V3 region (Figure 28 E-G). Compared with other donors, SVB049 had longer V4 loop lengths (Figure 28 F). In addition, donor SVB043 semen V5 loops were longer than those from blood and semen of other donors (Figure 28 G).

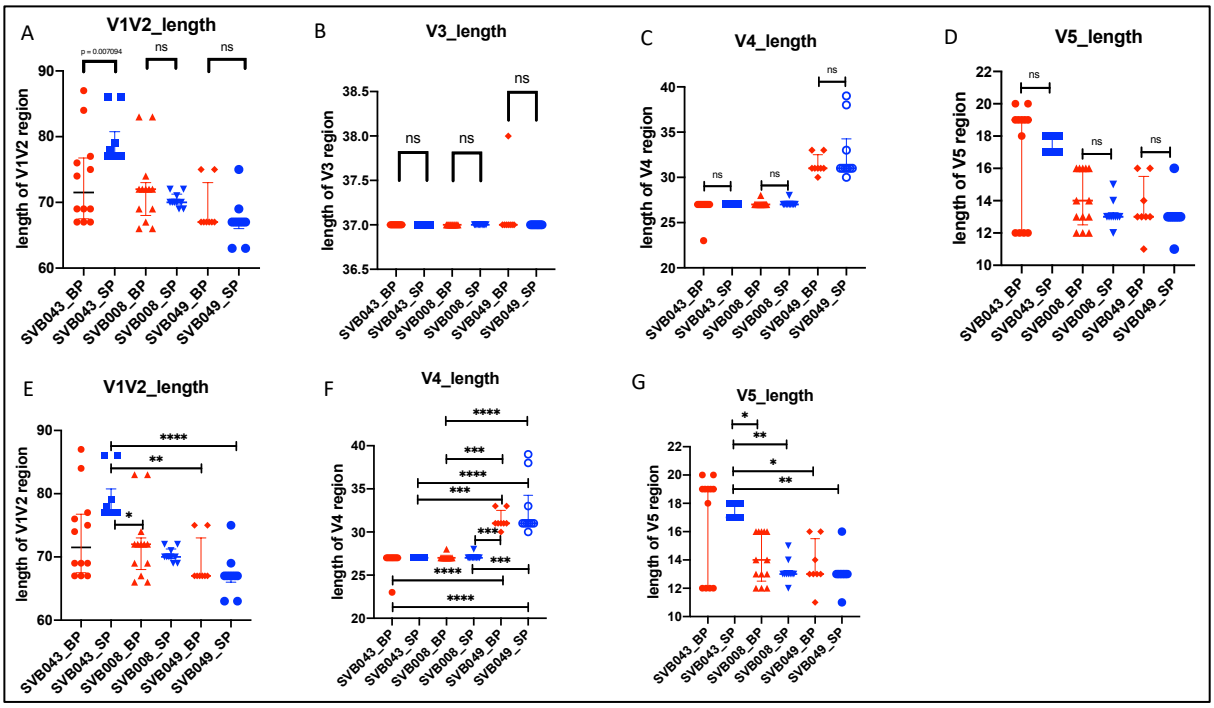


Figure 28: Variable loops polymorphisms. SVB043 had semen clones with overall longer V1V2 variable loops than those from blood which was not observed in the other two donors (A). There were no significant differences in loop polymorphisms in other variable loops between blood and genital compartment (B-D) although between donor variations occurred except in V3 region (E-G).

3.4.3.2 Predicted potential N-glycosylation sites (PNGS) in plasma- and semen-derived viral populations

Previous study had observed that the transmitted/founder (T/F) virus was underglycosylated in subtype C [459] and not in subtype B [525] HIV-1. In this section of the thesis, this property was investigated for donor semen variants and compared to those in blood. Indeed, the number of PNGS varied between the two compartments among the different donors but the direction of the difference was inconsistent (Figure 29). For the most severely compartmentalized donor, SVB043, semen variants had significantly higher number of PNGS compared to the blood counterparts (Figure 29 A). In contrast, for the second most compartmentalized donor, SVB008, sequences from the genital tract had lower number of PNGS compared to the blood sequences (Figure 30 B). There was no differences in number of PNGS between semen and blood variants for donor SVB049 (Figure 29 C).

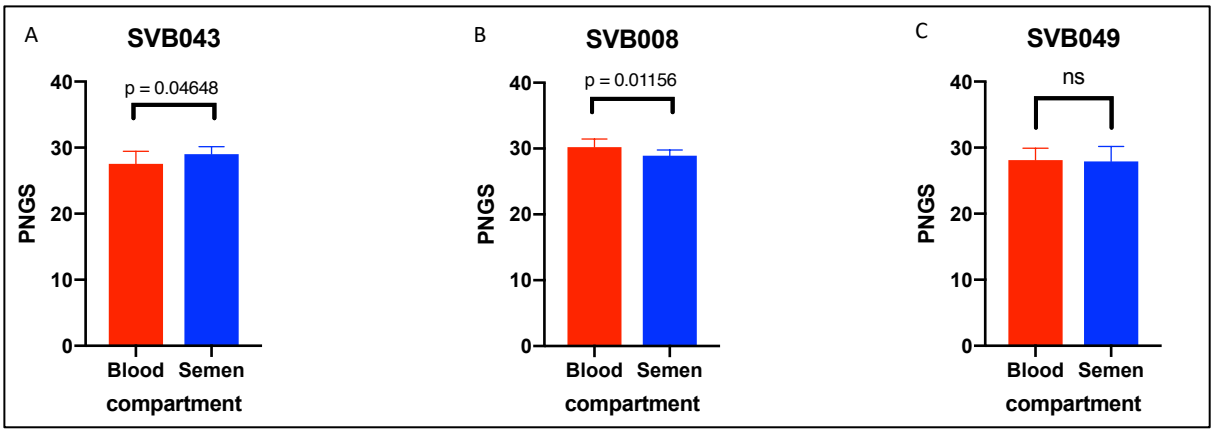


Figure 29: Predicted N-glycosylation sites (PNGS) in the variable loops. The semen variants for donor SVB043 had more PNGS than the variants from blood. In contrast donor SVB008 who had the most compartmentalized viral populations between blood and semen after donor SVB043 had less PNGS in semen than in blood. There was no significant difference in the number of PNGS between blood and semen variants for donor SVB049

Considering number of PNGS in individual variable loops, variations were also observed. Generally, there was no difference in the number of PNGS between blood and semen in all the variable loops except V3 of donor SVB008 (Figure 30). Similarly between donor variations of the variable loops were observed and mainly occurred in V3, V4 and V5 and rarely in V1V2 (Figure 30 E-H).

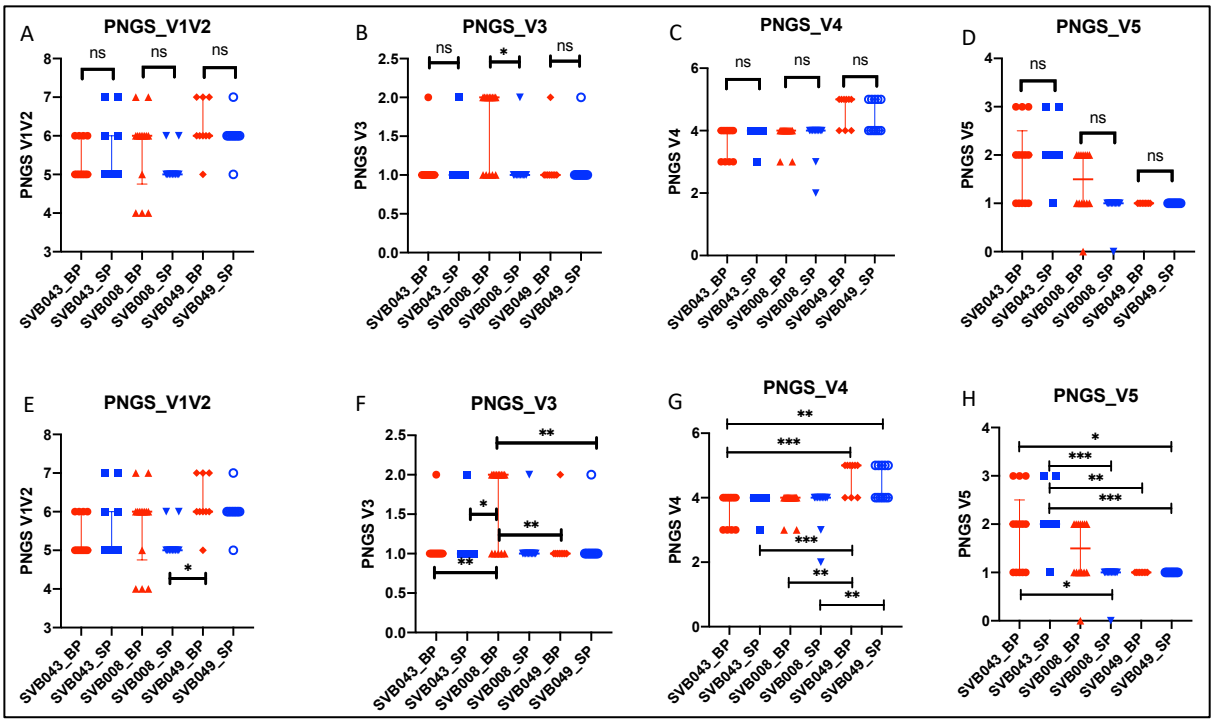


Figure 30: Within (A-D) and between (E-H) Predicted N-glycosylation sites (PNGS) in the variable loops. There were no within donor variations in number of PNGS between blood and semen in all the variable regions except V3 of donor SVB008 (A-D). Between donor variations majorly occurred in V3, V4 and V5 and rarely in V1V2 (E-H)

3.4.4 Blood and semen viral compartmentalization and susceptibility to entry inhibitors

Entry of HIV-1 into the target cells occurs during the initial steps of the lifecycle. Entry inhibition assays allows understanding of the differences in change of viral envelope function for variants in blood versus those in the genital tract. The coreceptor tropism was studied through sensitivity to maraviroc (CCR5 antagonist), PSC-RANTES (CCR5 competitive inhibitor), JM2987 (CXCR4 inhibitor) and enfuvirtide (T-20). Sensitivity to maraviroc varied between blood and semen variants among the three donors. Donor SVB043 who had the most compartmentalized viral populations between blood and semen had semen variants that were more insensitive to maraviroc (Figure 31 A). This was the opposite of what was found for donor SVB008 where the semen variants were sensitive to maraviroc (Figure 31 B). There was however a wide range of sensitivities of blood clones to maraviroc for this donor (Figure 31 B). In contrast, there was a narrow range of sensitivities of semen clones to maraviroc for this particular donor (Figure 31 B). On the other hand, there was no difference in sensitivities to maraviroc between variants derived from blood and those from semen for donor SVB049 (Figure 31 C).

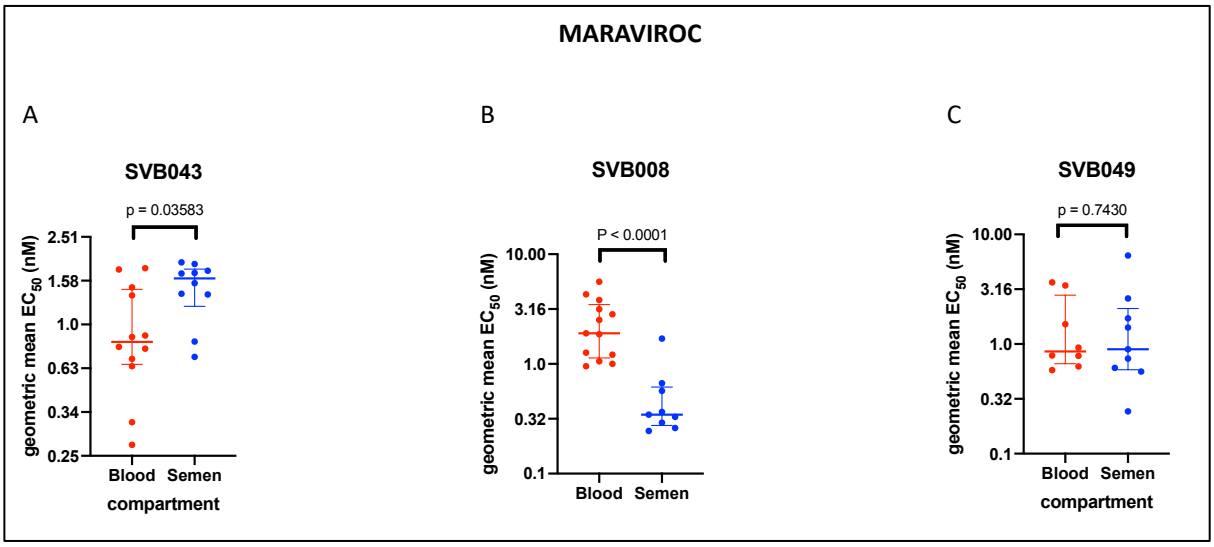


Figure 31: Sensitivities of blood and semen variants to maraviroc (CCR5 allosteric inhibitor). Donor SVB043 had semen variants that were more resistant to maraviroc inhibition than the blood counterparts (A). In contrast, donor SVB008 had semen variants that were more sensitive to maraviroc inhibition (B). There was no difference in sensitivities to Maraviroc between variants from blood versus those from semen for donor SVB049 (C).

Sensitivities of blood and semen variants to another form of CCR5 inhibitor, PSC-RANTES varied among the three donors (Figure 32). There were no significant differences in sensitivity of blood and semen variants to PSC-RANTES for donors SVB043 and SVB049 (Figure 32 A & C). In contrast, semen variants for donor SVB008 were more sensitive to PSC-RANTES than the variants from semen (Figure 32 B).

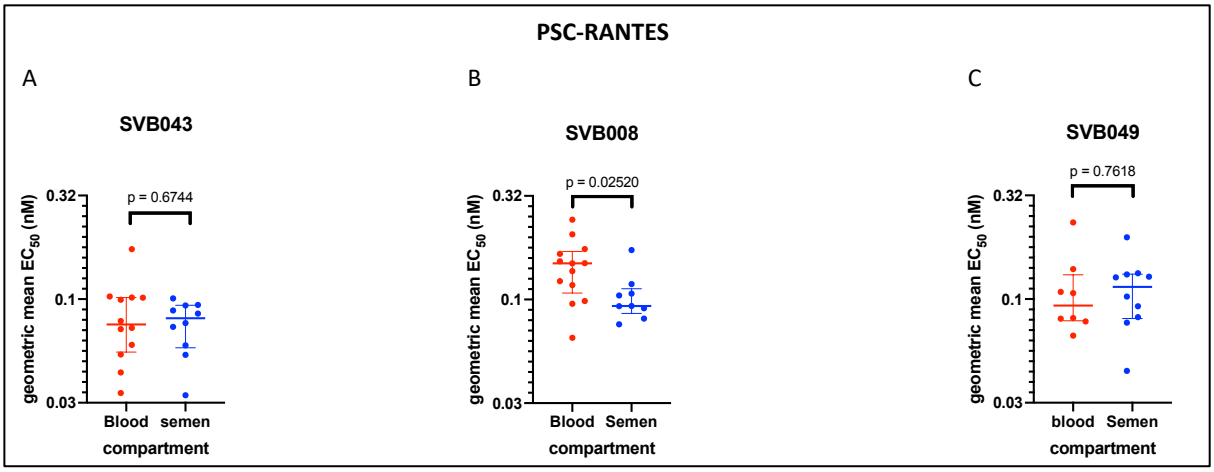


Figure 32: Sensitivities of blood and semen variants to PSC-RANTES (CCR5 competitive inhibitor). Only donor SVB008 had differences in sensitivity to PSC-RANTES where semen variants were more sensitive to PSC-RANTES than variants from blood.

Of note, donor SVB043 blood variants had a strong correlation in potency between maraviroc and PSC-RANTES despite differences in their mode of action $r^2 = 0.5206$; $p = 0.008079$) (Figure 33 A). This however was not the case with the semen variants of the same donor ($r^2 = 0.07265$; $p = 0.4514$) (Figure 33 B). Donor SVB008 did not show a correlation in potential between Maraviroc and PSC_RANTES for both blood and semen variants (Figure 33 C & D). On the other hand, donor SVB049 had strong correlation between potencies of PSC-RANTES and Maraviroc for both blood $r^2 = 0.5468$; $p = 0.03603$) and semen ($r^2 = 0.8072$; $p = 0.0009940$) (Figure 33 E & F).

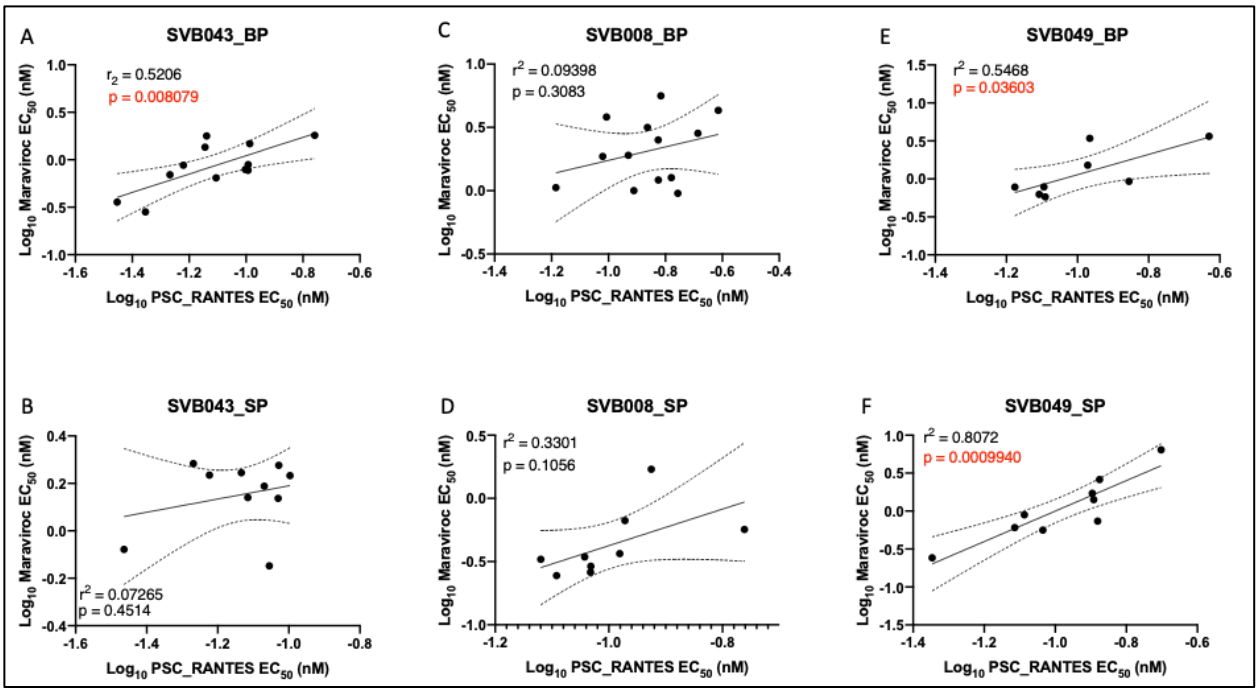


Figure 33: Correlation analysis of sensitivities to Maraviroc and PSC-RANTES. Solid line is the regression line while the dotted line is the 95% confidence intervals.

All the three donors' variants from blood and semen were resistant to JM2987, a CXCR4 inhibitor suggesting that they were all R5-tropic virus (Figure 34 A – C).

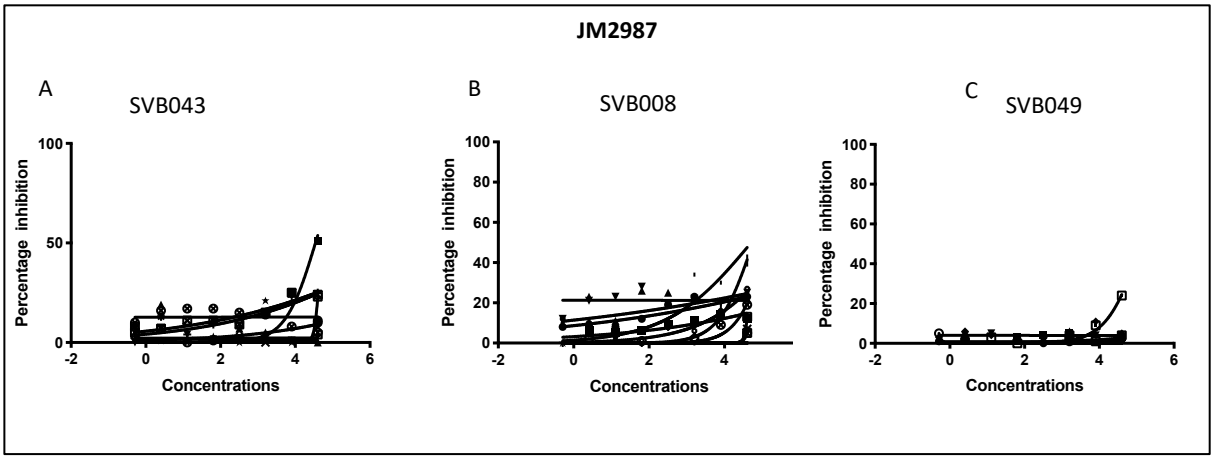


Figure 34: Sensitivities of compartmentalized blood and semen variants to JM2987. All the clones from blood and semen in all the three donors were insensitive to CXCR4 inhibitor.

There were no differences in sensitivities of blood and semen variants to enfirvitide (T-20), a fusion inhibitor (Figure 35). However, semen variants for donor SVB049 had wide ranges of

sensitivity to T-20 compared to blood variants (semen geometric mean EC_{50} 3.9 - 28.4 vs 4.0 – 17.4 for blood) (Figure 35 C).

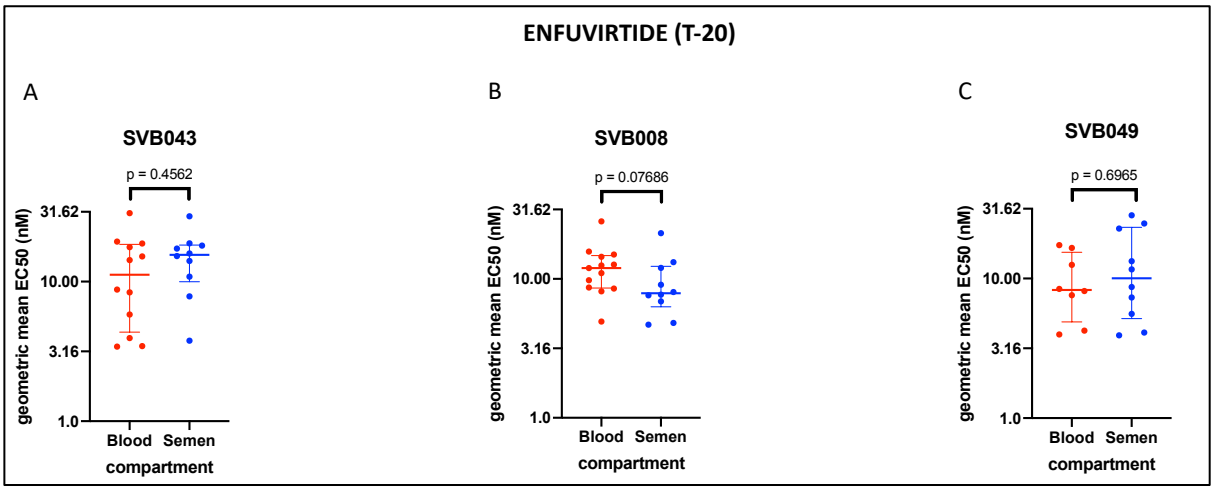


Figure 35: Sensitivities of compartmentalized blood and semen variants to T-20 (fusion inhibitor). There were no differences in sensitivities to T-20 between blood and semen for all the three donors who had the highest extent of compartmentalization.

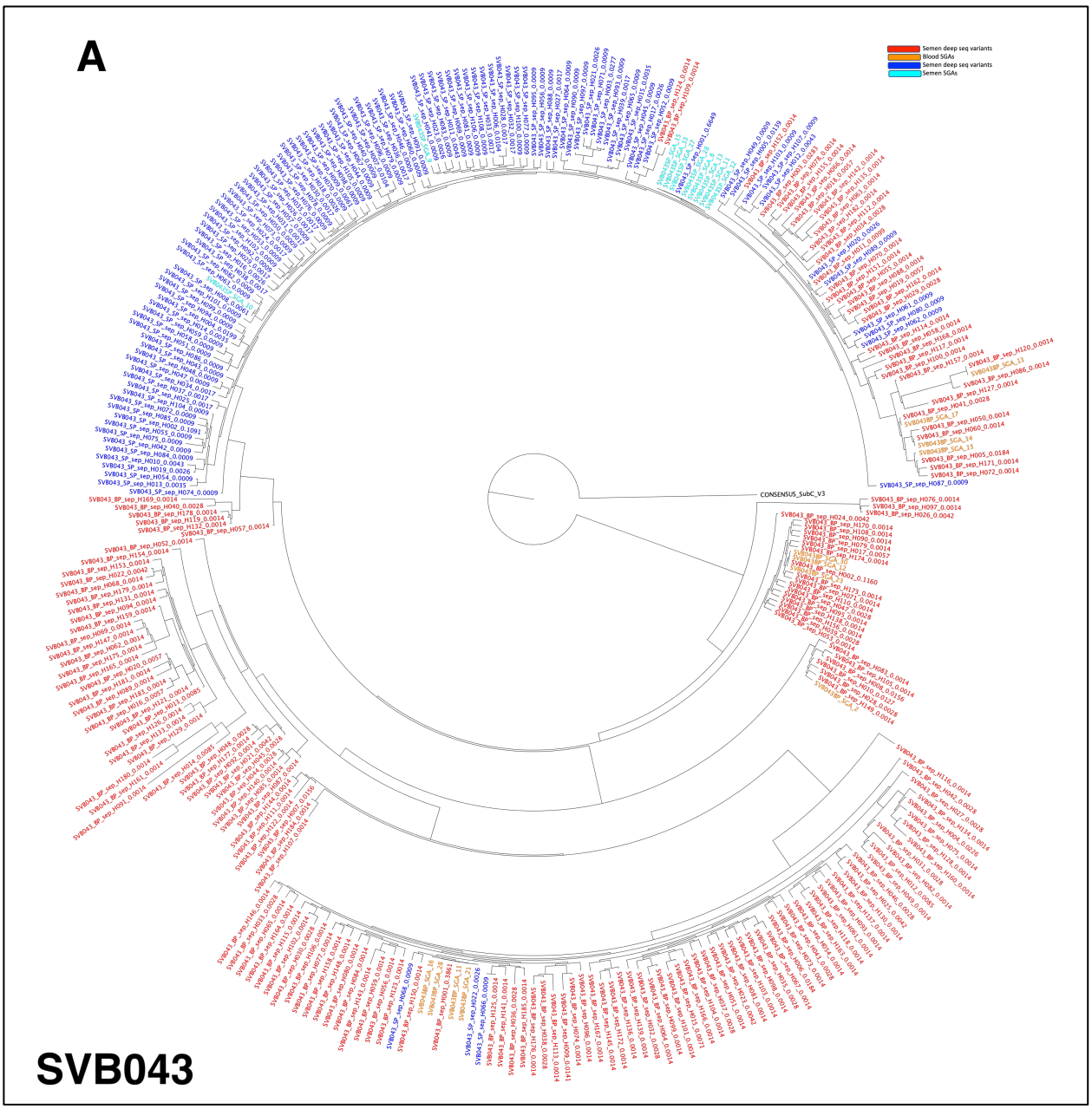
In sum, all the viruses tested in this part of the experiment were R5-tropic and the sensitivities to the entry inhibitors varied in each donor and in each compartment.

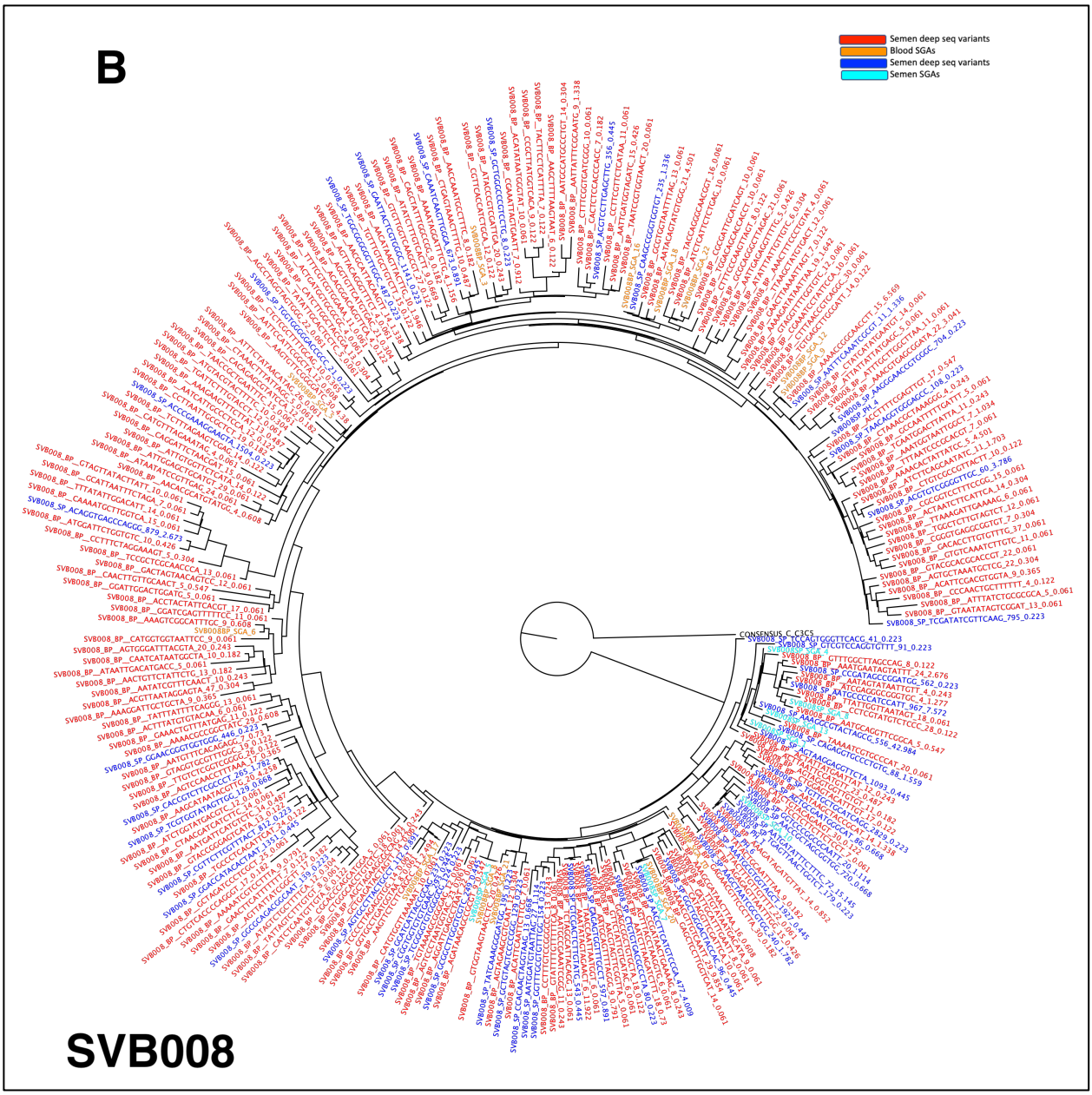
3.4.5 Phenotypic assays isolates were sampled from across the diversity found in the original HIV-1 population

To study the diversity of the clones used for phenotypic assays (neutralization and sensitivity to entry inhibitors) and how they fit within the overall population in an infected donor, the data generated from deep sequencing for the three study participants (Chapter 2 of this thesis) was combined with single genome amplification (SGA) sequencing data (Figure 36). The isolates used for phenotypic assay (chapter 3) and neutralization (chapter 4) were represented well in the diversity observed from deep sequencing of the individual donor's blood and genital tract (Figure 36).

Most of the clones (8 out of 10) from semen of donor SVB043 clustered around one semen variant from deep sequencing data (Figure 36 A). As would be expected, this was the most dominant variant found in the genital tract of this donor representing 66% of all the sequences from this compartment (Chapter 2, Figure 24). Similarly 4 out of the 12 blood

clones came from a variant represented at 38.6% (Figure 36). Clones from donors SVB008 and donor SVB049 did not form part of clonal amplification clusters.




B
Semen deep seq variants
Blood SGAs
Semen deep seq variants
Semen SGAs
SVB008

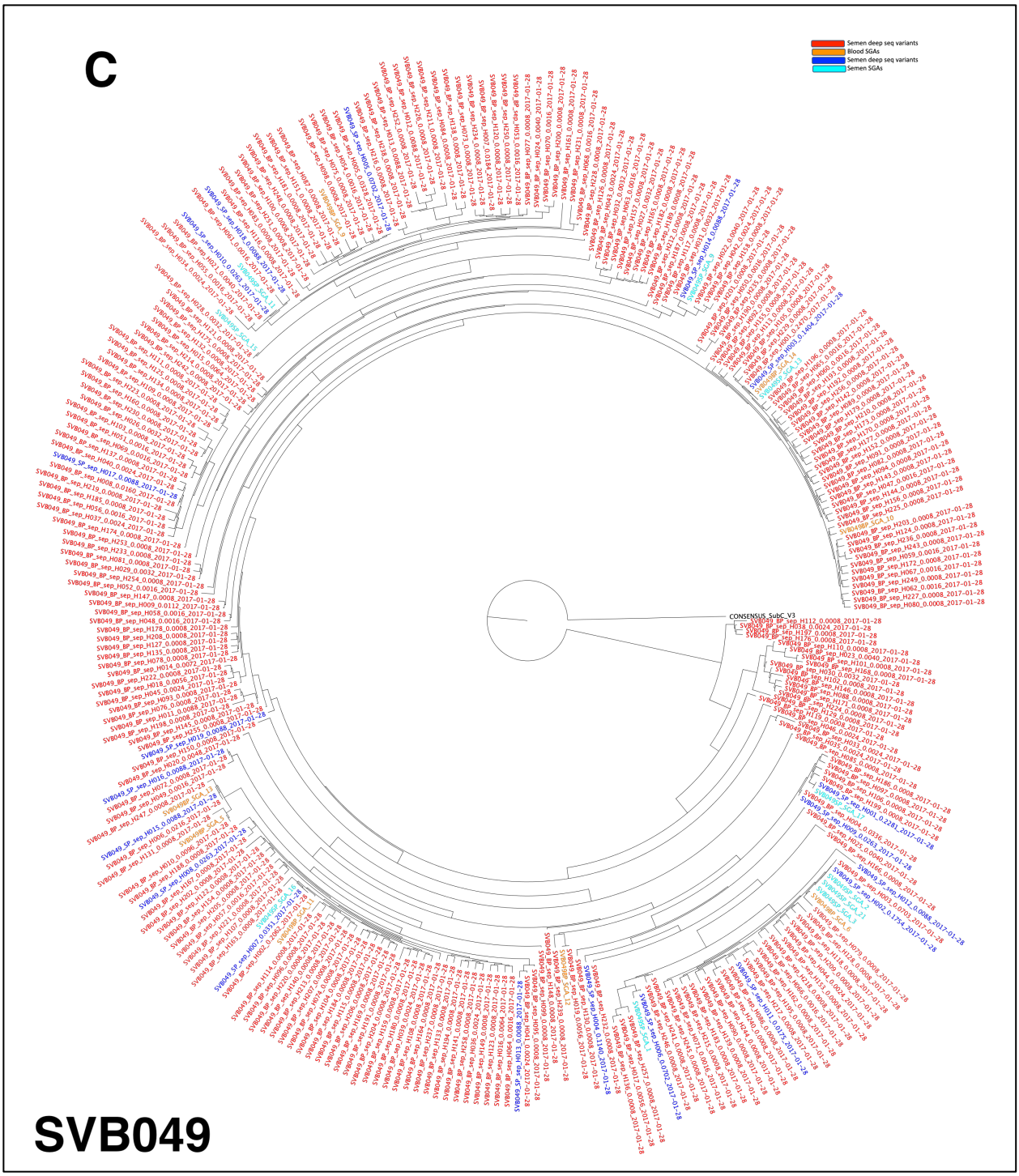


Figure 36: Phylogenetic trees generated from combined deep sequencing data and the sequences of isolates used for neutralization experiment. The diversity of neutralization isolates was well represented within the overall population of the infected donors.

3.5 Discussion

The viruses produced in the male genital tract are genetically distinct from the variants found in the blood [289, 377, 381, 396, 490]. This could reflect the tissue-specific differences or differential pressures experienced by the virus in these anatomical sites. Such changes may manifest in the variable loop characteristics (V1-V5 length and PNGS density). In addition, sensitivity to entry inhibitors provides information about the change of function of the HIV-1 envelopes as the virus moves from systemic circulation into the genital compartment. This information will dovetail the observations from study of transmitted/founder virus to find out if some of the changes might have started in the genital tract of transmitting donors.

Several studies have shown that transmitted founder (T/F) or acute HIV-1 is compact [459, 506, 510, 595] and underglycosylated [459, 506, 510, 525, 595]. Over the course of infection and disease progression, the length of the variable loops and the number of glycosylation sites increases [506, 596]. However only a handful of these studies managed to directly compare the viruses with those found in the genital tract of donors [292, 409, 490]. In order to completely understand the HIV-1 transmission bottleneck, the virus compartmentalized in the donor's genital tract need to be understood as it is the source of the potential T/F virus. Here three donors infected with HIV-1 subtype C and with highly compartmentalized viral variants in blood and semen were studied for genotypic and phenotypic variations between viruses in blood and the male genital compartment. The objective in this part of the study was to investigate whether viral isolates compartmentalized in semen and blood vary in variable loop characteristics or in interaction with viral receptors by displaying differing sensitivities to entry inhibitors.

Shorter variable loops have been reported to be a characteristic of the T/F viruses in HIV-1 subtypes C [459], A [510, 595] and D [595] but not B [510, 520, 597]. Only donor SVB043 who had the most severe compartmentalization among the donors studied had significant difference in the length of V1V2 between blood and semen. Variants from semen had longer V1V2 loops compared to variants derived from blood. The difference was in the opposite direction from what had previously been reported for T/F virus among HIV-1 subtype C [459]. Nonetheless, our observation supports the theory of local production of the virus within the

genital tract and that it responds to local selective pressures which are distinct from those in blood. This observation however neither explains the previous observation that the T/F virus in heterosexual transmissions selects for shorter variable loops [459] nor expansion of V1-V2 envelope loop over the course of infection [596]. The V1V2 region is known to contain sites important for immune recognition [598] and therefore it makes it difficult to interpret this result in this manner since as will be shown later in Chapter 4 of this thesis, semen isolates from this donor were sensitive to donor autologous antibodies. For the rest of the donors, there were no overall differences in variable loop-lengths between variants from blood and those found in the genital tract. However, multiple comparisons of inter-donor loop-lengths showed donor-to-donor variations in variable loops lengths except in the V3 region which had a constant 37 amino acids. This suggest that selective pressures within blood and semen vary from donor to donor and between anatomical sites.

The HIV-1 extracellular domain contains approximately 30 glycosylation sites, which have both structural and immune evasion roles [599, 600]. The virus is under pressure to conserve them and therefore a dynamic nature of absence and presence of the glycans in certain sites is established, which has led to the term 'evolving glycan shield', which protect the vulnerable epitopes beneath [516]. Previous studies have shown that the T/F virus is underglycosylated in subtypes A, C and D [459, 510, 595]and not B [510, 525]. The explanation is that in an immunologically naïve recipient, the glycan shield is not necessary and therefore such viruses are outcompeted by strains with a more compact and less glycosylated envelope. In this study, variation in number of glycosylation sites varied both between blood and semen compartments as well as between donors. Considering blood-semen pairs from all donors in aggregate, there was no difference in overall number of PNGS between blood and semen variants. However, considering differences per individual loop, PNGS in the V3 region reached significance overall where the semen variants were underglycosylated (Mann-Whitney $p = 0.026$). Considering donor pair-per-pair, differences in PNGS were observed for donor SVB043 (Mann-Whitney $p = 0.04648$) and SVB008 (Mann-Whitney $p = 0.01156$) although in opposite directions and no difference for donor SVB049. SVB043 who had the most compartmentalized viral populations between blood and semen, had semen sequences with more PNGS overall than the blood counterparts. This is consistent with expansion of the variable loop of the same donor detailed above which accumulates PNGS as well. In contrast, donor SVB008 had

semen variants that were underglycosylated without significant reduction in loop lengths. These observations are consistent with previous observations involving database blood-semen pairs for subtype B viruses where differences were observed in the majority of the pairs, but the direction of the difference was inconsistent [409]. Our findings are also consistent with observations of Frost *et al* in a subtype B cohort involving men-who-have-sex-with-men (MSM) [525] where only 1/6 of the epidemiologically-linked recipients had variants that were underglycosylated. It is possible for the observed discrepancy in direction of the difference that fewer glycosylation and by extent shorter loop lengths are indirect modifications that provide the structural change to the backbone that benefit the T/F virus. In support of this idea, V1V2 is known to modulate the CD4 and coreceptor binding sites altering interaction of the virus with the target cells [601, 602]. In addition, our data differs with studies of T/F or acute vs chronic viruses in the sense that these viruses might be a step further in the selection phase than variants compartmentalized in the genital tract. The T/F has gone through the selective effect in the donor's genital tract, recipient's genital tract and mucosal surfaces as well as selection to produce the disseminated infection. Each of these environments contributes to the selective process [468, 508]. It is also difficult to rule out cohort-to-cohort variations. Derdeyn et al studied a Zambian cohort of heterosexual transmissions while this thesis worked on a South African cohort in a clinic that predominantly gives care to men who have sex with men. One explanation given is that glycosylation is affected by the mode of HIV-1 transmission [506]. The HIV-1 subtype B cohort studied by Frost *et al* was dominated by male-to-male transmissions while the cohort studied by Derdeyn *et al* was subtype C and the transmissions were primarily heterosexual.

After glycans are processed in the endoplasmic reticulum and the Golgi body, the final one left is mannose-rich which is known to be neutralized by mannose-binding lectin (MBL) dominant in the female genital tract [603]. While donor SVB043 contrasts what Derdeyn et al [459] observed for HIV-1 subtype C in a cohort of transmitting partners in a heterosexual transmission setting, our overall findings agrees with these findings when transmitting pairs involving a male donor were considered in isolation. Derdeyn *et al* found underglycosylation as a feature of T/F virus in 5/8 of the study participants i.e. in all the 4 female donors and one male donor. Indeed, a more stringent transmission bottleneck for female-to-male transmission was also confirmed by the work on Carlson *et al* [489]. Our data therefore adds

to the theory that underglycosylation feature is more of a female-to-male transmission feature and not the other way.

We also considered the sensitivities of blood and semen variants to entry inhibitors by evaluating them against Maraviroc, PSC-RANTES, JM2987 and T-20. In HIV-1 subtype C virus, CXCR4 coreceptor usage is rare and our cohort was not an exception; we found that all the donors had variants that were insensitive to CXCR4 inhibitors suggesting that they were R5-tropic. This is in line with previous studies that showed that the virus found in acute or early infection is CCR5-tropic as individuals with a defective CCR5 coreceptor are protected against HIV-1 infection [457, 465, 502, 505, 506, 511, 604, 605]. The X4 viruses have also been reported to be underrepresented in semen suggesting that they are CCR5-tropic [409].

To evaluate whether semen- and blood-derived isolates vary in their dependence on CCR5 coreceptors, we performed sensitivity analysis against Maraviroc and PSC-RANTES. There were differences in sensitivity of blood and semen variants to Maraviroc, but the direction of the difference varied from donor to donor. Donor SVB043 who had the most compartmentalized viral populations between blood and semen, had semen variants that were more insensitive to maraviroc compared to blood variants. A study by Ping *et al* [506] comparing pseudotyped viruses constructed from acute and chronic infections found that the transmitted viruses were sensitive to CCR5 antagonist. Similar observations had been reported in previous studies either working on acute or the T/F virus [137, 457, 465, 502, 505, 605]; suggesting the virus establishing infection uses CCR5 coreceptor for entry which seems to contradict observations for donor SVB043 semen variants. On the other hand, SVB008 had semen variants that were sensitive to maraviroc suggesting the shaping of the potentially T/F virus may start in the genital tract. This is consistent with previous observations involving subtype C acute vs chronic viruses [502, 505, 506, 605] and also mother-to-infant transmission [606].

Susceptibility to PSC-RANTES, another CCR5 competitive inhibitor only had significant differences with donor SVB008 where the semen variants were more sensitive to inhibition than blood variants. The discordance for donor SVB043 in sensitivity to CCR5 inhibitors (Maraviroc vs PSC-RANTES) could be attributed to the difference in the mechanism of action

[607]. Maraviroc is an allosteric non-competitive inhibitor of CCR5 while PSC-RANTES acts using several mechanisms such as competitive inhibition or downregulation via internalizing the CCR5 receptor [608]. In addition, the CCR5 receptor has been reported to exist in two forms, maraviroc-sensitive and resistant-forms when the target cells have high concentration of CCR5 [506]; TZM-bl cells used in this experiment had high concentrations of CCR5 [609]. The T/F virus is known to use the maraviroc-sensitive conformation of the receptor while the chronic viruses are more promiscuous in the use of this receptor [506]. Therefore, it is possible that the genital tract HIV-1 of donor SVB043 had the maraviroc-resistant CCR5-trophic phenotype which is unlike what had been observed previously for transmitted viruses [506]. This means that they were able to use the maraviroc-bound CCR5 receptors or an alternative form that is not inhibited by maraviroc at all.

There were no differences in sensitivities to inhibition by T-20 between variants from blood and those from the genital compartment in all the 3 donors. This suggest that the structural intermediate bound by T-20 does not vary between variants from blood and those from semen.

Potential limitation of this study is the use of data from only three study participants. Even though these were the participants showing the highest extent of compartmentalization, generalization of the findings becomes a challenge.

3.6 Conclusion

The Transmitted/founder virus has been characterized with viral properties that include utilization of CCR5, reduction in lengths of the variable loops and underglycosylation. The findings have been inconsistent, and most studies lacked the information on the virus compartmentalized in the genital tract. Here we found variations in variable loop characteristics and viral entry phenotype between variants from blood and semen for three of the most compartmentalized donors; although the observed differences were not consistent in any direction. This is indicative of discordant selective pressures operating on the virus in the male genital compartment which are distinctive from the experiences in blood which also varied from donor to donor. This suggest that for a subset of the recipients, the features observed for the transmitted virus may be acquired in the genital compartment

before the virus establishes in the recipient. This information helps to understand the nature of the virus compartmentalized in semen and potentially transmitted to new recipients.

4 CHAPTER 4

Increased neutralization sensitivity of genital tract HIV-1 isolates compared to blood-derived isolates: a shift in the selective milieu shapes the population available for transmission?

4.1 Abstract

Human Immunodeficiency Virus type 1 (HIV-1) transmission by the sexual route is known to involve a transmission bottleneck, in which only one or a very few variants result in the disseminated infection in new recipients. Further, the new disseminated infection in a recipient often appears to be derived from a minor variant in the donor. This has led many scientists to the believe that the transmitted founder virus is selected to be better suited for successful transmission. Hence it might have some identifing features. Previous reports had indicated that the T/F virus is sensitive to transmitting donor antibodies in at least a portion of transmission events. Interestingly, it remains possible that selection for the traits associated with the T/F virus begins before transmission, in the transmitting donor. Therefore, we hypothesized that there may be a shift of selective milieu when the virus moves from systemic circulation into the genital tract of infected male donors. If the male genital tract is partially shielded from circulating antibodies, it would be expected that the virus found in the genital tract would be more sensitive to the donor's systemic circulating antibodies.

Blood and semen samples were collected from HIV-1 infected men. Pseudoviruses were constructed from blood and semen isolates of three donors who had the highest extent of viral compartmentalization between the two anatomical sites. Neutralization using autologous serum obtained from blood at the time of sampling was done. Heterologous neutralization using a pool of HIV-1 subtype C serum was also done.

One of the three donors, donor SVB043 with marked compartmentalization between blood and genital tract (F_{ST} = 0.47, p = 0.001, Snn = 0.9838, p = 0.001, chapter 2) had semen variants that were 1.82 (95%CI 1.06-3.13) more sensitive to donor antibodies than blood variants (Mann-Whitney test, p = 0.004429). This however was not the case for the other two subjects who had lower genetic variation between variants from blood and semen.

These results suggest that in a subset of the donors there is a shift of selective milieu when the virus moves from systemic circulation into genital tract. The change in selective milieu might result from partial shielding of the genital tract from circulating antibodies. In contrast,

the neutralization sensitivity of semen variants compared to blood variants to neutralization by antibodies from a pool of sera was not different in donor SVB043, suggesting little change in overall/global neutralization sensitivity.

These results will help in understanding the shaping of the virus found in the male genital tract that is potentially transmitted to the recipients.

4.2 Background

The Human Immunodeficiency Virus type 1 (HIV-1) found in newly established infections following successful sexual transmission can be traced back to one or very few isolates that themselves are derived from the minor populations in the presumed donor [457, 458, 469, 472]. Acutely infected subjects have a homogeneous viral population compared to the heterogenous populations found in the presumed donor which indicates a "transmission bottleneck" [459, 468, 508]. There has been continuous effort to study and characterize isolates before and after the HIV-1 transmission bottleneck, particularly to understand to what extent it reflects stochastic nonspecific barriers that restricts all isolates equally, versus selection for some identifiable traits that reflect relative advantage to establish disseminated infection in the recipients. Knowledge of such characteristics would be beneficial in understanding the process of establishing HIV-1 infections in newly-infected individuals, and perhaps in designing the HIV-1 intervention strategies.

Traditionally, the bottleneck has been associated with the virus's inefficiency in penetrating the mucosal surfaces of the recipient [164, 469]. However, more recent studies indicate that there are other processes that may play a role in selecting which virus from the donor's circulating variants becomes the transmitted/founder (T/F) [468, 508]. The physical barriers and selective pressures act at various steps of the cycle of transmission by selecting the isolates that reach the donor's genital and transmission fluid, shaping the virus within the recipient's genital tract and when crossing the recipient's genital mucosa and finally during the establishment of systemic infection in the recipient [468, 508]. Together, these barriers and selective pressures block the majority of the viruses at the transmission bottleneck [457, 458, 472]. These bottlenecks are incompletely defined and many studies appear to disagree

as to precisely what features are associated with T/F viruses [457, 467, 468, 473, 504, 505, 508, 511, 610-620]. Nonetheless, the transmitted/founder viruses appear to be heavily selected, as acute viruses tend to have different characteristics from chronic viruses in transmitting donors [459, 489, 525, 526, 621] or when acute and chronic viruses were compared [510, 511, 615].

Genital tract virus comes from local sources or shipped from blood. Using macaque models infected with simian immunodeficiency virus (SIV), it has been demonstrated that the viral variants in semen and other tissues of the male genital tract are often shipped from blood [433]. In addition, viral variants in the male genital tract have also been shown to be produced locally in addition to being imported from blood [381]. Thus, it appears that the variants available in the male genital tract may reflect an equilibrium between local production and trafficking between compartments. Indeed, trafficking back and forth between blood and the male genital tract has been confirmed by analyses by Chaillon and colleagues [303, 530]. Nonetheless, HIV-1 populations within an infected donor's blood and male genital tract compartmentalize [289, 292, 303, 377-379, 381, 385, 396, 490], meaning that the two populations are at least partially distinct. A lower genetic diversity and lower rate of molecular evolution in seminal sequences relative to those from blood have been observed [409]. In addition, blood plasma and seminal plasma viral loads correlate poorly, with some individuals having higher viral load in semen than blood suggesting independent replication of the virus in the genital compartment [205, 377, 392, 411, 622]. More surprising is that many or most of the sequences in the male genital tract are clonally amplified [381, 490]. This can explain the observation that the quasispecies in the blood of chronically infected individuals are very diverse, but the diversity seems to be reduced in the genital tract and within the transmission fluid [490, 508]. Compartmentalization may also reflect physical constraints against migration between the compartments. That there are physical constraints to migration between blood and the genital tract is suggested by the fact that some small molecules such as antiretroviral drugs appear not to readily penetrate into the genital tract compared to others [442, 444]. Together, these observations support the idea that there is independent replication of the virus in the genital tract possibly due to changed selective milieu as the virus moves from systemic circulation into the male genital tract.

There is evidence that isolates from the newly established infections are sometimes more sensitive to neutralization by antibodies from the presumed HIV-1 donor. One prominent study observed that the newly transmitted viruses in the recipients in a Zambian heterosexual transmission cohort infected with HIV-1 subtype C were more sensitive to the donor's antibodies compared to isolates derived from blood [459]. Similar observations were reported from the same cohort when infectious molecular clones were used [526]. In a similar study involving HIV-1 subtype B transmission among men who have sex with men (MSM), Frost *et al* observed 2/8 of the recipient's isolates to be sensitive to donor antibodies [525]. The authors of Frost *et al* study suggest that the transmissions with a difference in neutralization sensitivity tend to be transmissions from donors with chronic HIV-1 infections, noting that there is much less variability of HIV-1 populations in transmission donors with acute infections. It is also possible that mode of transmission and inter-clade variation might have had a role. We considered the possibility that these neutralization sensitive variants may arise much earlier in the donor's genital tract before they are transmitted into the new host.

In this study, we proposed that HIV-1 from the seminal plasma is in a genital tract site relatively protected from the antibodies and can therefore evolve into more neutralization-sensitive variants. That changes in the HIV-1 populations occur during partially separate evolution of the virus in the genital tract may explain some of the systemic changes that appear to occur during viral transmission to a new host. We are unaware of any neutralization study involving seminal HIV-1 isolates. Such information will add to the understanding of the transition from donor virus to early acute viruses in a recipient derived from the transmitted virus.

4.3 Methods

4.3.1 Study participants

The same study participants described in chapter 3 of this thesis were used. These study participants had the highest extend of compartmentalization as described in chapter 2.

4.3.2 RNA extraction and complementary DNA (cDNA) synthesis

The RNA extraction was done using QIAamp Viral RNA Mini Kit (Qiagen) and following the manufacturer's instructions as described in chapter 3 of this thesis. The extracted RNA was transcribed to cDNA using SuperScript First-Strand Synthesis System Kit (Invitrogen) as described in Chapter 3 of this thesis.

4.3.3 Single genome amplifications (SGA)

The single genome amplification (SGA) technique was used to amplify single genomes as adapted from previous studies [456-458]. The description is as detailed in Chapter 3 of this thesis.

4.3.4 Cloning and construction of pseudoviruses

The generated envelope amplicons were cloned into a mammalian expression vector pcDNA3.1D/V5.His.TOPO (Invitrogen, Carlsbad, CA, USA) as described in Chapter 3 of the thesis. Pseudoviruses were constructed using established methods [592] as described in Chapter 3 of the thesis.

4.3.5 Sequencing of the full-length envelopes from blood and semen isolates

Sequencing of the clones obtained from blood and semen was done at Stellenbosch University Sequencing facility (SUN). Twelve primers encompassing the whole of the HIV-1 envelope were used as described in Chapter 3 of the thesis.

4.3.6 Sequence analysis

The assembled sequences were aligned using MAFFT [594] and inspected in Aliview [553]. Blood and semen sequences were then combined in one alignment and maximum likelihood trees constructed using the generalized Time Reversible model (GTR) in FastTree [554]. In

order to determine how the pseudovirus *env* clones fit into the HIV-1 diversity of the study participant from which they came from, full-length envelope sequences were trimmed at either C3-V5 or V3 regions (depending upon the region analysed by Miseq deep sequencing from that study participant (see Chapter 2), combined with the sequences generated from the same donors by deep sequencing of these regions and a maximum likelihood tree was constructed. Trees were visualized in FigTree [560].

4.3.7 Serum for autologous neutralization

Serum collected from blood at the same time blood and semen samples were collected was used as the source of the study participant's own (autologous) neutralizing antibodies. The serum was heat-inactivated at 56^{o}C for an hour and stored at -80^{o}C for later use. For autologous neutralization patient-based antibodies were evaluated against same patient's blood and semen pseudoviruses.

4.3.8 Construction of HIV-1 subtype C sera pool for heterologous neutralization

A pool of heterologous sera was constructed from HIV-1 subtype C-infected South African's sera that had previously been mapped and published [623]. To make the subtype C pool, a total of 24 sera were selected from a total of 177 representing between 45th - 90th percentiles for potency, and a similar proportion of sera as the overall population of sera with dominant neutralizing antibodies specific for the membrane proximal external region (MPER), V2 apex and V3 glycans which has been previously reported [623]. TZM-bl-based neutralization assays were performed over a series of dilutions of serum in duplicate wells.

4.3.9 Neutralization HIV-1 assays

Neutralization assays were performed according to established methods [592]. Briefly, serum was first diluted 1:25 (for Donor SVB049 autologous neutralization, the isolated IgGs start concentration was 2150µg/ml) and put in duplicate in the first row of a flat-bottomed 96-well plate. A three-fold dilution was then performed down the 96-well plate. The serum was then pre-incubated for 1 hour with 50µl of the pseudovirus. Fifty microliters of freshly prepared

TZM-bl cells were then added at a final concentration of 10,000 cells/well. The cells were constituted in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal bovine serum (FBS), HEPES, antibiotics and 7.5μg/ml DEAE-dextran. Controls included TZM-bl cells alone (cell controls) and TZM-bl cells with virus (virus controls). Unrelated Murine Leukaemia virus (MLV) and a lab-adapted neutralization sensitive strain SF-162 that is very sensitive to neutralization were also used. The MLV control helped to identify the presence of viral inhibitors in serum. Incubation was done for 48 hours and then developed in a luciferase assay system. To measure luciferase activity, 100μl of the supernatant was removed followed by addition of similar volume of Bright-Glo luciferase assay substrate (Promega, Madison, USA). After a 2-minute incubation, 100μl of the mixture was transferred to 96-well opaque plate, and luminescence was measured using a VERITAS luminometer (Turner Biosystems).

4.3.10 Measurement of antiretroviral levels from donor SVB049 donor plasma

Donor SVB049 serum inhibited unrelated Murine Leukaemia Virus (MLV). Therefore levels of the most common component of first line and protease inhibitor antiretroviral therapy (ART) used in South Africa at the time of sample collection were tested. A mass spectroscopy method was used to check for the commonly used antiretroviral drugs with the longest half-lives in first line and second line antiretroviral therapy (ART), efavirenz, and lopinavir, respectively. The assay was qualitative with lower detection limits of 0.02μg/ml, substantially lower than the minimum trough blood plasma concentrations needed for efficacy, 1μg/ml for both drugs.

4.3.11 Purification of immunoglobin G (IgG) from donor SVB049

To remove the inhibitor (efavirenz) from the serum of donor SVB049, the IgGs from donor SVB049 were purified using Nab Protein G Spin Kit (Thermo Scientific) and following the recommendations of the manufacturer. Briefly, 0.5ml of serum was mixed with equal volume of PBS before loading into the provided affinity column with protein G agarose resin after equilibrating the column at room temperature. The column with the sample was then incubated at room temperature for 10 minutes with end-to-end mixing. Three washes with

the provided binding buffer were done with one-minute centrifugation after each wash. Elution was then done with the provided elution buffer into tubes containing (a tenth of the elution buffer volume) neutralization buffer. Multiple elution were done to make sure that most of the bound antibodies were eluted. Quantification of the IgGs from the elute fractions was then done using nanodrop by measuring the relative absorbance at 280nm. The fractions with the highest amount of antibody after quantification were combined and desalting done using Zeba Spin Desalting Columns (Thermo Scientific) and following the manufacturer's instructions. The purified IgGs were used for the autologous neutralization assay for donor SVB049.

4.3.12 Determination of ID_{50} values

The 50% inhibitory dilution (ID_{50}), defined as the dilution of serum that resulted in 50% inhibition of viral entry, was then determined using GraphPad Prism version 8 (GraphPad, La Jolla, USA) for donors SVB008 and SVB043. Values are displayed as the geometric mean of the ID_{50} from two independent experiments. Values for purified antibody from SVB049 are displayed as IC_{50} values, defined as the antibody concentration that resulted in 50% inhibition of viral entry. Lower IC_{50} and higher ID_{50} values indicate higher neutralization activity.

4.4 Results

4.4.1 Study participants

RNA extracted from blood and semen samples of three study participants who showed the highest extent of compartmentalization after deep sequencing and compartmentalization analysis (Table 10, and Chapter 2 of this thesis) were used to construct the pseudoviruses. As previously indicated SVB043 had the most severe compartmentalized viral populations between blood and semen followed by SVB008 and SVB049.

Table 11: Study participants with CD4 counts, viral loads in blood and semen, number of SGAs from each compartment and status of compartmentalization. Three standard techniques were used to evaluate compartmentalization; F_{ST} = Wrights measure of population subdivision, S_{nn} = Nearest neighbor statistic, SMT = Slatkin Maddison test.

SampleID	CD4 count (cells/μl)	Blood viral load (cp/ml)	Semen viral load (cp/sample)	Number of SGAs from Blood	Number of SGAs from Semen	F_{ST} pvalue	S_{nn} p value	SMT p value
SVB043	509	122297	5556596	12	10	<0.001	<0.001	<0.001
SVB008	257	309388	38640	13	10	<0.001	<0.001	<0.001
SVB049	423	85711	7915	8	10	<0.001	<0.001	<0.001

4.4.2 Donor SVB049 tested positive for efavirenz

Donor SVB049 serum neutralized unrelated MLV control during the autologous neutralization assay which was not the case for donor SVB043 and SVB008 (Figure 37). On testing for drugs with long half-lives commonly used in combination for ART in South Africa, donor SVB049 was positive for efavirenz. On purification of IgGs, the inhibitor was successfully removed (Figure 37 D). Efavirenz is a non-nucleoside reverse transcriptase inhibitor (NNRTI) that prevent the reverse transcriptase enzyme from working. This means that the drug pressure will presumably reflect more on the *pol* and not the *env* where antibody pressure mainly targets. Nonetheless, it is difficult to rule out founder effects, in which partially random *env* genes are selected because they are in the same genome as the drug selected *pol* gene. It was possible to recover viruses from this donor's blood and semen which may suggest that these viruses were resistant to efavirenz, although this was not empirically tested. It is therefore difficult to differentiate between antibody and drug pressure for donor SVB049.

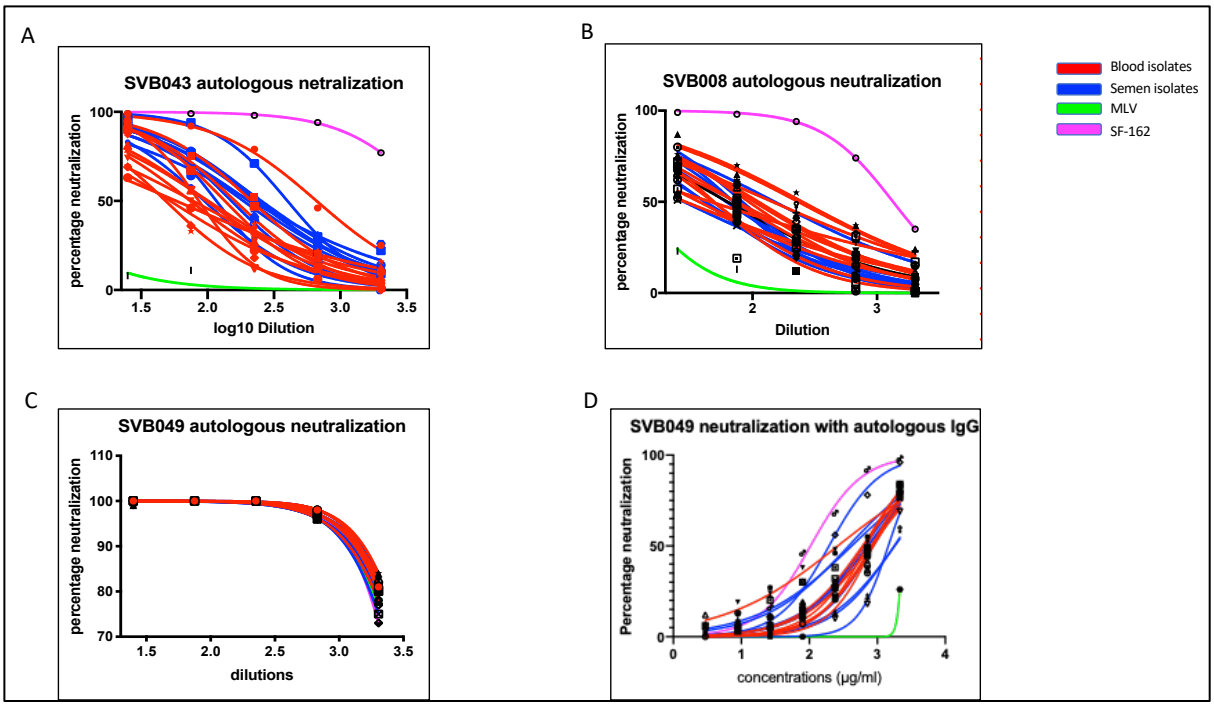


Figure 37: Neutralization profiles using donor antibodies. Pseudoviruses expressing the envelope proteins from the HIV-1 unrelated Murine Leukaemia virus (MLV) and from a neutralization-sensitive strain SF-162 pseudoviruses were used as internal controls in the neutralization experiments. The use of MLV also helped to identify the presence of viral inhibitor in donor sera (usually antiretroviral drugs). Panels A-C shows the neutralization profile for the three donors used in the neutralization experiments. Donor SVB049 serum inhibited MLV and therefore purification of IgGs was used to separate the neutralizing antibodies from other components of the serum, and neutralization repeated on successful purification (panel D).

4.4.3 Increased sensitivity to autologous neutralization was observed for the most severe blood vs semen compartmentalized donor

A total of 63 pseudoviruses were generated from blood and semen from the three donors who had the highest extent of compartmentalization determined using three standard statistical techniques as detailed in chapter 2 of this thesis. Here, pseudoviruses generated from blood and semen were tested for differences in sensitivities to autologous antibodies that were collected at the same time the samples were collected. The aim was to investigate if there was partial shielding of viral variants compartmentalized in the male genital tract viruses from autologous antibodies. This would test the change of selective milieu between variants from blood and the male genital compartment.

Sensitivities of blood and genital tract pseudoviruses to autologous antibodies varied among the highly compartmentalized donors. Donor SVB043 who exhibited particularly severe compartmentalization (F_{ST} = 0.47, p = 0.001, Snn = 0.9838, p = 0.001) had semen variants that

were 1.75 (95%CI 1.11-2.78) more sensitive to autologous antibodies (p = 0.004) (Figure 38 A). Pseudoviruses generated from the genital tract and blood for donors SVB008 and SVB049 had no differences in sensitivities to autologous antibodies (Figure 38).

Of note is that donor SVB049 who was positive for efavirenz, had blood and semen isolates with similar sensitivities to autologous IgGs. It is however difficult to distinguish between genital vs blood compartmentalization and/or selection driven by partial shielding from drugs vs the one resulting from circulating antibodies.

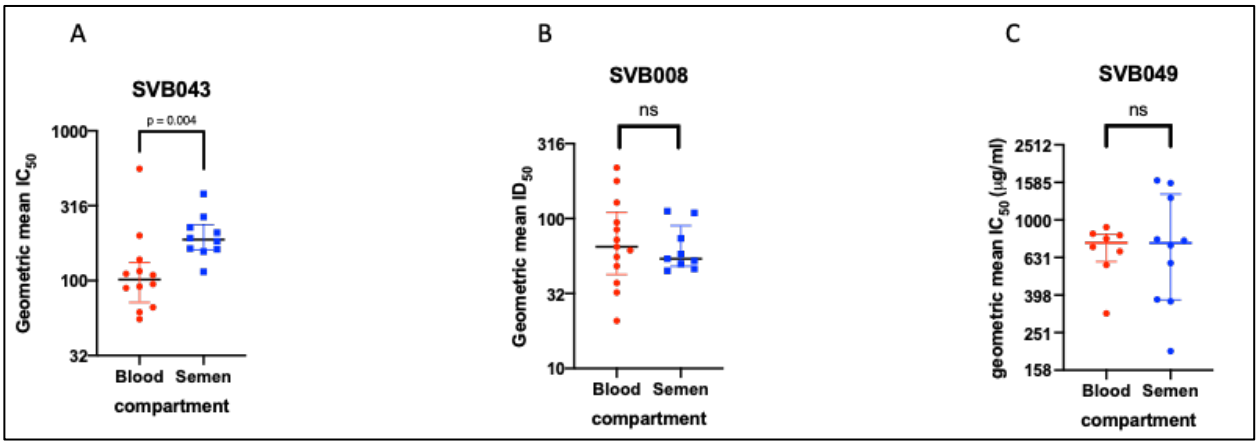


Figure 38: Sensitivities of pseudoviruses generated from blood and semen to autologous antibodies collected from donors at the same time the samples were obtained. Donor SVB043 who also showed severe compartmentalization had genital tract variants that were more sensitive to autologous antibodies compared to variants from blood. The other two donors, SVB008 and SVB049 did not show differences in sensitivities to autologous antibodies between variants from blood and genital tract. The x-axis represents the compartment where the isolates came from while the y-axis represent the log to base ten of the geometric mean calculated from two replicates of the experiments. Red colour represents the isolates from blood and blue the isolates from semen.

4.4.4 Genital tract HIV-1 isolates of SVB043 were not inherently more sensitive to neutralizing antibodies

We considered that the genital tract variants were globally more sensitive to neutralizing antibodies. There is a range of overall sensitivity to neutralizing antibodies of HIV-1 isolates [624] which has led to characterising HIV-1 isolates by "tier," with tier 1 being the most sensitive and tier 3 isolates being the least sensitive. In order to address whether changes in the overall neutralization sensitivity could explain the difference in neutralization sensitivity

to autologous antibodies observed among the clones isolated from blood vs semen of donor SVB043 [624], a HIV-1 subtype C sera pool was constructed from a cohort recruited in South Africa as previously stated. There was no overall difference in neutralization sensitivities to the subtype C serum pool between blood and semen variants for donor SVB043 and SVB008 (Figure 39). Unexpectedly, for donor SVB049, variants derived from semen were more sensitive to neutralization by heterologous antibodies compared to variants derived from blood (Figure 39). The meaning of this is not clear because in this case, there was no detected difference in sensitivity to neutralization by autologous antibodies.

Overall, these results indicate that sensitivities of genital tract strains of SVB043 to antibodies is driven by autologous antibodies and does not reflect a change in overall neutralization sensitivity to common neutralizing antibodies.

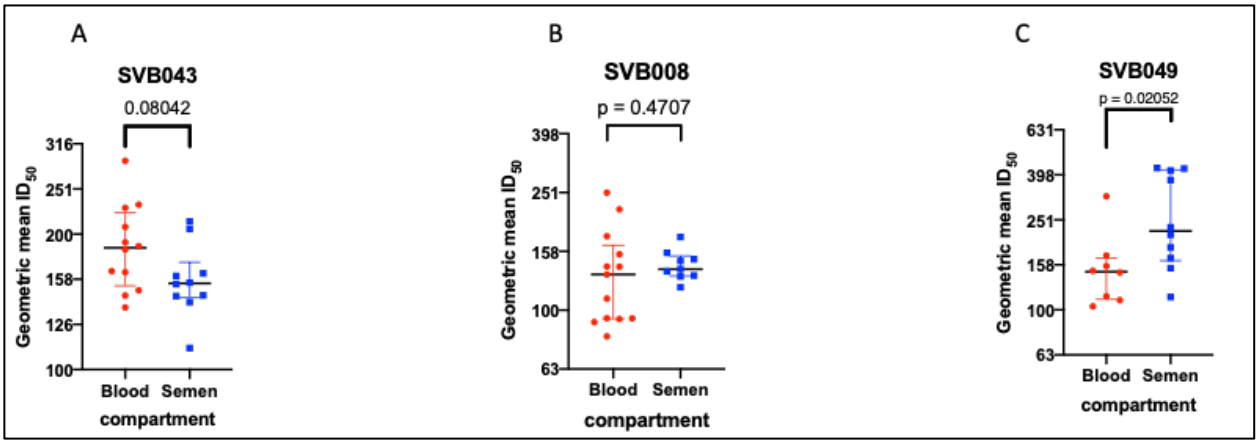


Figure 39: Sensitivities of blood and semen isolates to HIV-1 subtype C sera pool from a South African cohort. There were no differences in sensitivities of blood vs semen isolates to heterologous antibodies for donors SVB043 and SVB008. Donor SVB049 had semen variants that were more sensitive to heterologous antibodies compared to those from blood. The values on the y-axis represents the log to base 10 of the geometric mean of two replicate experiments while the x-axis represent the compartment from which the isolates came from.

4.4.5 Phylogenetic analysis of viral isolates from blood and semen plasma

The relationship between HIV-1 populations in blood and semen of the subjects who were used for neutralization experiments was done by comparing full-length envelope sequences (Figure 40). Phylogenetic reconstruction of donor SVB043 isolates from blood and semen revealed that seminal plasma isolates formed a monophyletic group that was distinct from the isolates from blood (Figure 40 A). This data is highly suggestive of a shift in the selective

milieu of the genital tract of donor SVB043, resulting from the partial shielding of genital tract HIV-1 variants from circulating antibodies**.** Alternatively, the monophyletic clade might have resulted from expansion of an infected T cell within the genital tract. This would mean that the observed selection is of the T cell receptor found on the infected T cell and not necessarily selction of HIV in them hence clonal amplification. The viral variants compartmentalized in the genital tract of donor SVB043 are evolving independently from those in blood due to possible change in the selective environment between the two compartments.

While donor SVB043 clustering was much more severe, donors SVB008 and SVB049 showed some intermingling of blood and semen sequences (Figure 40 B & C). Furthermore, donor SVB049 had more intermingled blood and semen sequences compared to donor SVB008, who had one cluster supported by a strong bootstrap value carrying more than 50% of the blood plasma sequences. The extent of phylogenetic compartmentalization represented here (SVB043 > SVB008 > SVB049) is very similar to the one shown in chapter 2 using deep sequencing data.

The antibody pressure in donors SVB008 and SVB049 did not act perceptibly differently upon the isolates from blood and those from the genital compartment. This is suggestive that the variants are recent in one compartment and have not had enough time to evolve consistently in response to the changed selective environments. Alternatively, it could also mean that there is movement of the variants between blood and the genital compartment such that any differential effect within either compartment is negated.

Figure 40: Phylogenetic trees of full length env clones isolated from donors SVB043, SVB008 and SVB049. Donor SVB043 had the most compartmentalized viral sequences between blood and genital tract. Semen sequences formed a distinct cluster represented by high bootstrap supports from blood sequences of this donor. On the other hand, donors SVB008 and SVB049

had blood and semen sequences that were partially intermingled together. The red and blue colours represent variants from blood and semen respectively. Percentages at each node reflect the level of bootstrap support for the geometry of that node.

4.5 Discussion

The nature of the HIV-1 transmission bottleneck is mainly presumed to include effects from the integrity of the mucosal surfaces of the recipient's genital tract [469]. However, other characteristics of the isolate may also be important in determining its transmissibility [468, 508]. Three previous reports have suggested that the HIV-1 isolates found in the newly infected recipient are sometimes sensitive to neutralization by antibodies derived from transmitting donor [459, 525, 526]. These studies however did not test whether selection for the neutralization sensitivity phenotype of transmitted/ founder (T/F) virus might have started much earlier in the transmitting donor. The virus that must be antagonized by a successful vaccine or by mucosal interventions during sexual transmission is the one found in the genital tract.

Donor SVB043 who had the most severe compartmentalized viral populations between blood and semen, had semen variants that were more sensitive to autologous neutralization than their blood counterparts. This observation supports the idea of partial shielding of the genital tract viral variants from circulating antibodies, i.e. that the release of antibody pressure from isolates in the genital tract leads to variants that are more sensitive to donor antibodies than circulating blood isolates. Importantly, it thus appears that neutralization sensitivity of transmitted isolates may arise in the donor's genital tract before transmission to the new host, at least in some of the donors. To our knowledge, this has not been reported previously. Of the three reports of increased sensitivity of newly transmitted variants to neutralization by antibodies of the donor, one study observed it in all transmission pairs studied [459], one observed it in 2/8 transmission pairs studied [525] and the third study was only able to detect the difference by aggregating data from several transmission pairs [526]. It thus seems likely that not all transmission pairs exhibit a change in neutralization sensitivity upon transmission. This matches our results, in which only one out of three study participants exhibited increased autologous neutralization sensitivity of semen isolates compared to blood isolates. It is thus

possible (although speculative) that the increased sensitivity in recipients only occurs if there is increased sensitivity of the genital tract isolates in the transmission donor.

This work therefore broadens the previous observations by Derdeyn et al [459], Deymier et al [526] and Frost et al [597] who observed neutralization sensitivity of recipient viruses to antibodies from the transmitting donor. In Derdeyn et al work [459], four of the transmitting pairs involved male donors and two had envelope clones that were obtained from plasma (the other two came from PBMC source). Based on the observations in this thesis, it is possible that selection of the neutralization sensitivity to donor antibodies phenotype in some recipients might have started in the genital tract of the donors; at least for the two donors in Derdeyn et al [459] whose source of clones was plasma.

On the other hand, donors SVB008 and SVB049 who also had high extent of compartmentalization but less severe than that of SVB043, exhibited no detectable difference in sensitivities of blood and semen viruses to autologous antibodies. The discrepancy could be explained by the fact that the barrier between blood and the genital compartment was not completely tight as evidenced by the intermingling of some blood and semen sequences in the phylogenetic trees. It is possible that the semen variants of these donors were recent in the genital tract and had not had enough time to experience the changed selective milieu within the genital compartment. It is also possible that these donors are more recently infected and have less neutralising antibody with which to select viral populations in the blood, the reason that Frost *et al* [525] cited to explain why, in their study, only some of the HIV-1 transmissions were associated with an increased in sensitivity of isolates from the recipient's disseminated infection. An additional confounder for donor SVB049 might be the use of antiretroviral drug which makes it difficult to single out drug and antibody pressure.

There was no difference in sensitivity of blood and semen variants to heterologous antibodies (from HIV-1 subtype C pool) for donor SVB043 who previously had semen variants that were sensitive to autologous antibodies. This suggest that the change in sensitivity to antibodies of semen-derived isolates versus those in blood does not reflect overall change in neutralization. Though the viruses were steps further in selection towards transmitted/founder viruses, the work of Derdeyn et al [459] and Ping et al [506] had reported lack of differences in sensitivity

of donor and recipient viral variants to heterologous subtype C sera in the context of heterosexual transmission. Our data therefore suggest that the overall sensitivities to antibodies for blood vs genital tract variants is donor-specific rather than a HIV-1 subtype C population-based feature. The only donor who exhibited a difference in inherent neutralization to a serum pool was donor SVB049, who was on antiretroviral therapy, which could have resulted in selective pressures that affected the inherent sensitivity of viruses in one or the other compartment.

In conclusion, this study is the first to show that in some donors, the neutralization sensitivity feature of potentially transmitted/founder viruses begins in the donor's genital tract. In addition, we show that the genital tract has altered antibody selective milieu that shape the virus on the way to becoming a transmitted/founder virus in one among the three study participants that we analysed. This strongly suggests that the selection of HIV-1 populations in the genital tract can substantially and meaningfully alter which viral isolates can transmit. Therefore the trait of selection of donor antibody-sensitive viruses during transmission observed in previous studies may well had started in the genital tract of the transmitting donors in a subset of the cases.

Candidates work and contributions from others

All experiments were done by the candidate (blood and semen sample processing, library preparation for the next generation sequencing, single genome amplifications, cloning, neutralization experiments and phenotypic assays).
Sample collection was done by a clinician in an accredited facility, Viral RNA sequencing was done at Stellenbosch University (Sanger sequencing) and NICD's Sequencing core facility (high throughput sequencing) of which they are referenced in this thesis.
In addition, the deep sequence data analysis pipeline was made by other colleagues in the lab, Colin Anthony and Dave Marten of which are acknowledged in this thesis.
Partly, I received advice from UCT's computation Biology Unit under Prof Darren Martin.

5 References

1. Fauci AS, Morens DM: **The perpetual challenge of infectious diseases**. *N Engl J Med* 2012, **366**(5):454-461.
2. UNAIDS: **Global HIV & AIDS statistics — 2018 fact sheet**. In.; 2018.
3. Baggaley RF, White RG, Hollingsworth TD, Boily MC: **Heterosexual HIV-1 infectiousness and antiretroviral use: systematic review of prospective studies of discordant couples**. *Epidemiology (Cambridge, Mass)* 2013, **24**(1):110-121.
4. Yoshimura K: **Current status of HIV/AIDS in the ART era**. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy* 2017, **23**(1):12-16.
5. Kaul R, Prodger J, Joag V, Shannon B, Yegorov S, Galiwango R, McKinnon L: **Inflammation and HIV Transmission in Sub-Saharan Africa**. *Current HIV/AIDS reports* 2015, **12**(2):216-222.
6. Nash D, Yotebieng M, Sohn AH: **Treating all people living with HIV in sub-Saharan Africa: a new era calling for new approaches**. *Journal of virus eradication* 2018, **4**(Suppl 2):1-4.
7. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A: **Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency**. *N Engl J Med* 1981, **305**(24):1425-1431.
8. **Persistent, generalized lymphadenopathy among homosexual males**. *MMWR Morbidity and mortality weekly report* 1982, **31**(19):249-251.
9. **Opportunistic infections and Kaposi's sarcoma among Haitians in the United States**. *MMWR Morbidity and mortality weekly report* 1982, **31**(26):353-354, 360-351.
10. Gottlieb MS: **Pneumocystis pneumonia--Los Angeles. 1981**. *American journal of public health* 2006, **96**(6):980-981; discussion 982-983.

11. Popovic M, Sarngadharan MG, Read E, Gallo RC: **Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS**. *Science* 1984, **224**(4648):497-500.
12. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B *et al*: **Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS**. *Science* 1984, **224**(4648):500-503.
13. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S *et al*: **Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS)**. *Science* 1983, **220**(4599):865-867.
14. Levy JA, Hoffman AD, Kramer SM, Landis JA, Shimabukuro JM, Oshiro LS: **Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS**. *Science* 1984, **225**(4664):840-842.
15. Duesberg PH: **AIDS epidemiology: inconsistencies with human immunodeficiency virus and with infectious disease**. *Proceedings of the National Academy of Sciences of the United States of America* 1991, **88**(4):1575-1579.
16. Chigwedere P, Seage GR, 3rd, Gruskin S, Lee TH, Essex M: **Estimating the lost benefits of antiretroviral drug use in South Africa**. *Journal of acquired immune deficiency syndromes* 2008, **49**(4):410-415.
17. Small CB, Klein RS, Friedland GH, Moll B, Emeson EE, Spigland I: **Community-acquired opportunistic infections and defective cellular immunity in heterosexual drug abusers and homosexual men**. *The American journal of medicine* 1983, **74**(3):433-441.
18. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA: **The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus**. *Nature* 1984, **312**(5996):763-767.
19. Kohler G, Milstein C: **Continuous cultures of fused cells secreting antibody of predefined specificity**. *Nature* 1975, **256**(5517):495-497.
20. Gallo RC, Montagnier L: **The discovery of HIV as the cause of AIDS**. *N Engl J Med* 2003, **349**(24):2283-2285.
21. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C *et al*: **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)**. *Science* 1983, **220**(4599):868-871.
22. Lever AM, Berkhout B: **2008 Nobel prize in medicine for discoverers of HIV**. *Retrovirology* 2008, **5**:91.
23. Karpas A: **Human retroviruses in leukaemia and AIDS: reflections on their discovery, biology and epidemiology**. *Biological reviews of the Cambridge Philosophical Society* 2004, **79**(4):911-933.
24. de-The G, Gessain A, Gazzolo L, Robert-Guroff M, Najberg G, Calender A, Peti M, Brubaker G, Bensliman A, Fabry F *et al*: **Comparative seroepidemiology of HTLV-I and HTLV-III in the French West Indies and some African countries**. *Cancer research* 1985, **45**(9 Suppl):4633s-4636s.
25. (CDC) CfDC: **Epidemiologic Notes and Reports Pneumocystis carinii Pneumonia among Persons with Hemophilia A**. In.; 1982.
26. Dimopoulos Y, Moysi E, Petrovas C: **The Lymph Node in HIV Pathogenesis**. *Current HIV/AIDS reports* 2017, **14**(4):133-140.
27. Brun-Vezinet F, Rouzioux C, Barre-Sinoussi F, Klatzmann D, Saimot AG, Rozenbaum W, Christol D, Gluckmann JC, Montagnier L, Chermann JC: **Detection of IgG antibodies to lymphadenopathy-associated virus in patients with AIDS or lymphadenopathy syndrome**. *Lancet* 1984, **1**(8389):1253-1256.

28. Schupbach J, Popovic M, Gilden RV, Gonda MA, Sarngadharan MG, Gallo RC: **Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS**. *Science* 1984, **224**(4648):503-505.
29. Broder CC, Collman RG: **Chemokine receptors and HIV**. *Journal of leukocyte biology* 1997, **62**(1):20-29.
30. Moore JP, Trkola A, Dragic T: **Co-receptors for HIV-1 entry**. *Current opinion in immunology* 1997, **9**(4):551-562.
31. Schmitt FA, Bigley JW, McKinnis R, Logue PE, Evans RW, Drucker JL: **Neuropsychological outcome of zidovudine (AZT) treatment of patients with AIDS and AIDS-related complex**. *N Engl J Med* 1988, **319**(24):1573-1578.
32. Fischl MA, Richman DD, Grieco MH, Gottlieb MS, Volberding PA, Laskin OL, Leedom JM, Groopman JE, Mildvan D, Schooley RT *et al*: **The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial**. *N Engl J Med* 1987, **317**(4):185-191.
33. Barre-Sinoussi F, Ross AL, Delfraissy JF: **Past, present and future: 30 years of HIV research**. *Nature reviews Microbiology* 2013, **11**(12):877-883.
34. Clavel F, Mansinho K, Chamaret S, Guetard D, Favier V, Nina J, Santos-Ferreira MO, Champalimaud JL, Montagnier L: **Human immunodeficiency virus type 2 infection associated with AIDS in West Africa**. *N Engl J Med* 1987, **316**(19):1180-1185.
35. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C *et al*: **Isolation of a new human retrovirus from West African patients with AIDS**. *Science* 1986, **233**(4761):343-346.
36. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V, Damond F, Robertson DL, Simon F: **A new human immunodeficiency virus derived from gorillas**. *Nature medicine* 2009, **15**(8):871-872.
37. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C *et al*: **HIV-1 nomenclature proposal**. *Science* 2000, **288**(5463):55-56.
38. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, St Louis D, Burke DS, McCutchan FE: **Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand**. *J Virol* 1996, **70**(9):5935-5943.
39. Tongo M, Dorfman JR, Martin DP: **High Degree of HIV-1 Group M (HIV-1M) Genetic Diversity within Circulating Recombinant Forms: Insight into the Early Events of HIV-1M Evolution**. *J Virol* 2016, **90**(5):2221-2229.
40. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J *et al*: **HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations**. *Science (New York, NY)* 2014, **346**(6205):56.
41. Tongo M, Martin DP, Zembe L, Mpoudi-Ngole E, Williamson C, Burgers WA: **Characterization of HIV-1 gag and nef in Cameroon: further evidence of extreme diversity at the origin of the HIV-1 group M epidemic**. *Virology journal* 2013, **10**(1):29.
42. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, Sema H, Tshimanga K, Bongo B, Delaporte E: **Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa**. *J Virol* 2000, **74**(22):10498-10507.
43. Damond F, Worobey M, Campa P, Farfara I, Colin G, Matheron S, Brun-Vezinet F, Robertson DL, Simon F: **Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification**. *AIDS research and human retroviruses* 2004, **20**(6):666-672.
44. Matheron S, Pueyo S, Damond F, Simon F, Lepretre A, Campa P, Salamon R, Chene G, Brun-Vezinet F, French HIVCSG: **Factors associated with clinical progression in HIV-2 infected-patients: the French ANRS cohort**. *Aids* 2003, **17**(18):2593-2601.

45. Popper SJ, Sarr AD, Travers KU, Gueye-Ndiaye A, Mboup S, Essex ME, Kanki PJ: **Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2**. *The Journal of infectious diseases* 1999, **180**(4):1116-1121.
46. Nyamweya S, Hegedus A, Jaye A, Rowland-Jones S, Flanagan KL, Macallan DC: **Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis**. *Rev Med Virol* 2013, **23**(4):221-240.
47. Azevedo-Pereira JM, Santos-Costa Q: **HIV Interaction With Human Host: HIV-2 As a Model of a Less Virulent Infection**. *AIDS reviews* 2016, **18**(1):44-53.
48. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pepin J *et al*: **HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations**. *Science* 2014, **346**(6205):56-61.
49. Piot P, Quinn TC, Taelman H, Feinsod FM, Minlangu KB, Wobin O, Mbendi N, Mazebo P, Ndangi K, Stevens W *et al*: **Acquired immunodeficiency syndrome in a heterosexual population in Zaire**. *Lancet* 1984, **2**(8394):65-69.
50. Van de Perre P, Rouvroy D, Lepage P, Bogaerts J, Kestelyn P, Kayihigi J, Hekker AC, Butzler JP, Clumeck N: **Acquired immunodeficiency syndrome in Rwanda**. *Lancet* 1984, **2**(8394):62-65.
51. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD: **An African HIV-1 sequence from 1959 and implications for the origin of the epidemic**. *Nature* 1998, **391**(6667):594-597.
52. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, Muyembe JJ, Kabongo JM, Kalengayi RM, Van Marck E *et al*: **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960**. *Nature* 2008, **455**(7213):661-664.
53. Hahn BH, Shaw GM, De Cock KM, Sharp PM: **AIDS as a zoonosis: scientific and public health implications**. *Science* 2000, **287**(5453):607-614.
54. Chen Z, Telfier P, Gettie A, Reed P, Zhang L, Ho DD, Marx PA: **Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop**. *J Virol* 1996, **70**(6):3617-3627.
55. Sharp PM, Bailes E, Gao F, Beer BE, Hirsch VM, Hahn BH: **Origins and evolution of AIDS viruses: estimating the time-scale**. *Biochemical Society transactions* 2000, **28**(2):275-282.
56. Peeters M, Jung M, Ayouba A: **The origin and molecular epidemiology of HIV**. *Expert review of anti-infective therapy* 2013, **11**(9):885-896.
57. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T: **Timing the ancestor of the HIV-1 pandemic strains**. *Science* 2000, **288**(5472):1789-1796.
58. Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, Li Y, Rudicell RS, Learn GH, Neel C *et al*: **Origin and biology of simian immunodeficiency virus in wild-living western gorillas**. *J Virol* 2009, **83**(4):1635-1648.
59. D'Arc M, Ayouba A, Esteban A, Learn GH, Boue V, Liegeois F, Etienne L, Tagg N, Leendertz FH, Boesch C *et al*: **Origin of the HIV-1 group O epidemic in western lowland gorillas**. *Proceedings of the National Academy of Sciences of the United States of America* 2015, **112**(11):E1343-1352.
60. De Leys R, Vanderborght B, Vanden Haesevelde M, Heyndrickx L, van Geel A, Wauters C, Bernaerts R, Saman E, Nijs P, Willems B *et al*: **Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin**. *J Virol* 1990, **64**(3):1207-1216.
61. Gurtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, Tsague JM, Kaptue L: **A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon**. *J Virol* 1994, **68**(3):1581-1585.

62. Charneau P, Borman AM, Quillent C, Guetard D, Chamaret S, Cohen J, Remy G, Montagnier L, Clavel F: **Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group**. *Virology* 1994, **205**(1):247-253.
63. Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barre-Sinoussi F, Brun-Vezinet F: **Identification of a new human immunodeficiency virus type 1 distinct from group M and group O**. *Nature medicine* 1998, **4**(9):1032-1037.
64. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, Mbanya D, Kaptue L, Ndembi N, Gurtler L *et al*: **Confirmation of putative HIV-1 group P in Cameroon**. *J Virol* 2011, **85**(3):1403-1407.
65. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR: **An African primate lentivirus (SIVsm) closely related to HIV-2**. *Nature* 1989, **339**(6223):389-392.
66. UNAIDS: **UNAIDS data 2018**. In. Edited by UNAIDS; 2018.
67. Johnston LG, Sabin ML, Prybylski D, Sabin K, McFarland W, Baral S, Kim AA, Raymond HF: **The importance of assessing self-reported HIV status in bio-behavioural surveys**. *Bulletin of the World Health Organization* 2016, **94**(8):605-612.
68. Mumtaz GR, Weiss HA, Thomas SL, Riome S, Setayesh H, Riedner G, Semini I, Tawil O, Akala FA, Wilson D *et al*: **HIV among people who inject drugs in the Middle East and North Africa: systematic review and data synthesis**. *PLoS medicine* 2014, **11**(6):e1001663.
69. Al-Abri S, Mokhbat JE: **HIV in the MENA Region: Cultural and Political Challenges**. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2016, **44**:64-65.
70. UNAIDS: **FACT SHEET – JULY 2018**. In.; 2018.
71. DeHovitz J, Uuskula A, El-Bassel N: **The HIV epidemic in Eastern Europe and Central Asia**. *Current HIV/AIDS reports* 2014, **11**(2):168-176.
72. Masharsky AE, Dukhovlinova EN, Verevochkin SV, Toussova OV, Skochilov RV, Anderson JA, Hoffman I, Cohen MS, Swanstrom R, Kozlov AP: **A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia**. *The Journal of infectious diseases* 2010, **201**(11):1697-1702.
73. 2018 GAm: **indicators for monitoring the 2016 United Nations Political Declaration on HIV and AIDS**. In.; 2018.
74. Hemelaar J, Gouws E, Ghys PD, Osmanov S, Isolation W-UNfH, Characterisation: **Global trends in molecular epidemiology of HIV-1 during 2000-2007**. *Aids* 2011, **25**(5):679-689.
75. Parczewski M, Urbanska A, Grzeszczuk A, Maciejewska K, Witak-Jedra M, Leszczyszyn-Pynka M: **Phylogeographic analysis on the travel-related introduction of HIV-1 non-B subtypes to Northern Poland**. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2014, **27**:121-130.
76. Mumtaz G, Hilmi N, Akala FA, Semini I, Riedner G, Wilson D, Abu-Raddad LJ: **HIV-1 molecular epidemiology evidence and transmission patterns in the Middle East and North Africa**. *Sexually transmitted infections* 2011, **87**(2):101-106.
77. Perrin L, Kaiser L, Yerly S: **Travel and the spread of HIV-1 genetic variants**. *The Lancet infectious diseases* 2003, **3**(1):22-27.
78. Simon V, Ho DD, Abdool Karim Q: **HIV/AIDS epidemiology, pathogenesis, prevention, and treatment**. *Lancet* 2006, **368**(9534):489-504.
79. Weiss RA: **How does HIV cause AIDS?** *Science* 1993, **260**(5112):1273-1279.
80. Moir S, Chun TW, Fauci AS: **Pathogenic mechanisms of HIV disease**. *Annual review of pathology* 2011, **6**:223-248.
81. Shattock RJ, Moore JP: **Inhibiting sexual transmission of HIV-1 infection**. *Nature reviews Microbiology* 2003, **1**(1):25-34.
82. Svicher V, Ceccherini-Silberstein F, Antinori A, Aquaro S, Perno CF: **Understanding HIV compartments and reservoirs**. *Current HIV/AIDS reports* 2014, **11**(2):186-194.

83. Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, Beilman GJ, Nguyen PL, Khoruts A, Larson M, Haase AT *et al*: **CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract**. *The Journal of experimental medicine* 2004, **200**(6):749-759.
84. Hait SH, Darc M, Machado ES, Soares EA, Sprinz E, Soares MA: **Conservation of the alpha4beta7 lymphocyte homing receptor in HIV-infected patients with distinct transmission routes and disease progression profiles**. *AIDS research and human retroviruses* 2014, **30**(5):493-497.
85. Mehandru S, Poles MA, Tenner-Racz K, Horowitz A, Hurley A, Hogan C, Boden D, Racz P, Markowitz M: **Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract**. *The Journal of experimental medicine* 2004, **200**(6):761-770.
86. Health MFfAIaS: **HIV in primary care: an essential guide for GPs, practice nurses and other members of the primary healthcare team**. *Foundation for A IDS and Sexual Health* 2011.
87. Macneal RJ, Dinulos JG: **Acute retroviral syndrome**. *Dermatologic clinics* 2006, **24**(4):431-438, v.
88. Suthar AB, Granich RM, Kato M, Nsanzimana S, Montaner JS, Williams BG: **Programmatic Implications of Acute and Early HIV Infection**. *The Journal of infectious diseases* 2015, **212**(9):1351-1360.
89. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH *et al*: **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection**. *Aids* 2003, **17**(13):1871-1879.
90. Piatak M, Jr., Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, Hahn BH, Shaw GM, Lifson JD: **High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR**. *Science* 1993, **259**(5102):1749-1754.
91. Naif HM: **Pathogenesis of HIV Infection**. *Infectious disease reports* 2013, **5**(Suppl 1):e6.
92. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ *et al*: **Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes**. *Science* 1999, **283**(5403):857-860.
93. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, Farthing C, Ho DD: **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome**. *J Virol* 1994, **68**(7):4650-4655.
94. Haynes BF, Pantaleo G, Fauci AS: **Toward an understanding of the correlates of protective immunity to HIV infection**. *Science* 1996, **271**(5247):324-328.
95. Goto T, Nakai M, Ikuta K: **The life-cycle of human immunodeficiency virus type 1**. *Micron (Oxford, England : 1993)* 1998, **29**(2-3):123-138.
96. Stannard LM, van der Riet FD, Moodie JW: **The morphology of human immunodeficiency virus particles by negative staining electron microscopy**. *The Journal of general virology* 1987, **68 (Pt 3)**:919-923.
97. Nakai M, Goto T, Imura S: **Ultrastructural features of the AIDS virus (HIV) and its morphogenesis**. *Journal of electron microscopy technique* 1989, **12**(2):95-100.
98. Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ: **Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus**. *Nature* 1985, **313**(6002):450-458.
99. Bryant M, Ratner L: **Myristoylation-dependent replication and assembly of human immunodeficiency virus 1**. *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(2):523-527.
100. Gottlinger HG, Sodroski JG, Haseltine WA: **Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1**.

Proceedings of the National Academy of Sciences of the United States of America 1989, **86**(15):5781-5785.
101. Gallay P, Swingler S, Song J, Bushman F, Trono D: **HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase**. *Cell* 1995, **83**(4):569-576.
102. Thali M, Bukovsky A, Kondo E, Rosenwirth B, Walsh CT, Sodroski J, Gottlinger HG: **Functional association of cyclophilin A with HIV-1 virions**. *Nature* 1994, **372**(6504):363-365.
103. Paxton W, Connor RI, Landau NR: **Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis**. *J Virol* 1993, **67**(12):7229-7237.
104. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE: **Characterization of ribosomal frameshifting in HIV-1 gag-pol expression**. *Nature* 1988, **331**(6153):280-283.
105. Parkin NT, Chamorro M, Varmus HE: **Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo**. *J Virol* 1992, **66**(8):5147-5151.
106. Chiu TK, Davies DR: **Structure and function of HIV-1 integrase**. *Current topics in medicinal chemistry* 2004, **4**(9):965-977.
107. Bernstein HB, Tucker SP, Kar SR, McPherson SA, McPherson DT, Dubay JW, Lebowitz J, Compans RW, Hunter E: **Oligomerization of the hydrophobic heptad repeat of gp41**. *J Virol* 1995, **69**(5):2745-2750.
108. Landau NR, Warton M, Littman DR: **The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4**. *Nature* 1988, **334**(6178):159-162.
109. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM *et al*: **Identification of a major co-receptor for primary isolates of HIV-1**. *Nature* 1996, **381**(6584):661-666.
110. Van den Bergh R, Morin S, Sass HJ, Grzesiek S, Vekemans M, Florence E, Tran HT, Imiru RG, Heyndrickx L, Vanham G *et al*: **Monocytes contribute to differential immune pressure on R5 versus X4 HIV through the adipocytokine visfatin/NAMPT**. *PloS one* 2012, **7**(4):e35074.
111. Kamp C: **Understanding the HIV coreceptor switch from a dynamical perspective**. *BMC evolutionary biology* 2009, **9**:274.
112. Ruben S, Perkins A, Purcell R, Joung K, Sia R, Burghoff R, Haseltine WA, Rosen CA: **Structural and functional characterization of human immunodeficiency virus tat protein**. *J Virol* 1989, **63**(1):1-8.
113. Harrich D, Ulich C, Gaynor RB: **A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription**. *J Virol* 1996, **70**(6):4017-4027.
114. Feinberg MB, Baltimore D, Frankel AD: **The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation**. *Proceedings of the National Academy of Sciences of the United States of America* 1991, **88**(9):4045-4049.
115. Kao SY, Calman AF, Luciw PA, Peterlin BM: **Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product**. *Nature* 1987, **330**(6147):489-493.
116. Zapp ML, Green MR: **Sequence-specific RNA binding by the HIV-1 Rev protein**. *Nature* 1989, **342**(6250):714-716.
117. Bartel DP, Zapp ML, Green MR, Szostak JW: **HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA**. *Cell* 1991, **67**(3):529-536.
118. Malim MH, Bohnlein S, Hauber J, Cullen BR: **Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function**. *Cell* 1989, **58**(1):205-214.
119. Hope TJ, McDonald D, Huang XJ, Low J, Parslow TG: **Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus**. *J Virol* 1990, **64**(11):5360-5366.

120. Wen W, Meinkoth JL, Tsien RY, Taylor SS: **Identification of a signal for rapid export of proteins from the nucleus**. *Cell* 1995, **82**(3):463-473.
121. Cohen EA, Dehni G, Sodroski JG, Haseltine WA: **Human immunodeficiency virus vpr product is a virion-associated regulatory protein**. *J Virol* 1990, **64**(6):3097-3099.
122. Zack JA, Arrigo SJ, Weitsman SR, Go AS, Haislip A, Chen IS: **HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure**. *Cell* 1990, **61**(2):213-222.
123. Heinzinger NK, Bukinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M, Emerman M: **The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells**. *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**(15):7311-7315.
124. Vodicka MA, Koepp DM, Silver PA, Emerman M: **HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection**. *Genes & development* 1998, **12**(2):175-185.
125. Rogel ME, Wu LI, Emerman M: **The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection**. *J Virol* 1995, **69**(2):882-888.
126. Sato A, Igarashi H, Adachi A, Hayami M: **Identification and localization of vpr gene product of human immunodeficiency virus type 1**. *Virus genes* 1990, **4**(4):303-312.
127. Schubert U, Bour S, Ferrer-Montiel AV, Montal M, Maldarell F, Strebel K: **The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains**. *J Virol* 1996, **70**(2):809-819.
128. Guatelli JC: **Interactions of viral protein U (Vpu) with cellular factors**. *Current topics in microbiology and immunology* 2009, **339**:27-45.
129. Pereira EA, daSilva LL: **HIV-1 Nef: Taking Control of Protein Trafficking**. *Traffic (Copenhagen, Denmark)* 2016, **17**(9):976-996.
130. Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM: **Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein**. *Nature medicine* 1996, **2**(3):338-342.
131. Garcia JV, Miller AD: **Downregulation of cell surface CD4 by nef**. *Research in virology* 1992, **143**(1):52-55.
132. Aiken C, Konner J, Landau NR, Lenburg ME, Trono D: **Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain**. *Cell* 1994, **76**(5):853-864.
133. Ray N, Doms RW: **HIV-1 coreceptors and their inhibitors**. *Current topics in microbiology and immunology* 2006, **303**:97-120.
134. Shen R, Richter HE, Smith PD: **Early HIV-1 target cells in human vaginal and ectocervical mucosa**. *American journal of reproductive immunology* 2011, **65**(3):261-267.
135. Earl PL, Doms RW, Moss B: **Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein**. *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(2):648-652.
136. Weiss RA: **Thirty years on: HIV receptor gymnastics and the prevention of infection**. *BMC biology* 2013, **11**:57.
137. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, Malnati MS, Plebani A, Siccardi AG, Littman DR *et al*: **In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression**. *Nature medicine* 1997, **3**(11):1259-1265.
138. Fanales-Belasio E, Raimondo M, Suligoi B, Butto S: **HIV virology and pathogenetic mechanisms of infection: a brief overview**. *Annali dell'Istituto superiore di sanita* 2010, **46**(1):5-14.
139. Melikyan GB: **HIV entry: a game of hide-and-fuse?** *Current opinion in virology* 2014, **4**:1-7.

140. Gotte M, Li X, Wainberg MA: **HIV-1 reverse transcription: a brief overview focused on structure-function relationships among molecules involved in initiation of the reaction**. *Archives of biochemistry and biophysics* 1999, **365**(2):199-210.
141. Miller MD, Farnet CM, Bushman FD: **Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition**. *J Virol* 1997, **71**(7):5382-5390.
142. Piller SC, Caly L, Jans DA: **Nuclear import of the pre-integration complex (PIC): the Achilles heel of HIV?** *Current drug targets* 2003, **4**(5):409-429.
143. Liao Z, Cimakasky LM, Hampton R, Nguyen DH, Hildreth JE: **Lipid rafts and HIV pathogenesis: host membrane cholesterol is required for infection by HIV type 1**. *AIDS research and human retroviruses* 2001, **17**(11):1009-1019.
144. Gouws E, White PJ, Stover J, Brown T: **Short term estimates of adult HIV incidence by mode of transmission: Kenya and Thailand as examples**. *Sexually transmitted infections* 2006, **82 Suppl 3**:iii51-55.
145. Galvin SR, Cohen MS: **The role of sexually transmitted diseases in HIV transmission**. *Nature reviews Microbiology* 2004, **2**(1):33-42.
146. Shaw GM, Hunter E: **HIV transmission**. *Cold Spring Harbor perspectives in medicine* 2012, **2**(11).
147. Saldanha VP, Buse K: **AIDS in eastern Europe and central Asia: time to face the facts**. *Lancet* 2016, **388**(10061):2737-2738.
148. Liu JF, Liu G, Li ZG: **Factors responsible for mother to child transmission (MTCT) of HIV-1 - a review**. *European review for medical and pharmacological sciences* 2017, **21**(4 Suppl):74-78.
149. Hurst SA, Appelgren KE, Kourtis AP: **Prevention of mother-to-child transmission of HIV type 1: the role of neonatal and infant prophylaxis**. *Expert review of anti-infective therapy* 2015, **13**(2):169-181.
150. Lynch NG, Johnson AK: **Congenital HIV: Prevention of Maternal to Child Transmission**. *Advances in neonatal care : official journal of the National Association of Neonatal Nurses* 2018, **18**(5):330-340.
151. Kourtis AP, Bulterys M: **Mother-to-child transmission of HIV: pathogenesis, mechanisms and pathways**. *Clinics in perinatology* 2010, **37**(4):721-737, vii.
152. Gumede-Moyo S, Filteau S, Munthali T, Todd J, Musonda P: **Implementation effectiveness of revised (post-2010) World Health Organization guidelines on prevention of mother-to-child transmission of HIV using routinely collected data in sub-Saharan Africa: A systematic literature review**. *Medicine* 2017, **96**(40):e8055.
153. Adetokunboh OO, Oluwasanu M: **Eliminating mother-to-child transmission of the human immunodeficiency virus in sub-Saharan Africa: The journey so far and what remains to be done**. *Journal of infection and public health* 2016, **9**(4):396-407.
154. Organization WH: **programmatic update: Use of AntiretrovirAl DrUgs for treAting PregnAnt Women AnD Preventing Hiv infection in infAnts**. In.; 2012.
155. Kelley CF, Kraft CS, de Man TJ, Duphare C, Lee HW, Yang J, Easley KA, Tharp GK, Mulligan MJ, Sullivan PS *et al*: **The rectal mucosa and condomless receptive anal intercourse in HIV-negative MSM: implications for HIV transmission and prevention**. *Mucosal immunology* 2017, **10**(4):996-1007.
156. Cavarelli M, Scarlatti G: **HIV-1 infection: the role of the gastrointestinal tract**. *American journal of reproductive immunology* 2014, **71**(6):537-542.
157. Varghese B, Maher JE, Peterman TA, Branson BM, Steketee RW: **Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use**. *Sexually transmitted diseases* 2002, **29**(1):38-43.
158. Page-Shafer K, Shiboski CH, Osmond DH, Dilley J, McFarland W, Shiboski SC, Klausner JD, Balls J, Greenspan D, Greenspan JS: **Risk of HIV infection attributable to oral sex among men who have sex with men and in the population of men who have sex with men**. *Aids* 2002, **16**(17):2350-2352.

159. Wilson TE, Massad LS, Riester KA, Barkan S, Richardson J, Young M, Gurtman A, Greenblatt R: **Sexual, contraceptive, and drug use behaviors of women with HIV and those at high risk for infection: results from the Women's Interagency HIV Study**. *Aids* 1999, **13**(5):591-598.
160. Royce RA, Sena A, Cates W, Jr., Cohen MS: **Sexual transmission of HIV**. *N Engl J Med* 1997, **336**(15):1072-1078.
161. Stieh DJ, Maric D, Kelley ZL, Anderson MR, Hattaway HZ, Beilfuss BA, Rothwangl KB, Veazey RS, Hope TJ: **Vaginal challenge with an SIV-based dual reporter system reveals that infection can occur throughout the upper and lower female reproductive tract**. *PLoS pathogens* 2014, **10**(10):e1004440.
162. Stieh DJ, Matias E, Xu H, Fought AJ, Blanchard JL, Marx PA, Veazey RS, Hope TJ: **Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques**. *Cell host & microbe* 2016, **19**(4):529-540.
163. Iwasaki A: **Antiviral immune responses in the genital tract: clues for vaccines**. *Nature reviews Immunology* 2010, **10**(10):699-711.
164. Carias AM, McCoombe S, McRaven M, Anderson M, Galloway N, Vandergrift N, Fought AJ, Lurain J, Duplantis M, Veazey RS *et al*: **Defining the interaction of HIV-1 with the mucosal barriers of the female reproductive tract**. *J Virol* 2013, **87**(21):11388-11400.
165. Mall AS, Habte H, Mthembu Y, Peacocke J, de Beer C: **Mucus and Mucins: do they have a role in the inhibition of the human immunodeficiency virus?** *Virology journal* 2017, **14**(1):192.
166. Hladik F, McElrath MJ: **Setting the stage: host invasion by HIV**. *Nature reviews Immunology* 2008, **8**(6):447-457.
167. Miller CJ, Shattock RJ: **Target cells in vaginal HIV transmission**. *Microbes and infection* 2003, **5**(1):59-67.
168. Witkin SS: **Immunology of the vagina**. *Clinical obstetrics and gynecology* 1993, **36**(1):122-128.
169. Norvell MK, Benrubi GI, Thompson RJ: **Investigation of microtrauma after sexual intercourse**. *The Journal of reproductive medicine* 1984, **29**(4):269-271.
170. Guimaraes MD, Vlahov D, Castilho EA: **Postcoital vaginal bleeding as a risk factor for transmission of the human immunodeficiency virus in a heterosexual partner study in Brazil. Rio de Janeiro Heterosexual Study Group**. *Archives of internal medicine* 1997, **157**(12):1362-1368.
171. Brown JE, Ayowa OB, Brown RC: **Dry and tight: sexual practices and potential AIDS risk in Zaire**. *Social science & medicine (1982)* 1993, **37**(8):989-994.
172. Civic D, Wilson D: **Dry sex in Zimbabwe and implications for condom use**. *Social science & medicine (1982)* 1996, **42**(1):91-98.
173. Kell PD, Barton SE, Edmonds DK, Boag FC: **HIV infection in a patient with Meyer-Rokitansky-Kuster-Hauser syndrome**. *Journal of the Royal Society of Medicine* 1992, **85**(11):706-707.
174. Miller CJ, Alexander NJ, Vogel P, Anderson J, Marx PA: **Mechanism of genital transmission of SIV: a hypothesis based on transmission studies and the location of SIV in the genital tract of chronically infected female rhesus macaques**. *Journal of medical primatology* 1992, **21**(2-3):64-68.
175. Pudney J, Quayle AJ, Anderson DJ: **Immunological microenvironments in the human vagina and cervix: mediators of cellular immunity are concentrated in the cervical transformation zone**. *Biology of reproduction* 2005, **73**(6):1253-1263.
176. Olmsted SS, Padgett JL, Yudin AI, Whaley KJ, Moench TR, Cone RA: **Diffusion of macromolecules and virus-like particles in human cervical mucus**. *Biophysical journal* 2001, **81**(4):1930-1937.
177. Shukair SA, Allen SA, Cianci GC, Stieh DJ, Anderson MR, Baig SM, Gioia CJ, Spongberg EJ, Kauffman SM, McRaven MD *et al*: **Human cervicovaginal mucus contains an activity that hinders HIV-1 movement**. *Mucosal immunology* 2013, **6**(2):427-434.

178. Dezzutti CS, Guenthner PC, Cummins JE, Jr., Cabrera T, Marshall JH, Dillberger A, Lal RB: **Cervical and prostate primary epithelial cells are not productively infected but sequester human immunodeficiency virus type 1**. *The Journal of infectious diseases* 2001, **183**(8):1204-1213.
179. Berlier W, Bourlet T, Lawrence P, Hamzeh H, Lambert C, Genin C, Verrier B, Dieu-Nosjean MC, Pozzetto B, Delezay O: **Selective sequestration of X4 isolates by human genital epithelial cells: Implication for virus tropism selection process during sexual transmission of HIV**. *Journal of medical virology* 2005, **77**(4):465-474.
180. Yeaman GR, Asin S, Weldon S, Demian DJ, Collins JE, Gonzalez JL, Wira CR, Fanger MW, Howell AL: **Chemokine receptor expression in the human ectocervix: implications for infection by the human immunodeficiency virus-type I**. *Immunology* 2004, **113**(4):524-533.
181. Kwong PD, Wyatt R, Sattentau QJ, Sodroski J, Hendrickson WA: **Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus**. *J Virol* 2000, **74**(4):1961-1972.
182. Moulard M, Lortat-Jacob H, Mondor I, Roca G, Wyatt R, Sodroski J, Zhao L, Olson W, Kwong PD, Sattentau QJ: **Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 gp120**. *J Virol* 2000, **74**(4):1948-1960.
183. Vives RR, Imberty A, Sattentau QJ, Lortat-Jacob H: **Heparan sulfate targets the HIV-1 envelope glycoprotein gp120 coreceptor binding site**. *The Journal of biological chemistry* 2005, **280**(22):21353-21357.
184. Saba E, Grivel JC, Vanpouille C, Brichacek B, Fitzgerald W, Margolis L, Lisco A: **HIV-1 sexual transmission: early events of HIV-1 infection of human cervico-vaginal tissue in an optimized ex vivo model**. *Mucosal immunology* 2010, **3**(3):280-290.
185. Patterson BK, Landay A, Andersson J, Brown C, Behbahani H, Jiyamapa D, Burki Z, Stanislawski D, Czerniewski MA, Garcia P: **Repertoire of chemokine receptor expression in the female genital tract: implications for human immunodeficiency virus transmission**. *The American journal of pathology* 1998, **153**(2):481-490.
186. Wira CR, Fahey JV: **A new strategy to understand how HIV infects women: identification of a window of vulnerability during the menstrual cycle**. *Aids* 2008, **22**(15):1909-1917.
187. Reichelderfer PS, Coombs RW, Wright DJ, Cohn J, Burns DN, Cu-Uvin S, Baron PA, Coheng MH, Landay AL, Beckner SK *et al*: **Effect of menstrual cycle on HIV-1 levels in the peripheral blood and genital tract. WHS 001 Study Team**. *Aids* 2000, **14**(14):2101-2107.
188. Coombs RW, Reichelderfer PS, Landay AL: **Recent observations on HIV type-1 infection in the genital tract of men and women**. *Aids* 2003, **17**(4):455-480.
189. Wolff H, Anderson DJ: **Potential human immunodeficiency virus-host cells in human semen**. *AIDS research and human retroviruses* 1988, **4**(1):1-2.
190. Wolff H: **The biologic significance of white blood cells in semen**. *Fertility and sterility* 1995, **63**(6):1143-1157.
191. Houzet L, Perez-Losada M, Matusali G, Deleage C, Dereuddre-Bosquet N, Satie AP, Aubry F, Becker E, Jegou B, Le Grand R *et al*: **Seminal Simian Immunodeficiency Virus in Chronically Infected Cynomolgus Macaques Is Dominated by Virus Originating from Multiple Genital Organs**. *J Virol* 2018, **92**(14).
192. Montagnon D, Clavert A, Cranz C: **Fructose, proteins and coagulation in human seminal plasma**. *Andrologia* 1982, **14**(5):434-439.
193. Ribeiro RM, Qin L, Chavez LL, Li D, Self SG, Perelson AS: **Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection**. *J Virol* 2010, **84**(12):6096-6102.
194. Hussain LA, Lehner T: **Comparative investigation of Langerhans' cells and potential receptors for HIV in oral, genitourinary and rectal epithelia**. *Immunology* 1995, **85**(3):475-484.

195. McCoombe SG, Short RV: **Potential HIV-1 target cells in the human penis**. *Aids* 2006, **20**(11):1491-1495.
196. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR *et al*: **DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells**. *Cell* 2000, **100**(5):587-597.
197. Quinn TC: **Circumcision and HIV transmission**. *Current opinion in infectious diseases* 2007, **20**(1):33-38.
198. Pudney J, Anderson DJ: **Immunobiology of the human penile urethra**. *The American journal of pathology* 1995, **147**(1):155-165.
199. Miller CJ, Alexander NJ, Sutjipto S, Lackner AA, Gettie A, Hendrickx AG, Lowenstine LJ, Jennings M, Marx PA: **Genital mucosal transmission of simian immunodeficiency virus: animal model for heterosexual transmission of human immunodeficiency virus**. *J Virol* 1989, **63**(10):4277-4284.
200. Burgener A, McGowan I, Klatt NR: **HIV and mucosal barrier interactions: consequences for transmission and pathogenesis**. *Current opinion in immunology* 2015, **36**:22-30.
201. Cicala C, Arthos J, Fauci AS: **HIV-1 envelope, integrins and co-receptor use in mucosal transmission of HIV**. *Journal of translational medicine* 2011, **9 Suppl 1**:S2.
202. Moutsopoulos NM, Greenwell-Wild T, Wahl SM: **Differential mucosal susceptibility in HIV-1 transmission and infection**. *Advances in dental research* 2006, **19**(1):52-56.
203. Neutra MR: **HIV transmission and immune protection at mucosal surfaces**. *Advances in experimental medicine and biology* 1998, **452**:169-175.
204. Russo CL, Spurr-Michaud S, Tisdale A, Pudney J, Anderson D, Gipson IK: **Mucin gene expression in human male urogenital tract epithelia**. *Human reproduction* 2006, **21**(11):2783-2793.
205. Olivier AJ, Masson L, Ronacher K, Walzl G, Coetzee D, Lewis DA, Williamson AL, Passmore JA, Burgers WA: **Distinct cytokine patterns in semen influence local HIV shedding and HIV target cell activation**. *The Journal of infectious diseases* 2014, **209**(8):1174-1184.
206. Olivier AJ, Liebenberg LJ, Coetzee D, Williamson AL, Passmore JA, Burgers WA: **Isolation and characterization of T cells from semen**. *J Immunol Methods* 2012, **375**(1-2):223-231.
207. Kaul R, Pettengell C, Sheth PM, Sunderji S, Biringer A, MacDonald K, Walmsley S, Rebbapragada A: **The genital tract immune milieu: an important determinant of HIV susceptibility and secondary transmission**. *Journal of reproductive immunology* 2008, **77**(1):32-40.
208. Ganor Y, Bomsel M: **HIV-1 transmission in the male genital tract**. *American journal of reproductive immunology* 2011, **65**(3):284-291.
209. Botting RA, Rana H, Bertram KM, Rhodes JW, Baharlou H, Nasr N, Cunningham AL, Harman AN: **Langerhans cells and sexual transmission of HIV and HSV**. *Rev Med Virol* 2017, **27**(2).
210. de Witte L, Nabatov A, Pion M, Fluitsma D, de Jong MA, de Gruijl T, Piguet V, van Kooyk Y, Geijtenbeek TB: **Langerin is a natural barrier to HIV-1 transmission by Langerhans cells**. *Nature medicine* 2007, **13**(3):367-371.
211. Hladik F, Sakchalathorn P, Ballweber L, Lentz G, Fialkow M, Eschenbach D, McElrath MJ: **Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1**. *Immunity* 2007, **26**(2):257-270.
212. Auvert B, Buonamico G, Lagarde E, Williams B: **Sexual behavior, heterosexual transmission, and the spread of HIV in sub-Saharan Africa: a simulation study**. *Computers and biomedical research, an international journal* 2000, **33**(1):84-96.
213. Gray R, Kigozi G, Kong X, Ssempiija V, Makumbi F, Wattya S, Serwadda D, Nalugoda F, Sewenkambo NK, Wawer MJ: **The effectiveness of male circumcision for HIV prevention and effects on risk behaviors in a posttrial follow-up study**. *Aids* 2012, **26**(5):609-615.

214. Fischetti L, Barry SM, Hope TJ, Shattock RJ: **HIV-1 infection of human penile explant tissue and protection by candidate microbicides**. *Aids* 2009, **23**(3):319-328.
215. Bomsel M: **Transcytosis of infectious human immunodeficiency virus across a tight human epithelial cell line barrier**. *Nature medicine* 1997, **3**(1):42-47.
216. Dosekun O, Fox J: **An overview of the relative risks of different sexual behaviours on HIV transmission**. *Current opinion in HIV and AIDS* 2010, **5**(4):291-297.
217. Sullivan PS, Salazar L, Buchbinder S, Sanchez TH: **Estimating the proportion of HIV transmissions from main sex partners among men who have sex with men in five US cities**. *Aids* 2009, **23**(9):1153-1162.
218. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH: **Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group**. *N Engl J Med* 2000, **342**(13):921-929.
219. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, Kiwanuka N, Kigozi G, Kiddugavu M, Lutalo T *et al*: **Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda**. *The Journal of infectious diseases* 2005, **191**(9):1403-1409.
220. Padian NS, McCoy SI, Balkus JE, Wasserheit JN: **Weighing the gold in the gold standard: challenges in HIV prevention research**. *Aids* 2010, **24**(5):621-635.
221. Winkelstein W, Jr., Samuel M, Padian NS, Wiley JA, Lang W, Anderson RE, Levy JA: **The San Francisco Men's Health Study: III. Reduction in human immunodeficiency virus transmission among homosexual/bisexual men, 1982-86**. *American journal of public health* 1987, **77**(6):685-689.
222. Bekker LG, Beyrer C, Quinn TC: **Behavioral and biomedical combination strategies for HIV prevention**. *Cold Spring Harbor perspectives in medicine* 2012, **2**(8).
223. Slutkin G, Okware S, Naamara W, Sutherland D, Flanagan D, Carael M, Blas E, Delay P, Tarantola D: **How Uganda reversed its HIV epidemic**. *AIDS and behavior* 2006, **10**(4):351-360.
224. Padian NS, Buve A, Balkus J, Serwadda D, Cates W, Jr.: **Biomedical interventions to prevent HIV infection: evidence, challenges, and way forward**. *Lancet* 2008, **372**(9638):585-599.
225. Bailey RC, Plummer FA, Moses S: **Male circumcision and HIV prevention: current knowledge and future research directions**. *The Lancet infectious diseases* 2001, **1**(4):223-231.
226. Templeton DJ: **Male circumcision to reduce sexual transmission of HIV**. *Current opinion in HIV and AIDS* 2010, **5**(4):344-349.
227. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, Williams CF, Campbell RT, Ndinya-Achola JO: **Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial**. *Lancet* 2007, **369**(9562):643-656.
228. Siegfried N, Muller M, Volmink J, Deeks J, Egger M, Low N, Weiss H, Walker S, Williamson P: **Male circumcision for prevention of heterosexual acquisition of HIV in men**. *Cochrane database of systematic reviews (Online)* 2003(3):CD003362.
229. Grulich AE, Hendry O, Clark E, Kippax S, Kaldor JM: **Circumcision and male-to-male sexual transmission of HIV**. *Aids* 2001, **15**(9):1188-1189.
230. Halperin DT, Bailey RC: **Male circumcision and HIV infection: 10 years and counting**. *Lancet* 1999, **354**(9192):1813-1815.
231. Weiss HA, Quigley MA, Hayes RJ: **Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis**. *Aids* 2000, **14**(15):2361-2370.
232. Patterson BK, Landay A, Siegel JN, Flener Z, Pessis D, Chaviano A, Bailey RC: **Susceptibility to human immunodeficiency virus-1 infection of human foreskin and cervical tissue grown in explant culture**. *The American journal of pathology* 2002, **161**(3):867-873.
233. Rodriguez-Diaz CE, Clatts MC, Jovet-Toledo GG, Vargas-Molina RL, Goldsamt LA, Garcia H: **More than foreskin: circumcision status, history of HIV/STI, and sexual risk in a clinic-based sample of men in Puerto Rico**. *The journal of sexual medicine* 2012, **9**(11):2933-2937.

234. Backes DM, Bleeker MC, Meijer CJ, Hudgens MG, Agot K, Bailey RC, Ndinya-Achola JO, Hayombe J, Hogewoning CJ, Moses S *et al*: **Male circumcision is associated with a lower prevalence of human papillomavirus-associated penile lesions among Kenyan men**. *International journal of cancer Journal international du cancer* 2012, **130**(8):1888-1897.
235. Passmore JA, Jaspan HB, Masson L: **Genital inflammation, immune activation and risk of sexual HIV acquisition**. *Current opinion in HIV and AIDS* 2016, **11**(2):156-162.
236. Rottingen JA, Cameron DW, Garnett GP: **A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known?** *Sexually transmitted diseases* 2001, **28**(10):579-597.
237. Sexton J, Garnett G, Rottingen JA: **Metaanalysis and metaregression in interpreting study variability in the impact of sexually transmitted diseases on susceptibility to HIV infection**. *Sexually transmitted diseases* 2005, **32**(6):351-357.
238. Quilter L, Dhanireddy S, Marrazzo J: **Prevention of Sexually Transmitted Diseases in HIV-Infected Individuals**. *Current HIV/AIDS reports* 2017, **14**(2):41-46.
239. Piot P, Laga M: **Genital ulcers, other sexually transmitted diseases, and the sexual transmission of HIV**. *Bmj* 1989, **298**(6674):623-624.
240. Gray RH, Wawer MJ: **Reassessing the hypothesis on STI control for HIV prevention**. *Lancet* 2008, **371**(9630):2064-2065.
241. Hayes R, Watson-Jones D, Celum C, van de Wijgert J, Wasserheit J: **Treatment of sexually transmitted infections for HIV prevention: end of the road or new beginning?** *Aids* 2010, **24 Suppl 4**:S15-26.
242. Wasserheit JN: **Epidemiological synergy. Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases**. *Sexually transmitted diseases* 1992, **19**(2):61-77.
243. Dickerson MC, Johnston J, Delea TE, White A, Andrews E: **The causal role for genital ulcer disease as a risk factor for transmission of human immunodeficiency virus. An application of the Bradford Hill criteria**. *Sexually transmitted diseases* 1996, **23**(5):429-440.
244. Gianella S, Morris SR, Vargas MV, Young JA, Callahan B, Richman DD, Little SJ, Smith DM: **Role of seminal shedding of herpesviruses in HIV Type 1 Transmission**. *The Journal of infectious diseases* 2013, **207**(2):257-261.
245. Johnson LF, Lewis DA: **The effect of genital tract infections on HIV-1 shedding in the genital tract: a systematic review and meta-analysis**. *Sexually transmitted diseases* 2008, **35**(11):946-959.
246. Anderson BL, Cu-Uvin S: **Determinants of HIV shedding in the lower genital tract of women**. *Current infectious disease reports* 2008, **10**(6):505-511.
247. Eron JJ, Jr., Gilliam B, Fiscus S, Dyer J, Cohen MS: **HIV-1 shedding and chlamydial urethritis**. *Jama* 1996, **275**(1):36.
248. Dyer JR, Eron JJ, Hoffman IF, Kazembe P, Vernazza PL, Nkata E, Costello Daly C, Fiscus SA, Cohen MS: **Association of CD4 cell depletion and elevated blood and seminal plasma human immunodeficiency virus type 1 (HIV-1) RNA concentrations with genital ulcer disease in HIV-1-infected men in Malawi**. *The Journal of infectious diseases* 1998, **177**(1):224-227.
249. Rotchford K, Strum AW, Wilkinson D: **Effect of coinfection with STDs and of STD treatment on HIV shedding in genital-tract secretions: systematic review and data synthesis**. *Sexually transmitted diseases* 2000, **27**(5):243-248.
250. Moss GB, Overbaugh J, Welch M, Reilly M, Bwayo J, Plummer FA, Ndinya-Achola JO, Malisa MA, Kreiss JK: **Human immunodeficiency virus DNA in urethral secretions in men: association with gonococcal urethritis and CD4 cell depletion**. *The Journal of infectious diseases* 1995, **172**(6):1469-1474.
251. Cohen MS, Hoffman IF, Royce RA, Kazembe P, Dyer JR, Daly CC, Zimba D, Vernazza PL, Maida M, Fiscus SA *et al*: **Reduction of concentration of HIV-1 in semen after treatment of**

urethritis: implications for prevention of sexual transmission of HIV-1. AIDSCAP Malawi Research Group. *Lancet* 1997, **349**(9069):1868-1873.
252. Cresswell F, Waters L, Briggs E, Fox J, Harbottle J, Hawkins D, Murchie M, Radcliffe K, Rafferty P, Rodger A *et al*: **UK guideline for the use of HIV Post-Exposure Prophylaxis Following Sexual Exposure, 2015**. *International journal of STD & AIDS* 2016, **27**(9):713-738.
253. Cardo DM, Culver DH, Ciesielski CA, Srivastava PU, Marcus R, Abiteboul D, Heptonstall J, Ippolito G, Lot F, McKibben PS *et al*: **A case-control study of HIV seroconversion in health care workers after percutaneous exposure. Centers for Disease Control and Prevention Needlestick Surveillance Group**. *N Engl J Med* 1997, **337**(21):1485-1490.
254. Garcia-Lerma JG, Cong ME, Mitchell J, Youngpairoj AS, Zheng Q, Masciotra S, Martin A, Kuklenyik Z, Holder A, Lipscomb J *et al*: **Intermittent prophylaxis with oral truvada protects macaques from rectal SHIV infection**. *Science translational medicine* 2010, **2**(14):14ra14.
255. Tsai CC, Emau P, Follis KE, Beck TW, Benveniste RE, Bischofberger N, Lifson JD, Morton WR: **Effectiveness of postinoculation (R)-9-(2-phosphonylmethoxypropyl) adenine treatment for prevention of persistent simian immunodeficiency virus SIVmne infection depends critically on timing of initiation and duration of treatment**. *J Virol* 1998, **72**(5):4265-4273.
256. Otten RA, Smith DK, Adams DR, Pullium JK, Jackson E, Kim CN, Jaffe H, Janssen R, Butera S, Folks TM: **Efficacy of postexposure prophylaxis after intravaginal exposure of pig-tailed macaques to a human-derived retrovirus (human immunodeficiency virus type 2)**. *J Virol* 2000, **74**(20):9771-9775.
257. Wade NA, Birkhead GS, Warren BL, Charbonneau TT, French PT, Wang L, Baum JB, Tesoriero JM, Savicki R: **Abbreviated regimens of zidovudine prophylaxis and perinatal transmission of the human immunodeficiency virus**. *N Engl J Med* 1998, **339**(20):1409-1414.
258. Sperling RS, Shapiro DE, Coombs RW, Todd JA, Herman SA, McSherry GD, O'Sullivan MJ, Van Dyke RB, Jimenez E, Rouzioux C *et al*: **Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group**. *N Engl J Med* 1996, **335**(22):1621-1629.
259. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, Sullivan AK, Clarke A, Reeves I, Schembri G *et al*: **Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial**. *Lancet* 2016, **387**(10013):53-60.
260. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, Goicochea P, Casapia M, Guanira-Carranza JV, Ramirez-Cardich ME *et al*: **Preexposure chemoprophylaxis for HIV prevention in men who have sex with men**. *N Engl J Med* 2010, **363**(27):2587-2599.
261. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, Tremblay C, Le Gall JM, Cua E, Pasquet A *et al*: **On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection**. *N Engl J Med* 2015, **373**(23):2237-2246.
262. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, Tappero JW, Bukusi EA, Cohen CR, Katabira E *et al*: **Antiretroviral prophylaxis for HIV prevention in heterosexual men and women**. *N Engl J Med* 2012, **367**(5):399-410.
263. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, Henderson FL, Pathak SR, Soud FA, Chillag KL *et al*: **Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana**. *N Engl J Med* 2012, **367**(5):423-434.
264. Marrazzo JM, Ramjee G, Richardson BA, Gomez K, Mgodi N, Nair G, Palanee T, Nakabiito C, van der Straten A, Noguchi L *et al*: **Tenofovir-based preexposure prophylaxis for HIV infection among African women**. *N Engl J Med* 2015, **372**(6):509-518.
265. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, Malahleha M, Owino F, Manongi R, Onyango J *et al*: **Preexposure prophylaxis for HIV infection among African women**. *N Engl J Med* 2012, **367**(5):411-422.

266. Choopanya K, Martin M, Suntharasamai P, Sangkum U, Mock PA, Leethochawalit M, Chiamwongpaet S, Kitisin P, Natrujirote P, Kittimunkong S *et al*: **Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial**. *Lancet* 2013, **381**(9883):2083-2090.
267. Reynolds MR, Weiler AM, Weisgrau KL, Piaskowski SM, Furlott JR, Weinfurter JT, Kaizu M, Soma T, Leon EJ, MacNair C *et al*: **Macaques vaccinated with live-attenuated SIV control replication of heterologous virus**. *The Journal of experimental medicine* 2008, **205**(11):2537-2550.
268. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premsri N, Namwat C, de Souza M, Adams E *et al*: **Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand**. *N Engl J Med* 2009, **361**(23):2209-2220.
269. Barouch DH, Tomaka FL, Wegmann F, Stieh DJ, Alter G, Robb ML, Michael NL, Peter L, Nkolola JP, Borducchi EN *et al*: **Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19)**. *Lancet* 2018, **392**(10143):232-243.
270. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JH *et al*: **Prevention of HIV-1 infection with early antiretroviral therapy**. *N Engl J Med* 2011, **365**(6):493-505.
271. Attia S, Egger M, Muller M, Zwahlen M, Low N: **Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis**. *Aids* 2009, **23**(11):1397-1404.
272. Cohen MS: **HIV treatment as prevention and "the Swiss statement": in for a dime, in for a dollar?** *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2010, **51**(11):1323-1324.
273. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG: **Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model**. *Lancet* 2009, **373**(9657):48-57.
274. Bacaer N, Pretorius C, Auvert B: **An age-structured model for the potential impact of generalized access to antiretrovirals on the South African HIV epidemic**. *Bulletin of mathematical biology* 2010, **72**(8):2180-2198.
275. Organization WH: **Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection**. In.; 2015.
276. Gangakhedkar RR: **Ending AIDS: The 90-90-90 strategy**. *Medical journal, Armed Forces India* 2017, **73**(1):3-4.
277. Blackard JT: **HIV compartmentalization: a review on a clinically important phenomenon**. *Current HIV research* 2012, **10**(2):133-142.
278. Smyth RP, Negroni M: **A step forward understanding HIV-1 diversity**. *Retrovirology* 2016, **13**:27.
279. Lauring AS, Andino R: **Quasispecies theory and the behavior of RNA viruses**. *PLoS pathogens* 2010, **6**(7):e1001005.
280. Mansky LM, Temin HM: **Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase**. *J Virol* 1995, **69**(8):5087-5094.
281. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH: **Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication**. *J Virol* 2010, **84**(19):9864-9878.
282. Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R: **Extremely High Mutation Rate of HIV-1 In Vivo**. *PLoS biology* 2015, **13**(9):e1002251.
283. Hemelaar J: **The origin and diversity of the HIV-1 pandemic**. *Trends in molecular medicine* 2012, **18**(3):182-192.

284. Quinones-Mateu ME, Arts EJ: **Virus fitness: concept, quantification, and application to HIV population dynamics**. *Current topics in microbiology and immunology* 2006, **299**:83-140.
285. Heath L, Fox A, McClure J, Diem K, van 't Wout AB, Zhao H, Park DR, Schouten JT, Twigg HL, Corey L *et al*: **Evidence for limited genetic compartmentalization of HIV-1 between lung and blood**. *PloS one* 2009, **4**(9):e6949.
286. Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, Tapia K, Holte S, Dragavon J, Coombs R *et al*: **Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations**. *PloS one* 2009, **4**(9):e7122.
287. Bull ME, Heath LM, McKernan-Mullin JL, Kraft KM, Acevedo L, Hitti JE, Cohn SE, Tapia KA, Holte SE, Dragavon JA *et al*: **Human immunodeficiency viruses appear compartmentalized to the female genital tract in cross-sectional analyses but genital lineages do not persist over time**. *The Journal of infectious diseases* 2013, **207**(8):1206-1215.
288. Chaudhary S, Noel RJ, Rodriguez N, Collado S, Munoz J, Kumar A, Yamamura Y: **Correlation between CD4 T cell counts and virus compartmentalization in genital and systemic compartments of HIV-infected females**. *Virology* 2011, **417**(2):320-326.
289. Diem K, Nickle DC, Motoshige A, Fox A, Ross S, Mullins JI, Corey L, Coombs RW, Krieger JN: **Male genital tract compartmentalization of human immunodeficiency virus type 1 (HIV)**. *AIDS research and human retroviruses* 2008, **24**(4):561-571.
290. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C, Philpott SM, Gao W, Robison E, Holman S *et al*: **HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(22):12972-12977.
291. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C, Weiser B: **Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination**. *J Virol* 2005, **79**(1):353-363.
292. Zhu T, Wang N, Carr A, Nam DS, Moor-Jankowski R, Cooper DA, Ho DD: **Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission**. *J Virol* 1996, **70**(5):3098-3107.
293. van Zyl G, Bale MJ, Kearney MF: **HIV evolution and diversity in ART-treated patients**. *Retrovirology* 2018, **15**(1):14.
294. Nickle DC, Shriner D, Mittler JE, Frenkel LM, Mullins JI: **Importance and detection of virus reservoirs and compartments of HIV infection**. *Current opinion in microbiology* 2003, **6**(4):410-416.
295. Simmonds P, Zhang LQ, McOmish F, Balfe P, Ludlam CA, Brown AJ: **Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis**. *J Virol* 1991, **65**(11):6266-6276.
296. Potter SJ, Dwyer DE, Saksena NK: **Differential cellular distribution of HIV-1 drug resistance in vivo: evidence for infection of CD8+ T cells during HAART**. *Virology* 2003, **305**(2):339-352.
297. Sarmati L, Nicastri E, Uccella I, D'Ettorre G, Parisi SG, Palmisano L, Galluzzo C, Concia E, Vullo V, Vella S *et al*: **Drug-associated resistance mutations in plasma and peripheral blood mononuclear cells of human immunodeficiency virus type 1-infected patients for whom highly active antiretroviral therapy is failing**. *Journal of clinical microbiology* 2003, **41**(4):1760-1762.
298. Fulcher JA, Hwangbo Y, Zioni R, Nickle D, Lin X, Heath L, Mullins JI, Corey L, Zhu T: **Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection**. *J Virol* 2004, **78**(15):7883-7893.
299. Ince WL, Harrington PR, Schnell GL, Patel-Chhabra M, Burch CL, Menezes P, Price RW, Eron JJ, Jr., Swanstrom RI: **Major coexisting human immunodeficiency virus type 1 env gene**

subpopulations in the peripheral blood are produced by cells with similar turnover rates and show little evidence of genetic compartmentalization. *J Virol* 2009, **83**(9):4068-4080.
300. Devereux HL, Loveday C, Youle M, Sabin CA, Burke A, Johnson M: **Substantial correlation between HIV type 1 drug-associated resistance mutations in plasma and peripheral blood mononuclear cells in treatment-experienced patients**. *AIDS research and human retroviruses* 2000, **16**(11):1025-1030.
301. Almeida SM: **Cerebrospinal fluid analysis in the HIV infection and compartmentalization of HIV in the central nervous system**. *Arquivos de neuro-psiquiatria* 2015, **73**(7):624-629.
302. Bednar MM, Sturdevant CB, Tompkins LA, Arrildt KT, Dukhovlinova E, Kincer LP, Swanstrom R: **Compartmentalization, Viral Evolution, and Viral Latency of HIV in the CNS**. *Current HIV/AIDS reports* 2015, **12**(2):262-271.
303. Chaillon A, Gianella S, Wertheim JO, Richman DD, Mehta SR, Smith DM: **HIV migration between blood and cerebrospinal fluid or semen over time**. *The Journal of infectious diseases* 2014, **209**(10):1642-1652.
304. de Almeida SM, Rotta I, Ribeiro CE, Smith D, Wang R, Judicello J, Potter M, Vaida F, Letendre S, Ellis RJ: **Blood-CSF barrier and compartmentalization of CNS cellular immune response in HIV infection**. *Journal of neuroimmunology* 2016, **301**:41-48.
305. Schnell G, Price RW, Swanstrom R, Spudich S: **Compartmentalization and clonal amplification of HIV-1 variants in the cerebrospinal fluid during primary infection**. *J Virol* 2010, **84**(5):2395-2407.
306. Anesten B, Yilmaz A, Hagberg L, Zetterberg H, Nilsson S, Brew BJ, Fuchs D, Price RW, Gisslen M: **Blood-brain barrier integrity, intrathecal immunoactivation, and neuronal injury in HIV**. *Neurology(R) neuroimmunology & neuroinflammation* 2016, **3**(6):e300.
307. Ballabh P, Braun A, Nedergaard M: **The blood-brain barrier: an overview: structure, regulation, and clinical implications**. *Neurobiology of disease* 2004, **16**(1):1-13.
308. Zhang YL, Ouyang YB, Liu LG, Chen DX: **Blood-brain barrier and neuro-AIDS**. *European review for medical and pharmacological sciences* 2015, **19**(24):4927-4939.
309. Eggers CC, van Lunzen J, Buhk T, Stellbrink HJ: **HIV infection of the central nervous system is characterized by rapid turnover of viral RNA in cerebrospinal fluid**. *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association* 1999, **20**(3):259-264.
310. Banks WA: **The blood-brain barrier in neuroimmunology: Tales of separation and assimilation**. *Brain, behavior, and immunity* 2015, **44**:1-8.
311. Joseph J, Colosi DA, Rao VR: **HIV-1 Induced CNS Dysfunction: Current Overview and Research Priorities**. *Current HIV research* 2016, **14**(5):389-399.
312. Spudich SS, Ances BM: **CROI 2017: Neurologic Complications of HIV Infection**. *Topics in antiviral medicine* 2017, **25**(2):69-76.
313. Zayyad Z, Spudich S: **Neuropathogenesis of HIV: from initial neuroinvasion to HIV-associated neurocognitive disorder (HAND)**. *Current HIV/AIDS reports* 2015, **12**(1):16-24.
314. Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, Clifford DB, Cinque P, Epstein LG, Goodkin K *et al*: **Updated research nosology for HIV-associated neurocognitive disorders**. *Neurology* 2007, **69**(18):1789-1799.
315. Resnick L, Berger JR, Shapshak P, Tourtellotte WW: **Early penetration of the blood-brain-barrier by HIV**. *Neurology* 1988, **38**(1):9-14.
316. Valcour V, Chalermchai T, Sailasuta N, Marovich M, Lerdlum S, Suttichom D, Suwanwela NC, Jagodzinski L, Michael N, Spudich S *et al*: **Central nervous system viral invasion and inflammation during acute HIV infection**. *The Journal of infectious diseases* 2012, **206**(2):275-282.
317. Spudich S, Gisslen M, Hagberg L, Lee E, Liegler T, Brew B, Fuchs D, Tambussi G, Cinque P, Hecht FM *et al*: **Central nervous system immune activation characterizes primary human**

immunodeficiency virus 1 infection even in participants with minimal cerebrospinal fluid viral burden. *The Journal of infectious diseases* 2011, **204**(5):753-760.
318. Brew BJ: **Evidence for a change in AIDS dementia complex in the era of highly active antiretroviral therapy and the possibility of new forms of AIDS dementia complex**. *Aids* 2004, **18 Suppl 1**:S75-78.
319. Ohagen A, Devitt A, Kunstman KJ, Gorry PR, Rose PP, Korber B, Taylor J, Levy R, Murphy RL, Wolinsky SM *et al*: **Genetic and functional analysis of full-length human immunodeficiency virus type 1 env genes derived from brain and blood of patients with AIDS**. *J Virol* 2003, **77**(22):12336-12345.
320. Power C, McArthur JC, Johnson RT, Griffin DE, Glass JD, Perryman S, Chesebro B: **Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences**. *J Virol* 1994, **68**(7):4643-4649.
321. Tang YW, Huong JT, Lloyd RM, Jr., Spearman P, Haas DW: **Comparison of human immunodeficiency virus type 1 RNA sequence heterogeneity in cerebrospinal fluid and plasma**. *Journal of clinical microbiology* 2000, **38**(12):4637-4639.
322. Strain MC, Letendre S, Pillai SK, Russell T, Ignacio CC, Gunthard HF, Good B, Smith DM, Wolinsky SM, Furtado M *et al*: **Genetic composition of human immunodeficiency virus type 1 in cerebrospinal fluid and blood without treatment and during failing antiretroviral therapy**. *J Virol* 2005, **79**(3):1772-1788.
323. Pillai SK, Pond SL, Liu Y, Good BM, Strain MC, Ellis RJ, Letendre S, Smith DM, Gunthard HF, Grant I *et al*: **Genetic attributes of cerebrospinal fluid-derived HIV-1 env**. *Brain : a journal of neurology* 2006, **129**(Pt 7):1872-1883.
324. Dahiya S, Irish BP, Nonnemacher MR, Wigdahl B: **Genetic variation and HIV-associated neurologic disease**. *Advances in virus research* 2013, **87**:183-240.
325. Stefic K, Chaillon A, Bouvin-Pley M, Moreau A, Braibant M, Bastides F, Gras G, Bernard L, Barin F: **Probing the compartmentalization of HIV-1 in the central nervous system through its neutralization properties**. *PloS one* 2017, **12**(8):e0181680.
326. Sturdevant CB, Joseph SB, Schnell G, Price RW, Swanstrom R, Spudich S: **Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection**. *PLoS pathogens* 2015, **11**(3):e1004720.
327. Schnell G, Joseph S, Spudich S, Price RW, Swanstrom R: **HIV-1 replication in the central nervous system occurs in two distinct cell types**. *PLoS pathogens* 2011, **7**(10):e1002286.
328. Joseph SB, Arrildt KT, Swanstrom AE, Schnell G, Lee B, Hoxie JA, Swanstrom R: **Quantification of entry phenotypes of macrophage-tropic HIV-1 across a wide range of CD4 densities**. *J Virol* 2014, **88**(4):1858-1869.
329. Mestecky J, Moldoveanu Z, Smith PD, Hel Z, Alexander RC: **Mucosal immunology of the genital and gastrointestinal tracts and HIV-1 infection**. *Journal of reproductive immunology* 2009, **83**(1-2):196-200.
330. Mehandru S, Poles MA, Tenner-Racz K, Manuelli V, Jean-Pierre P, Lopez P, Shet A, Low A, Mohri H, Boden D *et al*: **Mechanisms of gastrointestinal CD4+ T-cell depletion during acute and early human immunodeficiency virus type 1 infection**. *J Virol* 2007, **81**(2):599-612.
331. Hel Z, McGhee JR, Mestecky J: **HIV infection: first battle decides the war**. *Trends in immunology* 2006, **27**(6):274-281.
332. McGowan I, Elliott J, Fuerst M, Taing P, Boscardin J, Poles M, Anton P: **Increased HIV-1 mucosal replication is associated with generalized mucosal cytokine activation**. *Journal of acquired immune deficiency syndromes* 2004, **37**(2):1228-1236.
333. van Marle G, Gill MJ, Kolodka D, McManus L, Grant T, Church DL: **Compartmentalization of the gut viral reservoir in HIV-1 infected patients**. *Retrovirology* 2007, **4**:87.
334. Lamers SL, Salemi M, Galligan DC, de Oliveira T, Fogel GB, Granier SC, Zhao L, Brown JN, Morris A, Masliah E *et al*: **Extensive HIV-1 intra-host recombination is common in tissues with abnormal histopathology**. *PloS one* 2009, **4**(3):e5065.

335. Salemi M, Lamers SL, Huysentruyt LC, Galligan D, Gray RR, Morris A, McGrath MS: **Distinct patterns of HIV-1 evolution within metastatic tissues in patients with non-Hodgkins lymphoma**. *PloS one* 2009, **4**(12):e8153.
336. van der Hoek L, Sol CJ, Maas J, Lukashov VV, Kuiken CL, Goudsmit J: **Genetic differences between human immunodeficiency virus type 1 subpopulations in faeces and serum**. *The Journal of general virology* 1998, **79 (Pt 2)**:259-267.
337. van der Hoek L, Goudsmit J, Maas J, Sol CJ: **Human immunodeficiency virus type 1 in faeces and serum: evidence against independently evolving subpopulations**. *The Journal of general virology* 1998, **79 (Pt 10)**:2455-2459.
338. van der Hoek L, Sol CJ, Snijders F, Bartelsman JF, Boom R, Goudsmit J: **Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations**. *The Journal of general virology* 1996, **77 (Pt 10)**:2415-2425.
339. Gantt S, Carlsson J, Heath L, Bull M, Shetty A, Mutsvangwa J, Musingwini G, Woelk G, Zijenah L, Katzenstein D: **Genetic analyses of HIV-1 env sequences demonstrate limited compartmentalization in breast milk and suggest viral replication within the breast that increases with mastitis**. *J Virol* 2010, **84**:10812 - 10819.
340. Becquart P, Chomont N, Roques P, Ayouba A, Kazatchkine MD, Belec L, Hocini H: **Compartmentalization of HIV-1 between breast milk and blood of HIV-infected mothers**. *Virology* 2002, **300**(1):109-117.
341. Gantt S, Carlsson J, Heath L, Bull ME, Shetty AK, Mutsvangwa J, Musingwini G, Woelk G, Zijenah LS, Katzenstein DA *et al*: **Genetic analyses of HIV-1 env sequences demonstrate limited compartmentalization in breast milk and suggest viral replication within the breast that increases with mastitis**. *J Virol* 2010, **84**(20):10812-10819.
342. Salazar-Gonzalez JF, Salazar MG, Learn GH, Fouda GG, Kang HH, Mahlokozera T, Wilks AB, Lovingood RV, Stacey A, Kalilani L *et al*: **Origin and evolution of HIV-1 in breast milk determined by single-genome amplification and sequencing**. *J Virol* 2011, **85**(6):2751-2763.
343. Gray RR, Salemi M, Lowe A, Nakamura KJ, Decker WD, Sinkala M, Kankasa C, Mulligan CJ, Thea DM, Kuhn L *et al*: **Multiple independent lineages of HIV-1 persist in breast milk and plasma**. *Aids* 2011, **25**(2):143-152.
344. Penton PK, Blackard JT: **Analysis of HIV quasispecies suggests compartmentalization in the liver**. *AIDS research and human retroviruses* 2014, **30**(4):394-402.
345. Blackard JT, Ma G, Martin CM, Rouster SD, Shata MT, Sherman KE: **HIV variability in the liver and evidence of possible compartmentalization**. *AIDS research and human retroviruses* 2011, **27**(10):1117-1126.
346. Blackard JT, Sherman KE: **HCV/ HIV co-infection: time to re-evaluate the role of HIV in the liver?** *Journal of viral hepatitis* 2008, **15**(5):323-330.
347. Collins KR, Mayanja-Kizza H, Sullivan BA, Quinones-Mateu ME, Toossi Z, Arts EJ: **Greater diversity of HIV-1 quasispecies in HIV-infected individuals with active tuberculosis**. *Journal of acquired immune deficiency syndromes* 2000, **24**(5):408-417.
348. Nakata K, Weiden M, Harkin T, Ho D, Rom WN: **Low copy number and limited variability of proviral DNA in alveolar macrophages from HIV-1-infected patients: evidence for genetic differences in HIV-1 between lung and blood macrophage populations**. *Molecular medicine (Cambridge, Mass)* 1995, **1**(7):744-757.
349. Itescu S, Simonelli PF, Winchester RJ, Ginsberg HS: **Human immunodeficiency virus type 1 strains in the lungs of infected individuals evolve independently from those in peripheral blood and are highly conserved in the C-terminal region of the envelope V3 loop**. *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**(24):11378-11382.
350. Fierer DS, Klotman ME: **Kidney and central nervous system as reservoirs of HIV infection**. *Current opinion in HIV and AIDS* 2006, **1**(2):115-120.

351. Marras D, Bruggeman LA, Gao F, Tanji N, Mansukhani MM, Cara A, Ross MD, Gusella GL, Benson G, D'Agati VD *et al*: **Replication and compartmentalization of HIV-1 in kidney epithelium of patients with HIV-associated nephropathy**. *Nature medicine* 2002, **8**(5):522-526.
352. Blasi M, Carpenter JH, Balakumaran B, Cara A, Gao F, Klotman ME: **Identification of HIV-1 genitourinary tract compartmentalization by analyzing the env gene sequences in urine**. *Aids* 2015, **29**(13):1651-1657.
353. Haddad DN, Birch C, Middleton T, Dwyer DE, Cunningham AL, Saksena NK: **Evidence for late stage compartmentalization of HIV-1 resistance mutations between lymph node and peripheral blood mononuclear cells**. *Aids* 2000, **14**(15):2273-2281.
354. Haggerty S, Stevenson M: **Predominance of distinct viral genotypes in brain and lymph node compartments of HIV-1-infected individuals**. *Viral Immunol* 1991, **4**(2):123-131.
355. Reddy RT, Achim CL, Sirko DA, Tehranchi S, Kraus FG, Wong-Staal F, Wiley CA: **Sequence analysis of the V3 loop in brain and spleen of patients with HIV encephalitis**. *AIDS research and human retroviruses* 1996, **12**(6):477-482.
356. Delassus S, Cheynier R, Wain-Hobson S: **Nonhomogeneous distribution of human immunodeficiency virus type 1 proviruses in the spleen**. *J Virol* 1992, **66**(9):5642-5645.
357. Epstein LG, Kuiken C, Blumberg BM, Hartman S, Sharer LR, Clement M, Goudsmit J: **HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies**. *Virology* 1991, **180**(2):583-590.
358. Gratton S, Cheynier R, Dumaurier MJ, Oksenhendler E, Wain-Hobson S: **Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(26):14566-14571.
359. Baeten JM, Kahle E, Lingappa JR, Coombs RW, Delany-Moretlwe S, Nakku-Joloba E, Mugo NR, Wald A, Corey L, Donnell D *et al*: **Genital HIV-1 RNA predicts risk of heterosexual HIV-1 transmission**. *Science translational medicine* 2011, **3**(77):77ra29.
360. Rasheed S: **Infectivity and dynamics of HIV type 1 replication in the blood and reproductive tract of HIV type 1-infected women**. *AIDS research and human retroviruses* 1998, **14 Suppl 1**:S105-118.
361. Sullivan ST, Mandava U, Evans-Strickfaden T, Lennox JL, Ellerbrock TV, Hart CE: **Diversity, divergence, and evolution of cell-free human immunodeficiency virus type 1 in vaginal secretions and blood of chronically infected women: associations with immune status**. *J Virol* 2005, **79**(15):9799-9809.
362. Kovacs A, Wasserman SS, Burns D, Wright DJ, Cohn J, Landay A, Weber K, Cohen M, Levine A, Minkoff H *et al*: **Determinants of HIV-1 shedding in the genital tract of women**. *Lancet* 2001, **358**(9293):1593-1601.
363. Poss M, Rodrigo AG, Gosink JJ, Learn GH, de Vange Panteleeff D, Martin HL, Jr., Bwayo J, Kreiss JK, Overbaugh J: **Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1**. *J Virol* 1998, **72**(10):8240-8251.
364. Haaland RE, Sullivan ST, Evans-Strickfaden T, Lennox JL, Hart CE: **Female genital tract shedding of CXCR4-tropic HIV Type 1 is associated with a majority population of CXCR4-tropic HIV Type 1 in blood and declining CD4(+) cell counts**. *AIDS research and human retroviruses* 2012, **28**(11):1524-1532.
365. Andreoletti L, Skrabal K, Perrin V, Chomont N, Saragosti S, Gresenguet G, Moret H, Jacques J, Longo Jde D, Matta M *et al*: **Genetic and phenotypic features of blood and genital viral populations of clinically asymptomatic and antiretroviral-treatment-naive clade a human immunodeficiency virus type 1-infected women**. *Journal of clinical microbiology* 2007, **45**(6):1838-1842.

366. Ellerbrock TV, Lennox JL, Clancy KA, Schinazi RF, Wright TC, Pratt-Palmore M, Evans-Strickfaden T, Schnell C, Pai R, Conley LJ *et al*: **Cellular replication of human immunodeficiency virus type 1 occurs in vaginal secretions**. *The Journal of infectious diseases* 2001, **184**(1):28-36.
367. Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, Nyange P, Mandaliya K, Overbaugh J: **Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1**. *J Virol* 1995, **69**(12):8118-8122.
368. Adal M, Ayele W, Wolday D, Dagne K, Messele T, Tilahun T, Berkhout B, Mayaan S, Pollakis G, Dorigo-Zetsma W: **Evidence of genetic variability of human immunodeficiency virus type 1 in plasma and cervicovaginal lavage in ethiopian women seeking care for sexually transmitted infections**. *AIDS research and human retroviruses* 2005, **21**(7):649-653.
369. Bull ME, Learn GH, McElhone S, Hitti J, Lockhart D, Holte S, Dragavon J, Coombs RW, Mullins JI, Frenkel LM: **Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus**. *J Virol* 2009, **83**(12):6020-6028.
370. Zhang L, Rowe L, He T, Chung C, Yu J, Yu W, Talal A, Markowitz M, Ho DD: **Compartmentalization of surface envelope glycoprotein of human immunodeficiency virus type 1 during acute and chronic infection**. *J Virol* 2002, **76**(18):9465-9473.
371. Chakraborty H, Sen PK, Helms RW, Vernazza PL, Fiscus SA, Eron JJ, Patterson BK, Coombs RW, Krieger JN, Cohen MS: **Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: a probabilistic empiric model**. *Aids* 2001, **15**(5):621-627.
372. Tindall B, Evans L, Cunningham P, McQueen P, Hurren L, Vasak E, Mooney J, Cooper DA: **Identification of HIV-1 in semen following primary HIV-1 infection**. *Aids* 1992, **6**(9):949-952.
373. Kalichman SC, Di Berto G, Eaton L: **Human immunodeficiency virus viral load in blood plasma and semen: review and implications of empirical findings**. *Sexually transmitted diseases* 2008, **35**(1):55-60.
374. Kalichman SC, Rompa D, Cage M, Austin J, Luke W, Barnett T, Tharnish P, Mowrey J, Schinazi RF: **Sexual transmission risk perceptions and behavioural correlates of HIV concentrations in semen**. *AIDS care* 2002, **14**(3):343-349.
375. Lambert-Niclot S, Tubiana R, Beaudoux C, Lefebvre G, Caby F, Bonmarchand M, Naouri M, Schubert B, Dommergues M, Calvez V *et al*: **Detection of HIV-1 RNA in seminal plasma samples from treated patients with undetectable HIV-1 RNA in blood plasma on a 2002-2011 survey**. *Aids* 2012, **26**(8):971-975.
376. Coombs Robert W, Speck Carl E, Hughes James P, Lee W, Sampoleo R, Ross Susan O, Dragavon J, Peterson G, Hooton Thomas M, Collier Ann C *et al*: **Association between Culturable Human Immunodeficiency Virus Type 1 (HIV-1) in Semen and HIV-1 RNA Levels in Semen and Blood: Evidence for Compartmentalization of HIV-1 between Semen and Blood**. *The Journal of infectious diseases* 1998, **177**(2):320-330.
377. Gupta P, Leroux C, Patterson BK, Kingsley L, Rinaldo C, Ding M, Chen Y, Kulka K, Buchanan W, McKeon B *et al*: **Human immunodeficiency virus type 1 shedding pattern in semen correlates with the compartmentalization of viral Quasi species between blood and semen**. *The Journal of infectious diseases* 2000, **182**(1):79-87.
378. Kiessling AA, Fitzgerald LM, Zhang D, Chhay H, Brettler D, Eyre RC, Steinberg J, McGowan K, Byrn RA: **Human immunodeficiency virus in semen arises from a genetically distinct virus reservoir**. *AIDS research and human retroviruses* 1998, **14 Suppl 1**:S33-41.
379. Overbaugh J, Anderson RJ, Ndinya-Achola JO, Kreiss JK: **Distinct but related human immunodeficiency virus type 1 variant populations in genital secretions and blood**. *AIDS research and human retroviruses* 1996, **12**(2):107-115.
380. Paranjpe S, Craigo J, Patterson B, Ding M, Barroso P, Harrison L, Montelaro R, Gupta P: **Subcompartmentalization of HIV-1 quasispecies between seminal cells and seminal plasma**

indicates their origin in distinct genital tissues. *AIDS research and human retroviruses* 2002, **18**(17):1271-1280.
381. Anderson JA, Ping LH, Dibben O, Jabara CB, Arney L, Kincer L, Tang Y, Hobbs M, Hoffman I, Kazembe P *et al*: **HIV-1 Populations in Semen Arise through Multiple Mechanisms**. *PLoS pathogens* 2010, **6**(8):e1001053.
382. Ball JK, Curran R, Irving WL, Dearden AA: **HIV-1 in semen: determination of proviral and viral titres compared to blood, and quantification of semen leukocyte populations**. *Journal of medical virology* 1999, **59**(3):356-363.
383. Bourlet T, Cazorla C, Berthelot P, Grattard F, Cognasse F, Fresard A, Defontaine C, Lucht FR, Genin C, Pozzetto B: **Compartmentalization of HIV-1 according to antiretroviral therapy: viral loads are correlated in blood and semen but poorly in blood and saliva**. *Aids* 2001, **15**(2):284-285.
384. Bujan L, Daudin M, Alvarez M, Massip P, Puel J, Pasquier C: **Intermittent human immunodeficiency type 1 virus (HIV-1) shedding in semen and efficiency of sperm processing despite high seminal HIV-1 RNA levels**. *Fertility and sterility* 2002, **78**(6):1321-1323.
385. Byrn RA, Kiessling AA: **Analysis of human immunodeficiency virus in semen: indications of a genetically distinct virus reservoir**. *Journal of reproductive immunology* 1998, **41**(1-2):161-176.
386. Byrn RA, Zhang D, Eyre R, McGowan K, Kiessling AA: **HIV-1 in semen: an isolated virus reservoir**. *Lancet* 1997, **350**(9085):1141.
387. Chaix ML, Boufassa F, Meyzer C, Leruez-Ville M, Mahjoub N, Nere ML, Genet P, Duvivier C, Lascoux-Combes C, Lambotte O *et al*: **Detectable HIV-RNA in semen of HIV controllers**. *PloS one* 2017, **12**(8):e0183376.
388. Chan DJ: **Pathophysiology of HIV-1 in semen: current evidence for compartmentalisation and penetration by antiretroviral drugs**. *Current HIV research* 2005, **3**(3):207-222.
389. Chan DJ, Ray JE, McNally L, Batterham M, Smith DE: **Correlation between HIV-1 RNA load in blood and seminal plasma depending on antiretroviral treatment status, regimen and penetration of semen by antiretroviral drugs**. *Current HIV research* 2008, **6**(5):477-484.
390. Choudhury B, Pillay D, Taylor S, Cane PA: **Analysis of HIV-1 variation in blood and semen during treatment and treatment interruption**. *Journal of medical virology* 2002, **68**(4):467-472.
391. Coombs RW, Krieger JN: **Semen as the Way Forward to Understand HIV-1 Transmission**. *The Journal of infectious diseases* 2016, **214**(10):1473-1474.
392. Coombs RW, Speck CE, Hughes JP, Lee W, Sampoleo R, Ross SO, Dragavon J, Peterson G, Hooton TM, Collier AC *et al*: **Association between culturable human immunodeficiency virus type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: evidence for compartmentalization of HIV-1 between semen and blood**. *The Journal of infectious diseases* 1998, **177**(2):320-330.
393. Craigo JK, Patterson BK, Paranjpe S, Kulka K, Ding M, Mellors J, Montelaro RC, Gupta P: **Persistent HIV type 1 infection in semen and blood compartments in patients after long-term potent antiretroviral therapy**. *AIDS research and human retroviruses* 2004, **20**(11):1196-1209.
394. Curran R, Ball JK: **Concordance between semen-derived HIV-1 proviral DNA and viral RNA hypervariable region 3 (V3) envelope sequences in cases where semen populations are distinct from those present in blood**. *Journal of medical virology* 2002, **67**(1):9-19.
395. Dejucq-Rainsford N, Jegou B: **Viruses in semen and male genital tissues--consequences for the reproductive system and therapeutic perspectives**. *Current pharmaceutical design* 2004, **10**(5):557-575.

396. Delwart EL, Mullins JI, Gupta P, Learn GH, Jr., Holodniy M, Katzenstein D, Walker BD, Singh MK: **Human immunodeficiency virus type 1 populations in blood and semen**. *J Virol* 1998, **72**(1):617-623.
397. Dondero F, Rossi T, D'Offizi G, Mazzilli F, Rosso R, Sarandrea N, Pinter E, Aiuti F: **Semen analysis in HIV seropositive men and in subjects at high risk for HIV infection**. *Human reproduction* 1996, **11**(4):765-768.
398. Gallicano K: **Antiretroviral-drug concentrations in semen**. *Antimicrobial agents and chemotherapy* 2000, **44**(4):1117-1118.
399. Gantner P, Assoumou L, Leruez-Ville M, David L, Suzan-Monti M, Costagliola D, Rouzioux C, Ghosn J, Ghosn J, Suzan-Monti M *et al*: **HIV-1-RNA in seminal plasma correlates with detection of HIV-1-DNA in semen cells, but not with CMV shedding, among MSM on successful antiretroviral regimens**. *Journal of Antimicrobial Chemotherapy* 2016, **71**(11):3202-3205.
400. Ghosn J, Viard J-P, Katlama C, de Almeida M, Tubiana R, Letourneur F, Aaron L, Goujard C, Salmon D, Leruez-Ville M *et al*: **Evidence of genotypic resistance diversity of archived and circulating viral strains in blood and semen of pre-treated HIV-infected men**. *Aids* 2004, **18**(3):447-457.
401. Gianella S, Mehta SR, Strain MC, Young JA, Vargas MV, Little SJ, Richman DD, Kosakovsky Pond SL, Smith DM: **Impact of seminal cytomegalovirus replication on HIV-1 dynamics between blood and semen**. *Journal of medical virology* 2012, **84**(11):1703-1709.
402. Halfon P, Giorgetti C, Khiri H, Penaranda G, Terriou P, Porcu-Buisson G, Chabert-Orsini V: **Semen may harbor HIV despite effective HAART: another piece in the puzzle**. *PloS one* 2010, **5**(5):e10569.
403. Kalichman SC, Cage M, Barnett T, Tharnish P, Rompa D, Austin J, Luke W, O'Mowrey J, Schinazi RF: **Human immunodeficiency virus in semen and plasma: investigation of sexual transmission risk behavioral correlates**. *AIDS research and human retroviruses* 2001, **17**(18):1695-1703.
404. Kiessling AA: **Semen transmission of human immunodeficiency virus**. *Fertility and sterility* 1992, **58**(4):667-669.
405. Kiessling AA: **Isolation of human immunodeficiency virus type 1 from semen and vaginal fluids**. *Methods in molecular biology (Clifton, NJ)* 2005, **304**:71-86.
406. Lorello G, la Porte C, Pilon R, Zhang G, Karnauchow T, MacPherson P: **Discordance in HIV-1 viral loads and antiretroviral drug concentrations comparing semen and blood plasma**. *HIV medicine* 2009, **10**(9):548-554.
407. Osborne BJ, Sheth PM, Yi TJ, Kovacs C, Benko E, Porte C, Huibner S, Le AQ, Danroth R, Baraki B *et al*: **Impact of antiretroviral therapy duration and intensification on isolated shedding of HIV-1 RNA in semen**. *The Journal of infectious diseases* 2013, **207**(8):1226-1234.
408. Pilcher CD, Joaki G, Hoffman IF, Martinson FE, Mapanje C, Stewart PW, Powers KA, Galvin S, Chilongozi D, Gama S *et al*: **Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection**. *Aids* 2007, **21**(13):1723-1730.
409. Pillai SK, Good B, Pond SK, Wong JK, Strain MC, Richman DD, Smith DM: **Semen-specific genetic characteristics of human immunodeficiency virus type 1 env**. *J Virol* 2005, **79**(3):1734-1742.
410. Sabatte J, Remes Lenicov F, Cabrini M, Rodriguez Rodrigues C, Ostrowski M, Ceballos A, Amigorena S, Geffner J: **The role of semen in sexual transmission of HIV: beyond a carrier for virus particles**. *Microbes and infection* 2011, **13**(12-13):977-982.
411. Shen C, Ding M, Craigo JK, Tarwater P, Chatterjee R, Roy P, Guha SK, Saha B, Modak D, Neogi D *et al*: **Genetic characterization of HIV-1 from semen and blood from clade C-infected subjects from India and effect of therapy in these body compartments**. *Virology* 2010, **401**(2):190-196.

412. Sheth PM, Kovacs C, Kemal KS, Jones RB, Raboud JM, Pilon R, la Porte C, Ostrowski M, Loutfy M, Burger H *et al*: **Persistent HIV RNA shedding in semen despite effective antiretroviral therapy**. *Aids* 2009, **23**(15):2050-2054.
413. Sheth PM, Shahabi K, Rebbapragada A, Kovacs C, Dimayuga R, Chackalakkal S, MacDonald K, Mazzulli T, Kaul R: **HIV viral shedding in semen: lack of correlation with systemic virus-specific CD8 responses**. *Aids* 2004, **18**(16):2202-2205.
414. Sheth PM, Yi TJ, Kovacs C, Kemal KS, Jones RB, Osborne B, Pilon R, la Porte C, Ostrowski M, Mazzulli T *et al*: **Mucosal correlates of isolated HIV semen shedding during effective antiretroviral therapy**. *Mucosal immunology* 2012, **5**(3):248-257.
415. Sutthent R, Sumrangsurp K, Wirachsilp P, Chaisilwattana P, Roongpisuthipong A, Chaiyakul P, Nooma P, Honda M, Warachit P: **Diversity of HIV-1 subtype E in semen and cervicovaginal secretion**. *Journal of human virology* 2001, **4**(5):260-268.
416. Tachet A, Dulioust E, Salmon D, De Almeida M, Rivalland S, Finkielsztejn L, Heard I, Jouannet P, Sicard D, Rouzioux C: **Detection and quantification of HIV-1 in semen: identification of a subpopulation of men at high potential risk of viral sexual transmission**. *Aids* 1999, **13**(7):823-831.
417. Taylor S, Back DJ, Drake SM, Workman J, Reynolds H, Gibbons SE, White DJ, Pillay D: **Antiretroviral drug concentrations in semen of HIV-infected men: differential penetration of indinavir, ritonavir and saquinavir**. *The Journal of antimicrobial chemotherapy* 2001, **48**(3):351-354.
418. Taylor S, Jayasuriya AN, Berry A, Gilleran G, Dufty NE, Else L, Back DJ, Smit EJ: **Darunavir concentrations exceed the protein-corrected EC(5)(0) for wild-type HIV in the semen of HIV-1-infected men**. *Aids* 2010, **24**(16):2583-2587.
419. Taylor S, Pereira AS: **Antiretroviral drug concentrations in semen of HIV-1 infected men**. *Sexually transmitted infections* 2001, **77**(1):4-11.
420. Vernazza PL, Gilliam BL, Dyer J, Fiscus SA, Eron JJ, Frank AC, Cohen MS: **Quantification of HIV in semen: correlation with antiviral treatment and immune status**. *Aids* 1997, **11**(8):987-993.
421. Baeten JM, Overbaugh J: **Measuring the infectiousness of persons with HIV-1: opportunities for preventing sexual HIV-1 transmission**. *Current HIV research* 2003, **1**(1):69-86.
422. Dulioust E, Tachet A, De Almeida M, Finkielsztejn L, Rivalland S, Salmon D, Sicard D, Rouzioux C, Jouannet P: **Detection of HIV-1 in seminal plasma and seminal cells of HIV-1 seropositive men**. *Journal of reproductive immunology* 1998, **41**(1-2):27-40.
423. Gantner P, Assoumou L, Leruez-Ville M, David L, Suzan-Monti M, Costagliola D, Rouzioux C, Ghosn J: **HIV-1-RNA in seminal plasma correlates with detection of HIV-1-DNA in semen cells, but not with CMV shedding, among MSM on successful antiretroviral regimens**. *The Journal of antimicrobial chemotherapy* 2016, **71**(11):3202-3205.
424. Louissaint NA, Nimmagadda S, Fuchs EJ, Bakshi RP, Cao YJ, Lee LA, Goldsmith J, Caffo BS, Du Y, King KE *et al*: **Distribution of cell-free and cell-associated HIV surrogates in the colon after simulated receptive anal intercourse in men who have sex with men**. *Journal of acquired immune deficiency syndromes* 2012, **59**(1):10-17.
425. Houzet L, Matusali G, Dejucq-Rainsford N: **Origins of HIV-infected leukocytes and virions in semen**. *The Journal of infectious diseases* 2014, **210 Suppl 3**:S622-630.
426. Krieger JN, Nirapathpongporn A, Chaiyaporn M, Peterson G, Nikolaeva I, Akridge R, Ross SO, Coombs RW: **Vasectomy and human immunodeficiency virus type 1 in semen**. *The Journal of urology* 1998, **159**(3):820-825; discussion 825-826.
427. Ilaria G, Jacobs JL, Polsky B, Koll B, Baron P, MacLow C, Armstrong D, Schlegel PN: **Detection of HIV-1 DNA sequences in pre-ejaculatory fluid**. *Lancet* 1992, **340**(8833):1469.
428. Mostad SB, Kreiss JK: **Shedding of HIV-1 in the genital tract**. *Aids* 1996, **10**(12):1305-1315.

429. Liuzzi G, Bagnarelli P, Chirianni A, Clementi M, Piazza M: **Quantitative techniques and HIV-1 load in plasma and semen**. *Aids* 1998, **12**(6):673.
430. Liuzzi G, Chirianni A, Clementi M, Bagnarelli P, Valenza A, Cataldo PT, Piazza M: **Analysis of HIV-1 load in blood, semen and saliva: evidence for different viral compartments in a cross-sectional and longitudinal study**. *Aids* 1996, **10**(14):F51-56.
431. Vernazza PL, Eron JJ, Cohen MS, van der Horst CM, Troiani L, Fiscus SA: **Detection and biologic characterization of infectious HIV-1 in semen of seropositive men**. *Aids* 1994, **8**(9):1325-1329.
432. Kroodsma KL, Kozal MJ, Hamed KA, Winters MA, Merigan TC: **Detection of drug resistance mutations in the human immunodeficiency virus type 1 (HIV-1) pol gene: differences in semen and blood HIV-1 RNA and proviral DNA**. *The Journal of infectious diseases* 1994, **170**(5):1292-1295.
433. Fieni F, Stone M, Ma ZM, Dutra J, Fritts L, Miller CJ: **Viral RNA levels and env variants in semen and tissues of mature male rhesus macaques infected with SIV by penile inoculation**. *PloS one* 2013, **8**(10):e76367.
434. Coltart CE, Lindsey B, Ghinai I, Johnson AM, Heymann DL: **The Ebola outbreak, 2013-2016: old lessons for new epidemics**. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2017, **372**(1721).
435. White MK, Wollebo HS, David Beckham J, Tyler KL, Khalili K: **Zika virus: An emergent neuropathological agent**. *Annals of neurology* 2016, **80**(4):479-489.
436. Moreira J, Peixoto TM, Siqueira AM, Lamas CC: **Sexually acquired Zika virus: a systematic review**. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2017, **23**(5):296-305.
437. Thorson A, Formenty P, Lofthouse C, Broutet N: **Systematic review of the literature on viral persistence and sexual transmission from recovered Ebola survivors: evidence and recommendations**. *BMJ open* 2016, **6**(1):e008859.
438. Donnell D, Baeten JM, Kiarie J, Thomas KK, Stevens W, Cohen CR, McIntyre J, Lingappa JR, Celum C: **Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis**. *Lancet* 2010, **375**(9731):2092-2098.
439. Taylor S, Boffito M, Vernazza PL: **Antiretroviral therapy to reduce the sexual transmission of HIV**. *Journal of HIV therapy* 2003, **8**(3):55-66.
440. Craigo JK, Gupta P: **HIV-1 in genital compartments: vexing viral reservoirs**. *Current opinion in HIV and AIDS* 2006, **1**(2):97-102.
441. Ferraretto X, Estellat C, Damond F, Longuet P, Epelboin S, Demailly P, Yazbeck C, Llabador MA, Pasquet B, Yazdanpanah Y *et al*: **Timing of intermittent seminal HIV-1 RNA shedding in patients with undetectable plasma viral load under combination antiretroviral therapy**. *PloS one* 2014, **9**(3):e88922.
442. Politch JA, Mayer KH, Welles SL, O'Brien WX, Xu C, Bowman FP, Anderson DJ: **Highly active antiretroviral therapy does not completely suppress HIV in semen of sexually active HIV-infected men who have sex with men**. *Aids* 2012, **26**(12):1535-1543.
443. Marcelin AG, Tubiana R, Lambert-Niclot S, Lefebvre G, Dominguez S, Bonmarchand M, Vauthier-Brouzes D, Marguet F, Mousset-Simeon N, Peytavin G *et al*: **Detection of HIV-1 RNA in seminal plasma samples from treated patients with undetectable HIV-1 RNA in blood plasma**. *Aids* 2008, **22**(13):1677-1679.
444. Else LJ, Taylor S, Back DJ, Khoo SH: **Pharmacokinetics of antiretroviral drugs in anatomical sanctuary sites: the male and female genital tract**. *Antiviral therapy* 2011, **16**(8):1149-1167.
445. Taylor S, Davies S: **Antiretroviral drug concentrations in the male and female genital tract: implications for the sexual transmission of HIV**. *Current opinion in HIV and AIDS* 2010, **5**(4):335-343.
446. Cohen MS: **Sexually transmitted diseases enhance HIV transmission: no longer a hypothesis**. *Lancet* 1998, **351 Suppl 3**:5-7.

447. Cohen MS: **HIV and sexually transmitted diseases: lethal synergy**. *Topics in HIV medicine : a publication of the International AIDS Society, USA* 2004, **12**(4):104-107.
448. Schacker T: **The role of HSV in the transmission and progression of HIV**. *Herpes : the journal of the IHMF* 2001, **8**(2):46-49.
449. Nagot N, Ouedraogo A, Foulongne V, Konate I, Weiss HA, Vergne L, Defer MC, Djagbare D, Sanon A, Andonaba JB *et al*: **Reduction of HIV-1 RNA levels with therapy to suppress herpes simplex virus**. *N Engl J Med* 2007, **356**(8):790-799.
450. Wald A, Link K: **Risk of human immunodeficiency virus infection in herpes simplex virus type 2-seropositive persons: a meta-analysis**. *The Journal of infectious diseases* 2002, **185**(1):45-52.
451. Desai DV, Kulkarni SS: **Herpes Simplex Virus: The Interplay Between HSV, Host, and HIV-1**. *Viral Immunol* 2015, **28**(10):546-555.
452. Palu G, Benetti L, Calistri A: **Molecular basis of the interactions between herpes simplex viruses and HIV-1**. *Herpes : the journal of the IHMF* 2001, **8**(2):50-55.
453. Sadiq ST, McSorley J, Copas AJ, Bennett J, Edwards SJ, Kaye S, Kirk S, French P, Weller IV: **The effects of early syphilis on CD4 counts and HIV-1 RNA viral loads in blood and semen**. *Sexually transmitted infections* 2005, **81**(5):380-385.
454. Ramsey KH, Schneider H, Cross AS, Boslego JW, Hoover DL, Staley TL, Kuschner RA, Deal CD: **Inflammatory cytokines produced in response to experimental human gonorrhea**. *The Journal of infectious diseases* 1995, **172**(1):186-191.
455. Ho JL, He S, Hu A, Geng J, Basile FG, Almeida MG, Saito AY, Laurence J, Johnson WD, Jr.: **Neutrophils from human immunodeficiency virus (HIV)-seronegative donors induce HIV replication from HIV-infected patients' mononuclear cells and cell lines: an in vitro model of HIV transmission facilitated by Chlamydia trachomatis**. *The Journal of experimental medicine* 1995, **181**(4):1493-1505.
456. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N *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* 2009, **83**(8):3556-3567.
457. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H *et al*: **Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(21):7552-7557.
458. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S *et al*: **Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing**. *J Virol* 2008, **82**(8):3952-3970.
459. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH *et al*: **Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission**. *Science* 2004, **303**(5666):2019-2022.
460. Gottlieb G, Heath L, Nickle D, Wong K, Leach S, Jacobs B, Gezahegne S, van't Wout A, Jacobson L, Margolick J *et al*: **HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study**. *The Journal of infectious diseases* 2008, **197**:1011 - 1015.
461. Ritola K, Pilcher CD, Fiscus SA, Hoffman NG, Nelson JA, Kitrinos KM, Hicks CB, Eron JJ, Jr., Swanstrom R: **Multiple V1/V2 env variants are frequently present during primary infection with human immunodeficiency virus type 1**. *J Virol* 2004, **78**(20):11208-11218.
462. Sagar M, Kirkegaard E, Long EM, Celum C, Buchbinder S, Daar ES, Overbaugh J: **Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes**. *J Virol* 2004, **78**(13):7279-7283.

463. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P: **Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection**. *J Virol* 1993, **67**(6):3345-3356.
464. Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, Kunstman KJ, Furtado MR, Munoz JL: **Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants**. *Science* 1992, **255**(5048):1134-1137.
465. Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, Ho DD: **Genotypic and phenotypic characterization of HIV-1 patients with primary infection**. *Science* 1993, **261**(5125):1179-1181.
466. Wolfs TF, Zwart G, Bakker M, Goudsmit J: **HIV-1 genomic RNA diversification following sexual and parenteral virus transmission**. *Virology* 1992, **189**(1):103-110.
467. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, Bedard HE, Gladden AD, Seese AM, Amero MA *et al*: **Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus**. *PLoS pathogens* 2016, **12**(5):e1005619.
468. Kariuki SM, Selhorst P, Arien KK, Dorfman JR: **The HIV-1 transmission bottleneck**. *Retrovirology* 2017, **14**(1):22.
469. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM *et al*: **Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1**. *PLoS pathogens* 2009, **5**(1):e1000274.
470. Naranbhai V, Abdool Karim SS, Altfeld M, Samsunder N, Durgiah R, Sibeko S, Abdool Karim Q, Carr WH: **Innate immune activation enhances hiv acquisition in women, diminishing the effectiveness of tenofovir microbicide gel**. *The Journal of infectious diseases* 2012, **206**(7):993-1001.
471. Zhang ZQ, Wietgrefe SW, Li Q, Shore MD, Duan L, Reilly C, Lifson JD, Haase AT: **Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(15):5640-5645.
472. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N *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* 2009, **83**(8):3556-3567.
473. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, Decker JM, Wang S, Baalwa J, Kraus MH *et al*: **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection**. *The Journal of experimental medicine* 2009, **206**(6):1273-1289.
474. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC *et al*: **Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing**. *PloS one* 2010, **5**(8):e12303.
475. Bar K, Li H, Chamberland A, Tremblay C, Routy J, Grayson T, Sun C, Wang S, Learn G, Morgan C: **Wide variation in the multiplicity of HIV-1 infection among injection drug users**. *J Virol* 2010, **84**:6241 - 6247.
476. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, Salazar-Gonzalez JF, Salazar MG, Learn GH, Morgan CJ *et al*: **High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men**. *PLoS pathogens* 2010, **6**(5):e1000890.
477. Russell ES, Kwiek JJ, Keys J, Barton K, Mwapasa V, Montefiori DC, Meshnick SR, Swanstrom R: **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* 2011, **85**(16):8253-8262.

478. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, Sun C, Wang S, Learn GH, Morgan CJ *et al*: **Wide variation in the multiplicity of HIV-1 infection among injection drug users**. *J Virol* 2010, **84**(12):6241-6247.
479. Clark F: **Gaps remain in Russia's response to HIV/AIDS**. *Lancet* 2016, **388**(10047):857-858.
480. Dukhovlinova E, Masharsky A, Toussova O, Verevochkin S, Solovyeva T, Meringof M, Paintsil E, White E, Barbour R, Heimer R *et al*: **Two Independent HIV Epidemics in Saint Petersburg, Russia Revealed by Molecular Epidemiology**. *AIDS research and human retroviruses* 2015, **31**(6):608-614.
481. Li H, Bar K, Wang S, Decker J, Chen Y, Sun C, Salazar-Gonzalez J, Salazar M, Learn G, Morgan C: **High multiplicity infection by HIV-1 in Men Who have Sex with Men**. *PLoS pathogens* 2010, **6**:e1000890.
482. Dukhovlinova Elena, Masharsky Alexey, Verevochkin Sergei, Toussova Olga, Shevchenko Andrey, Montefiori David, Horfman Irving, Miller William, Kozlov Andrei, Ronald S: **AHI detection among people who inject drugs in Russia reveals the HIV-1 transmission bottleneck**. In: *Conference on Retroviruses and Opportunistic Infections (CROI): 2014; Boston, Massachusetts*: International Antiviral Sociey - USA; 2014.
483. Zule WA, Ticknor-Stellato KM, Desmond DP, Vogtsberger KN: **Evaluation of needle and syringe combinations**. *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association* 1997, **14**(3):294-295.
484. Greenfield L, Bigelow GE, Brooner RK: **HIV risk behavior in drug users: increased blood "booting" during cocaine injection**. *AIDS education and prevention : official publication of the International Society for AIDS Education* 1992, **4**(2):95-107.
485. Bobashev GV, Zule WA: **Modeling the effect of high dead-space syringes on the human immunodeficiency virus (HIV) epidemic among injecting drug users**. *Addiction* 2010, **105**(8):1439-1447.
486. Zule WA, Bobashev G: **High dead-space syringes and the risk of HIV and HCV infection among injecting drug users**. *Drug and alcohol dependence* 2009, **100**(3):204-213.
487. Zule WA, Desmond DP, Neff JA: **Syringe type and drug injector risk for HIV infection: a case study in Texas**. *Social science & medicine (1982)* 2002, **55**(7):1103-1113.
488. Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, Sun C, Chen Y, Yeh WW, Letvin NL *et al*: **Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1**. *The Journal of experimental medicine* 2009, **206**(5):1117-1134.
489. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, Deymier MJ, Ende ZS, Klatt NR, DeZiel CE *et al*: **HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck**. *Science* 2014, **345**(6193):1254031.
490. Boeras DI, Hraber PT, Hurlston M, Evans-Strickfaden T, Bhattacharya T, Giorgi EE, Mulenga J, Karita E, Korber BT, Allen S *et al*: **Role of donor genital tract HIV-1 diversity in the transmission bottleneck**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(46):E1156-1163.
491. Lifson JD, Nowak MA, Goldstein S, Rossio JL, Kinter A, Vasquez G, Wiltrout TA, Brown C, Schneider D, Wahl L *et al*: **The extent of early viral replication is a critical determinant of the natural history of simian immunodeficiency virus infection**. *J Virol* 1997, **71**(12):9508-9514.
492. Sagar M, Lavreys L, Baeten JM, Richardson BA, Mandaliya K, Chohan BH, Kreiss JK, Overbaugh J: **Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression**. *J Virol* 2003, **77**(23):12921-12926.
493. Theus SA, Harrich DA, Gaynor R, Radolf JD, Norgard MV: **Treponema pallidum, lipoproteins, and synthetic lipoprotein analogues induce human immunodeficiency virus type 1 gene**

expression in monocytes via NF-kappaB activation. *The Journal of infectious diseases* 1998, **177**(4):941-950.
494. Buckner LR, Amedee AM, Albritton HL, Kozlowski PA, Lacour N, McGowin CL, Schust DJ, Quayle AJ: **Chlamydia trachomatis Infection of Endocervical Epithelial Cells Enhances Early HIV Transmission Events**. *PloS one* 2016, **11**(1):e0146663.
495. Jarvis GA, Chang TL: **Modulation of HIV transmission by Neisseria gonorrhoeae: molecular and immunological aspects**. *Current HIV research* 2012, **10**(3):211-217.
496. Sagar M, Lavreys L, Baeten JM, Richardson BA, Mandaliya K, Ndinya-Achola JO, Kreiss JK, Overbaugh J: **Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population**. *Aids* 2004, **18**(4):615-619.
497. Nazli A, Chan O, Dobson-Belaire WN, Ouellet M, Tremblay MJ, Gray-Owen SD, Arsenault AL, Kaushic C: **Exposure to HIV-1 directly impairs mucosal epithelial barrier integrity allowing microbial translocation**. *PLoS pathogens* 2010, **6**(4):e1000852.
498. Selhorst P, Masson L, Ismail SD, Samsunder N, Garrett N, Mansoor LE, Abdool Karim Q, Abdool Karim SS, Passmore JS, Williamson C: **Cervicovaginal Inflammation Facilitates Acquisition of Less Infectious HIV Variants**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017, **64**(1):79-82.
499. Frange P, Meyer L, Jung M, Goujard C, Zucman D, Abel S, Hochedez P, Gousset M, Gascuel O, Rouzioux C *et al*: **Sexually-transmitted/founder HIV-1 cannot be directly predicted from plasma or PBMC-derived viral quasispecies in the transmitting partner**. *PloS one* 2013, **8**(7):e69144.
500. Rieder P, Joos B, Scherrer AU, Kuster H, Braun D, Grube C, Niederost B, Leemann C, Gianella S, Metzner KJ *et al*: **Characterization of human immunodeficiency virus type 1 (HIV-1) diversity and tropism in 145 patients with primary HIV-1 infection**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2011, **53**(12):1271-1279.
501. Fleming DT, Wasserheit JN: **From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection**. *Sexually transmitted infections* 1999, **75**(1):3-17.
502. Isaacman-Beck J, Hermann EA, Yi Y, Ratcliffe SJ, Mulenga J, Allen S, Hunter E, Derdeyn CA, Collman RG: **Heterosexual transmission of human immunodeficiency virus type 1 subtype C: Macrophage tropism, alternative coreceptor use, and the molecular anatomy of CCR5 utilization**. *J Virol* 2009, **83**(16):8208-8220.
503. Margolis L, Shattock R: **Selective transmission of CCR5-utilizing HIV-1: the 'gatekeeper' problem resolved?** *Nature reviews Microbiology* 2006, **4**(4):312-317.
504. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee FH, Didigu CA, Berro R, Klasse PJ, Lee B *et al*: **Transmitted/founder and chronic HIV-1 envelope proteins are distinguished by differential utilization of CCR5**. *J Virol* 2013, **87**(5):2401-2411.
505. Parrish NF, Wilen CB, Banks LB, Iyer SS, Pfaff JM, Salazar-Gonzalez JF, Salazar MG, Decker JM, Parrish EH, Berg A *et al*: **Transmitted/founder and chronic subtype C HIV-1 use CD4 and CCR5 receptors with equal efficiency and are not inhibited by blocking the integrin alpha4beta7**. *PLoS pathogens* 2012, **8**(5):e1002686.
506. Ping LH, Joseph SB, Anderson JA, Abrahams MR, Salazar-Gonzalez JF, Kincer LP, Treurnicht FK, Arney L, Ojeda S, Zhang M *et al*: **Comparison of viral Env proteins from acute and chronic infections with subtype C human immunodeficiency virus type 1 identifies differences in glycosylation and CCR5 utilization and suggests a new strategy for immunogen design**. *J Virol* 2013, **87**(13):7218-7233.
507. Joseph SB, Swanstrom R: **HIV/AIDS. A fitness bottleneck in HIV-1 transmission**. *Science* 2014, **345**(6193):136-137.

508. Joseph SB, Swanstrom R, Kashuba AD, Cohen MS: **Bottlenecks in HIV-1 transmission: insights from the study of founder viruses**. *Nature reviews Microbiology* 2015, **13**(7):414-425.
509. Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, Shen T, Gaschen B, Krishnamoorthy M, Li H *et al*: **Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections**. *PLoS pathogens* 2011, **7**(9):e1002209.
510. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, Overbaugh J: **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* 2005, **79**(10):6528-6531.
511. Wilen CB, Parrish NF, Pfaff JM, Decker JM, Henning EA, Haim H, Petersen JE, Wojcechowskyj JA, Sodroski J, Haynes BF *et al*: **Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins**. *J Virol* 2011, **85**(17):8514-8527.
512. Oberle CS, Joos B, Rusert P, Campbell NK, Beauparlant D, Kuster H, Weber J, Schenkel CD, Scherrer AU, Magnus C *et al*: **Tracing HIV-1 transmission: envelope traits of HIV-1 transmitter and recipient pairs**. *Retrovirology* 2016, **13**(1):62.
513. Cavrois M, Neidleman J, Santiago ML, Derdeyn CA, Hunter E, Greene WC: **Enhanced fusion and virion incorporation for HIV-1 subtype C envelope glycoproteins with compact V1/V2 domains**. *J Virol* 2014, **88**(4):2083-2094.
514. Curlin ME, Zioni R, Hawes SE, Liu Y, Deng W, Gottlieb GS, Zhu T, Mullins JI: **HIV-1 envelope subregion length variation during disease progression**. *PLoS pathogens* 2010, **6**(12):e1001228.
515. Delwart EL, Pan H, Sheppard HW, Wolpert D, Neumann AU, Korber B, Mullins JI: **Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS**. *J Virol* 1997, **71**(10):7498-7508.
516. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS *et al*: **Antibody neutralization and escape by HIV-1**. *Nature* 2003, **422**(6929):307-312.
517. Falkowska E, Le KM, Ramos A, Doores KJ, Lee JH, Blattner C, Ramirez A, Derking R, van Gils MJ, Liang CH *et al*: **Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers**. *Immunity* 2014, **40**(5):657-668.
518. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K *et al*: **Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9**. *Nature* 2011, **480**(7377):336-343.
519. Nawaz F, Cicala C, Van Ryk D, Block KE, Jelicic K, McNally JP, Ogundare O, Pascuccio M, Patel N, Wei D *et al*: **The genotype of early-transmitting HIV gp120s promotes alpha (4) beta(7)-reactivity, revealing alpha (4) beta(7) +/CD4+ T cells as key targets in mucosal transmission**. *PLoS pathogens* 2011, **7**(2):e1001301.
520. Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, Zhu H, Woodward AS, Maenza J, Stevens CE, Stekler J *et al*: **Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses**. *Virology* 2008, **374**(2):229-233.
521. Grivel JC, Shattock RJ, Margolis LB: **Selective transmission of R5 HIV-1 variants: where is the gatekeeper?** *Journal of translational medicine* 2011, **9 Suppl 1**:S6.
522. Gray ES, Madiga MC, Hermanus T, Moore PL, Wibmer CK, Tumba NL, Werner L, Mlisana K, Sibeko S, Williamson C *et al*: **The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection**. *J Virol* 2011, **85**(10):4828-4840.

523. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, Treurnicht F, Mlisana K, Shaw GM *et al*: **Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection**. *J Virol* 2007, **81**(12):6187-6196.
524. Bar KJ, Tsao CY, Iyer SS, Decker JM, Yang Y, Bonsignori M, Chen X, Hwang KK, Montefiori DC, Liao HX *et al*: **Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape**. *PLoS pathogens* 2012, **8**(5):e1002721.
525. Frost SD, Liu Y, Pond SL, Chappey C, Wrin T, Petropoulos CJ, Little SJ, Richman DD: **Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B**. *J Virol* 2005, **79**(10):6523-6527.
526. Deymier MJ, Ende Z, Fenton-May AE, Dilernia DA, Kilembe W, Allen SA, Borrow P, Hunter E: **Heterosexual Transmission of Subtype C HIV-1 Selects Consensus-Like Variants without Increased Replicative Capacity or Interferon-alpha Resistance**. *PLoS pathogens* 2015, **11**(9):e1005154.
527. Puffer BA, Pohlmann S, Edinger AL, Carlin D, Sanchez MD, Reitter J, Watry DD, Fox HS, Desrosiers RC, Doms RW: **CD4 independence of simian immunodeficiency virus Envs is associated with macrophage tropism, neutralization sensitivity, and attenuated pathogenicity**. *J Virol* 2002, **76**(6):2595-2605.
528. Edwards TG, Hoffman TL, Baribaud F, Wyss S, LaBranche CC, Romano J, Adkinson J, Sharron M, Hoxie JA, Doms RW: **Relationships between CD4 independence, neutralization sensitivity, and exposure of a CD4-induced epitope in a human immunodeficiency virus type 1 envelope protein**. *J Virol* 2001, **75**(11):5230-5239.
529. Brown RJ, Peters PJ, Caron C, Gonzalez-Perez MP, Stones L, Ankghuambom C, Pondei K, McClure CP, Alemnji G, Taylor S *et al*: **Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors**. *J Virol* 2011, **85**(12):6024-6037.
530. Chaillon A, Smith DM, Vanpouille C, Lisco A, Jordan P, Caballero G, Vargas M, Gianella S, Mehta SR: **HIV Trafficking Between Blood and Semen During Early Untreated HIV Infection**. *Journal of acquired immune deficiency syndromes* 2017, **74**(1):95-102.
531. Seifert D, Di Giallonardo F, Töpfer A, Singer J, Schmutz S, Günthard HF, Beerenwinkel N, Metzner KJ: **A Comprehensive Analysis of Primer IDs to Study Heterogeneous HIV-1 Populations**. *Journal of molecular biology* 2016, **428**(1):238-250.
532. Seifert D, Di Giallonardo F, Topfer A, Singer J, Schmutz S, Gunthard HF, Beerenwinkel N, Metzner KJ: **A Comprehensive Analysis of Primer IDs to Study Heterogeneous HIV-1 Populations**. *Journal of molecular biology* 2016, **428**(1):238-250.
533. Zhou S, Jones C, Mieczkowski P, Swanstrom R: **Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations**. *J Virol* 2015, **89**(16):8540-8555.
534. Keys JR, Zhou S, Anderson JA, Eron JJ, Jr., Rackoff LA, Jabara C, Swanstrom R: **Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain**. *AIDS research and human retroviruses* 2015, **31**(6):658-668.
535. Brodin J, Hedskog C, Heddini A, Benard E, Neher RA, Mild M, Albert J: **Challenges with using primer IDs to improve accuracy of next generation sequencing**. *PloS one* 2015, **10**(3):e0119123.
536. Sheward DJ, Murrell B, Williamson C: **Degenerate Primer IDs and the birthday problem**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(21):E1330; author reply E1331.
537. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R: **Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(50):20166-20171.

538. Eyre RC, Zheng G, Kiessling AA: **Multiple drug resistance mutations in human immunodeficiency virus in semen but not blood of a man on antiretroviral therapy**. *Urology* 2000, **55**(4):591.
539. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V: **Evolutionary and immunological implications of contemporary HIV-1 variation**. *British medical bulletin* 2001, **58**:19-42.
540. Butler DM, Delport W, Kosakovsky Pond SL, Lakdawala MK, Cheng PM, Little SJ, Richman DD, Smith DM: **The origins of sexually transmitted HIV among men who have sex with men**. *Science translational medicine* 2010, **2**(18):18re11.
541. Ghosn J, Viard JP, Katlama C, de Almeida M, Tubiana R, Letourneur F, Aaron L, Goujard C, Salmon D, Leruez-Ville M *et al*: **Evidence of genotypic resistance diversity of archived and circulating viral strains in blood and semen of pre-treated HIV-infected men**. *Aids* 2004, **18**(3):447-457.
542. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, Davidov O, Zhao LP, Mullins JI: **HIV quasispecies and resampling**. *Science* 1996, **273**(5274):415-416.
543. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Jr., Dewar RL *et al*: **Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis**. *Journal of clinical microbiology* 2005, **43**(1):406-413.
544. Long EM, Martin HL, Jr., Kreiss JK, Rainwater SM, Lavreys L, Jackson DJ, Rakwar J, Mandaliya K, Overbaugh J: **Gender differences in HIV-1 diversity at time of infection**. *Nature medicine* 2000, **6**(1):71-75.
545. Learn GH, Muthui D, Brodie SJ, Zhu T, Diem K, Mullins JI, Corey L: **Virus population homogenization following acute human immunodeficiency virus type 1 infection**. *J Virol* 2002, **76**(23):11953-11959.
546. Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, Paxinos E, Arts EJ, Robertson DL, Mimms L *et al*: **Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism**. *PloS one* 2012, **7**(11):e49602.
547. Archer J, Braverman MS, Taillon BE, Desany B, James I, Harrigan PR, Lewis M, Robertson DL: **Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing**. *Aids* 2009, **23**:1209 - 1218.
548. Swenson LC, Mo T, Dong WW, Zhong X, Woods CK, Jensen MA, Thielen A, Chapman D, Lewis M, James I *et al*: **Deep sequencing to infer HIV-1 co-receptor usage: application to three clinical trials of maraviroc in treatment-experienced patients**. *The Journal of infectious diseases* 2011, **203**(2):237-245.
549. Bunnik EM, Swenson LC, Edo-Matas D, Huang W, Dong W, Frantzell A, Petropoulos CJ, Coakley E, Schuitemaker H, Harrigan PR *et al*: **Detection of inferred CCR5- and CXCR4-using HIV-1 variants and evolutionary intermediates using ultra-deep pyrosequencing**. *PLoS pathogens* 2011, **7**(6):e1002106.
550. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E: **The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing**. *Journal of virological methods* 2010, **169**(1):248-252.
551. Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K *et al*: **Complete nucleotide sequence of the AIDS virus, HTLV-III**. *Nature* 1985, **313**(6000):277-284.
552. !!! INVALID CITATION !!! [23].
553. Larsson A: **AliView: a fast and lightweight alignment viewer and editor for large datasets**. *Bioinformatics* 2014, **30**(22):3276-3278.
554. Price MN, Dehal PS, Arkin AP: **FastTree 2--approximately maximum-likelihood trees for large alignments**. *PloS one* 2010, **5**(3):e9490.

555. Rognes T, Flouri T, Nichols B, Quince C, Mahe F: **VSEARCH: a versatile open source tool for metagenomics**. *PeerJ* 2016, **4**:e2584.
556. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A: **Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10**. *Virus evolution* 2018, **4**(1):vey016.
557. Posada D: **jModelTest: phylogenetic model averaging**. *Molecular biology and evolution* 2008, **25**(7):1253-1256.
558. Magee D, Suchard MA, Scotch M: **Bayesian phylogeography of influenza A/H3N2 for the 2014-15 season in the United States using three frameworks of ancestral state reconstruction**. *PLoS computational biology* 2017, **13**(2):e1005389.
559. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA: **Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7**. *Systematic biology* 2018, **67**(5):901-904.
560. Rambaut A: **FigTree v1.3.1.** *Institute of Evolutionary Biology, University of Edinburgh, Edinburgh* 2010.
561. Hudson RR: **A new statistic for detecting genetic differentiation**. *Genetics* 2000, **155**(4):2011-2014.
562. Hudson RR, Boos DD, Kaplan NL: **A statistical test for detecting geographic subdivision**. *Molecular biology and evolution* 1992, **9**(1):138-151.
563. Hudson RR, Slatkin M, Maddison WP: **Estimation of levels of gene flow from DNA sequence data**. *Genetics* 1992, **132**(2):583-589.
564. Slatkin M, Maddison WP: **A cladistic measure of gene flow inferred from the phylogenies of alleles**. *Genetics* 1989, **123**(3):603-613.
565. Pond SL, Frost SD, Muse SV: **HyPhy: hypothesis testing using phylogenies**. *Bioinformatics* 2005, **21**(5):676-679.
566. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**(9):1312-1313.
567. Paradis E, Claude J, Strimmer K: **APE: Analyses of Phylogenetics and Evolution in R language**. *Bioinformatics* 2004, **20**(2):289-290.
568. Huerta-Cepas J, Serra F, Bork P: **ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data**. *Molecular biology and evolution* 2016, **33**(6):1635-1638.
569. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Molecular biology and evolution* 2016, **33**(7):1870-1874.
570. Govender S, Otwombe K, Essien T, Panchia R, de Bruyn G, Mohapi L, Gray G, Martinson N: **CD4 counts and viral loads of newly diagnosed HIV-infected individuals: implications for treatment as prevention**. *PloS one* 2014, **9**(3):e90754.
571. Agthe M, Nemes E, Jacob RA, Abrahams F, Fainguem N, Tetang SM, Cappelli G, Colizzi V, Dorfman JR: **Lower anti-HIV-1 neutralization in HIV-infected children with CD4+ T cell depletion: opposite correlation to that in adults**. *Aids* 2014, **28**(11):1694-1696.
572. Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N *et al*: **Ancient hybridization and an Irish origin for the modern polar bear matriline**. *Current biology : CB* 2011, **21**(15):1251-1258.
573. Tongo M, Essomba RG, Nindo F, Abrahams F, Nanfack AJ, Fokam J, Takou D, Torimiro JN, Mpoudi-Ngole E, Burgers WA *et al*: **Phylogenetics of HIV-1 subtype G env: Greater complexity and older origins than previously reported**. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2015, **35**:9-18.
574. Curran R, Ball JK: **Concordance between semen-derived HIV-1 proviral DNA and viral RNA hypervariable region 3 (V3) envelope sequences in cases where semen populations are distinct from those present in blood**. *Journal of medical virology* 2002, **67**(1):9-19.
575. Kosakovsky Pond SL, Frost SDW: **Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection**. *Molecular biology and evolution* 2005, **22**(5):1208-1222.

576. Lee SK, Zhou S, Baldoni PL, Spielvogel E, Archin NM, Hudgens MG, Margolis DM, Swanstrom R: **Quantification of the Latent HIV-1 Reservoir Using Ultra Deep Sequencing and Primer ID in a Viral Outgrowth Assay**. *Journal of acquired immune deficiency syndromes* 2017, **74**(2):221-228.
577. Zarate S, Pond SL, Shapshak P, Frost SD: **Comparative study of methods for detecting sequence compartmentalization in human immunodeficiency virus type 1**. *J Virol* 2007, **81**(12):6643-6651.
578. Thompson JR, Marcelino LA, Polz MF: **Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'**. *Nucleic acids research* 2002, **30**(9):2083-2088.
579. Meyerhans A, Vartanian JP, Wain-Hobson S: **DNA recombination during PCR**. *Nucleic acids research* 1990, **18**(7):1687-1691.
580. Tongo M, Dorfman JR, Abrahams MR, Mpoudi-Ngole E, Burgers WA, Martin DP: **Near full-length HIV type 1M genomic sequences from Cameroon : Evidence of early diverging under-sampled lineages in the country**. *Evolution, medicine, and public health* 2015, **2015**(1):254-265.
581. Permar SR, Salazar MG, Gao F, Cai F, Learn GH, Kalilani L, Hahn BH, Shaw GM, Salazar-Gonzalez JF: **Clonal amplification and maternal-infant transmission of nevirapine-resistant HIV-1 variants in breast milk following single-dose nevirapine prophylaxis**. *Retrovirology* 2013, **10**:88.
582. Serwadda D, Gray RH, Sewankambo NK, Wabwire-Mangen F, Chen MZ, Quinn TC, Lutalo T, Kiwanuka N, Kigozi G, Nalugoda F *et al*: **Human immunodeficiency virus acquisition associated with genital ulcer disease and herpes simplex virus type 2 infection: a nested case-control study in Rakai, Uganda**. *The Journal of infectious diseases* 2003, **188**(10):1492-1497.
583. Rusert P, Kuster H, Joos B, Misselwitz B, Gujer C, Leemann C, Fischer M, Stiegler G, Katinger H, Olson WC *et al*: **Virus isolates during acute and chronic human immunodeficiency virus type 1 infection show distinct patterns of sensitivity to entry inhibitors**. *J Virol* 2005, **79**(13):8454-8469.
584. Moore P, Gray E, Wibmer C, Bhiman J, Nonyane M, Sheward D, Hermanus T, Bajimaya S, Tumba N, Abrahams M: **Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape**. *Nature medicine* 2012, **18**:1688 - 1692.
585. Veazey RS, Ketas TJ, Dufour J, Moroney-Rasmussen T, Green LC, Klasse PJ, Moore JP: **Protection of rhesus macaques from vaginal infection by vaginally delivered maraviroc, an inhibitor of HIV-1 entry via the CCR5 co-receptor**. *The Journal of infectious diseases* 2010, **202**(5):739-744.
586. Veazey RS, Klasse PJ, Schader SM, Hu Q, Ketas TJ, Lu M, Marx PA, Dufour J, Colonno RJ, Shattock RJ *et al*: **Protection of macaques from vaginal SHIV challenge by vaginally delivered inhibitors of virus-cell fusion**. *Nature* 2005, **438**(7064):99-102.
587. Veazey RS, Springer MS, Marx PA, Dufour J, Klasse PJ, Moore JP: **Protection of macaques from vaginal SHIV challenge by an orally delivered CCR5 inhibitor**. *Nature medicine* 2005, **11**(12):1293-1294.
588. Lederman MM, Jump R, Pilch-Cooper HA, Root M, Sieg SF: **Topical application of entry inhibitors as "virustats" to prevent sexual transmission of HIV infection**. *Retrovirology* 2008, **5**:116.
589. Lederman MM, Veazey RS, Offord R, Mosier DE, Dufour J, Mefford M, Piatak M, Jr., Lifson JD, Salkowitz JR, Rodriguez B *et al*: **Prevention of vaginal SHIV transmission in rhesus macaques through inhibition of CCR5**. *Science* 2004, **306**(5695):485-487.
590. Kraus MH, Parrish NF, Shaw KS, Decker JM, Keele BF, Salazar-Gonzalez JF, Grayson T, McPherson DT, Ping LH, Anderson JA *et al*: **A rev1-vpu polymorphism unique to HIV-1 subtype A and C strains impairs envelope glycoprotein expression from rev-vpu-env**

cassettes and reduces virion infectivity in pseudotyping assays. *Virology* 2010, **397**(2):346-357.
591. Tarr AW, Owsianka AM, Szwejk A, Ball JK, Patel AH: **Cloning, expression, and functional analysis of patient-derived hepatitis C virus glycoproteins**. *Methods in molecular biology (Clifton, NJ)* 2007, **379**:177-197.
592. Montefiori DC: **Measuring HIV neutralization in a luciferase reporter gene assay**. *Methods in molecular biology (Clifton, NJ)* 2009, **485**:395-405.
593. Kilby JM, Hopkins S, Venetta TM, DiMassimo B, Cloud GA, Lee JY, Alldredge L, Hunter E, Lambert D, Bolognesi D *et al*: **Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry**. *Nature medicine* 1998, **4**(11):1302-1307.
594. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Molecular biology and evolution* 2013, **30**(4):772-780.
595. Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, Serwadda D, Sewankambo NK, Shepherd JC, Toma J *et al*: **Selection of HIV variants with signature genotypic characteristics during heterosexual transmission**. *The Journal of infectious diseases* 2009, **199**(4):580-589.
596. 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* 2006, **80**(19):9586-9598.
597. Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, Paxinos E, Chappey C, Galovich J, Beauchaine J, Petropoulos CJ *et al*: **Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(51):18514-18519.
598. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL *et al*: **Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target**. *Science* 2009, **326**(5950):285-289.
599. Poon AF, Lewis FI, Pond SL, Frost SD: **Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope**. *PLoS computational biology* 2007, **3**(1):e11.
600. Moore PL, Murphy MK, Derdeyn CA: **Role of HIV Glycans in Transmission and Immune Escape**. In: *HIV glycans in infection and immunity.* edn. Edited by Pantophlet R. New York, NY: Springer New York; 2014: 85-115.
601. Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, Sodroski J: **A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding**. *Science* 1998, **280**(5371):1949-1953.
602. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA: **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody**. *Nature* 1998, **393**(6686):648-659.
603. Ji X, Gewurz H, Spear GT: **Mannose binding lectin (MBL) and HIV**. *Molecular immunology* 2005, **42**(2):145-152.
604. Berger EA, Murphy PM, Farber JM: **Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease**. *Annu Rev Immunol* 1999, **17**:657-700.
605. Alexander M, Lynch R, Mulenga J, Allen S, Derdeyn CA, Hunter E: **Donor and recipient envs from heterosexual human immunodeficiency virus subtype C transmission pairs require high receptor levels for entry**. *J Virol* 2010, **84**(8):4100-4104.
606. Kishko M, Somasundaran M, Brewster F, Sullivan JL, Clapham PR, Luzuriaga K: **Genotypic and functional properties of early infant HIV-1 envelopes**. *Retrovirology* 2011, **8**:67.
607. Lobritz MA, Ratcliff AN, Marozsan AJ, Dudley DM, Tilton JC, Arts EJ: **Multifaceted mechanisms of HIV inhibition and resistance to CCR5 inhibitors PSC-RANTES and Maraviroc**. *Antimicrobial agents and chemotherapy* 2013, **57**(6):2640-2650.

608. Pastore C, Picchio GR, Galimi F, Fish R, Hartley O, Offord RE, Mosier DE: **Two mechanisms for human immunodeficiency virus type 1 inhibition by N-terminal modifications of RANTES**. *Antimicrobial agents and chemotherapy* 2003, **47**(2):509-517.
609. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D: **Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1**. *J Virol* 1998, **72**(4):2855-2864.
610. Murray JM, Maher S, Mota T, Suzuki K, Kelleher AD, Center RJ, Purcell D: **Differentiating founder and chronic HIV envelope sequences**. *PloS one* 2017, **12**(2):e0171572.
611. Song H, Hora B, Giorgi EE, Kumar A, Cai F, Bhattacharya T, Perelson AS, Gao F: **Transmission of Multiple HIV-1 Subtype C Transmitted/founder Viruses into the Same Recipients Was not Determined by Modest Phenotypic Differences**. *Scientific reports* 2016, **6**:38130.
612. Stoddard MB, Li H, Wang S, Saeed M, Andrus L, Ding W, Jiang X, Learn GH, von Schaewen M, Wen J *et al*: **Identification, molecular cloning, and analysis of full-length hepatitis C virus transmitted/founder genotypes 1, 3, and 4**. *mBio* 2015, **6**(2):e02518.
613. Killick M, Capovilla A, Papathanasopoulos MA: **Generation and characterization of an HIV-1 subtype C transmitted and early founder virus consensus sequence**. *AIDS research and human retroviruses* 2014, **30**(10):1001-1005.
614. Jafari M, Guatelli J, Lewinski MK: **Activities of transmitted/founder and chronic clade B HIV-1 Vpu and a C-terminal polymorphism specifically affecting virion release**. *J Virol* 2014, **88**(9):5062-5078.
615. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, Zajic L, Iyer SS, Decker JM, Kumar A *et al*: **Phenotypic properties of transmitted founder HIV-1**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(17):6626-6633.
616. Frange P, Meyer L, Jung M, Goujard C, Zucman D, Abel S, Hochedez P, Gousset M, Gascuel O, Rouzioux C *et al*: **Sexually-transmitted/founder HIV-1 cannot be directly predicted from plasma or PBMC-derived viral quasispecies in the transmitting partner**. *PloS one* 2013, **8**(7):e69144.
617. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, Pellegrino P, Williams I, Cohen MS, Gao F *et al*: **Relative resistance of HIV-1 founder viruses to control by interferon-alpha**. *Retrovirology* 2013, **10**:146.
618. Baalwa J, Wang S, Parrish NF, Decker JM, Keele BF, Learn GH, Yue L, Ruzagira E, Ssemwanga D, Kamali A *et al*: **Molecular identification, cloning and characterization of transmitted/founder HIV-1 subtype A, D and A/D infectious molecular clones**. *Virology* 2013, **436**(1):33-48.
619. Ochsenbauer C, Edmonds TG, Ding H, Keele BF, Decker J, Salazar MG, Salazar-Gonzalez JF, Shattock R, Haynes BF, Shaw GM *et al*: **Generation of transmitted/founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages**. *J Virol* 2012, **86**(5):2715-2728.
620. Merbah M, Arakelyan A, Edmonds T, Ochsenbauer C, Kappes JC, Shattock RJ, Grivel JC, Margolis LB: **HIV-1 expressing the envelopes of transmitted/founder or control/reference viruses have similar infection patterns of CD4 T-cells in human cervical tissue ex vivo**. *PloS one* 2012, **7**(12):e50839.
621. Redd AD, Collinson-Streng AN, Chatziandreou N, Mullis CE, Laeyendecker O, Martens C, Ricklefs S, Kiwanuka N, Nyein PH, Lutalo T *et al*: **Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions**. *The Journal of infectious diseases* 2012, **206**(9):1433-1442.
622. Shepard RN, Schock J, Robertson K, Shugars DC, Dyer J, Vernazza P, Hall C, Cohen MS, Fiscus SA: **Quantitation of human immunodeficiency virus type 1 RNA in different biological compartments**. *Journal of clinical microbiology* 2000, **38**(4):1414-1418.
623. Jacob RA, Moyo T, Schomaker M, Abrahams F, Grau Pujol B, Dorfman JR: **Anti-V3/Glycan and Anti-MPER Neutralizing Antibodies, but Not Anti-V2/Glycan Site Antibodies, Are**

Strongly Associated with Greater Anti-HIV-1 Neutralization Breadth and Potency. *J Virol* 2015, **89**(10):5264-5275.

624. Seaman MS, Janes H, Hawkins N, Grandpre LE, Devoy C, Giri A, Coffey RT, Harris L, Wood B, Daniels MG *et al*: **Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies**. *J Virol* 2010, **84**(3):1439-1452.