HIV-1 GENETIC DIVERSITY IN NAIROBI FEMALE SEX WORKERS

ABISI HELLEN KIVANDI

H56/20125/2019

Thesis submitted in partial fulfillment for the degree of Master of Science in Biochemistry in the department of Biochemistry in the University of Nairobi

APRIL 2023

Dedication

This thesis is dedicated to my parents, Mr. Solomon Abisi and my late mother Bishop. Mary Abisi. Thank you for your unconditional love and unwavering support at every stage in my life.

Proverbs 23:24-25
Parents rejoice when their children turn out well; wise children become proud parents. So, make your father happy! Make your mother proud!

The most beautiful thing in this world is to see your parents smiling and knowing that you are reason behind it. Without the inspiration, drive, and support that you have given me, I might not be person I am today. I vividly remember the words of my mother on her death bed, she said, “Kitinda (lastborn), make me proud!” I am so happy to have made both you and dad proud. I know you are smiling down from heaven because of this milestone, my guardian angel.

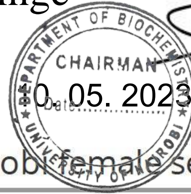
Mom and Dad, there are no words that could capture the gratitude I have for the sacrifices and dreams you let go off to give me a shot at achieving mine. I promise to achieve all the life dreams you couldn’t because you were too busy ensuring I achieved mine. Thank you for all your sacrifices. Thank you for your guidance. Thank you for your light.

I love you.

Forever and always.

SIMILARITY INDEX

Turnitin plagiarism check performed on 03/05/2023
by Dr. Martin Mulinge



HIV-1 genetic diversity in Nairobi female sex workers

ORIGINALITY REPORT

14%	8%	12%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

#	Source	%
1	Viral Genome Replication, 2009. Publication	<1%
2	"National Institute of Allergy and Infectious Diseases, NIH", Springer Science and Business Media LLC, 2009 Publication	<1%
3	www.prn.org Internet Source	<1%
4	Muthukumar Balasubramaniam, Eric O. Freed. "New Insights into HIV Assembly and Trafficking", Physiology, 2011 Publication	<1%
5	hivinsite.ucsf.edu Internet Source	<1%
6	Global Virology I - Identifying and Investigating Viral Diseases, 2015. Publication	<1%
7	www.karger.com Internet Source	<1%

scholarworks.gsu.edu

DECLARATON OF ORIGINALITY

UNIVERSITY OF NAIROBI

DECLARATION OF ORIGINALITY FORM

This form must be completed and signed for all works submitted to the University for examination.

Name of Student: ABISI HELLEN KIVANDI

Registration Number: H56/20125/2019

College: BIOLOGICAL SCIENCES

Faculty/ School/ Institute: FACULTY OF SCIENCE AND TECHNOLOGY

Department: BIOCHEMISTRY

Course Name: MASTER OF SCIENCE IN BIOCHEMISTRY

Title of the work

HIV-1 GENETIC DIVERSITY IN NAIROBI FEMALE SEX WORKERS

DECLARATION

1. I understand what Plagiarism is and I am aware of the University's policy in this regard
2. I declare that this THESIS (Thesis, project, essay, assignment paper, report. etc) is my original work and has not been submitted elsewhere for examination, award of a degree or publication. Where other people's work, or my own work has been used, this has properly been acknowledged and referenced in accordance with the University of Nairobi's requirements.
3. I have not sought or used the services of any professional agencies to produce this work
4. I have not allowed, and shall not allow anyone to copy my work with the intention of passing it off as his/her own work

5. I understand that any false claim in respect of this work shall result in disciplinary action, in accordance with University Plagiarism Policy.

Signature: ………………………………………………………………………………

Date: 04/05/2023……………………………………………………………………

Acknowledgement

First and foremost, praises and thanks to God, for the gift of life, good health, and provision. His good spirit has led me on level ground with wisdom, knowledge, understanding, counsel and might.

I am extremely grateful to my very able supervisors Dr. Martin Mulinge and Dr. Angeline Chepchirchir for their invaluable advice, continuous support, and patience at every stage of this project. Your immense knowledge and plentiful experience have encouraged me during this period. I am especially thankful for Dr. Mulinge's tenacious and resilient spirit, and brilliant mind.

I am deeply thankful to my family for all the love, support, and prayers. I am particularly grateful for my late mother who instilled discipline and other virtues in me that have been extremely valuable in my life and studies. My Godfather, Mr. Diah who always reminded that I am capable of more than I give myself credit for.

I would like to also thank my friends and colleagues for their unwavering support and belief in me throughout my studies. Thank you, Leon, for holding my hand in the lab and encouraging me even when there no bands on the gels. I am grateful for my friends and fellow graduate students Shamim, Jeff and Miriam. We have been each other's source of strength and encouragement even on very bleak days.

My gratitude also extends to the department of Biochemistry for all the support at every stage of this degree, the Partnership for Health Research Training in Kenya (P-HERT) program for the funding opportunity and UANDISHI program for the support and training in manuscript writing.

Contents

Acknowledgement 6
List of figures 11
List of tables 12
Abbreviation list 13
Abstract 15
Background: 15
Methods: 15
Results 15
Conclusion 15
Chapter 1 16
1.0 Introduction 16
Chapter 2 20
2.0 Literature review 20
2.1 Origin of HIV 21
2.2 Classification of HIV 23
2.3 Structure of the HIV-1 genome 28
2.3.1 Structural proteins 29
2.3.1.1 Gag 29
2.3.3.2 Gag-Pol Precursor 32
2.3.3.3 Protease (PR) 32
2.2.3.4 Reverse Transcriptase (RT) 33
2.2.4.5 Integrase (IN) 35
2.2.3.6 Env 35
2.3.2 Regulatory proteins 41
2.3.2.1 Tat (Transactivator of transcription) 41
2.3.2.2 Rev (Regulator of Expression of the Virion) 42
2.3.3 Accessory proteins 43
2.3.3.1 Viral Protein U (Vpu) 43
2.3.3.2 Viral Protein R (Vpr) 44
2.3.3.3 Virus infectivity factor (Vif) 45
2.3.3.4 Negative regulating factor (Nef) 45
2.4 HIV Transmission 47

2.5 Clinical staging of HIV infection 49
2.5.1 Acute phase 49
2.5.2 Asymptomatic phase 50
2.5.3 AIDS phase 51
2.6 HIV pathogenesis 52
2.6.1 Fusion and entry 52
2.6.2 Reverse transcription and nuclear import 53
2.6.3 Integration 55
2.6.4 Transcription and translation 57
2.6.5 Assembly, budding, and maturation 58
2.6.5.1 Viral Assembly 58
2.6.5.2 Viral Budding 59
2.6.5.3 Viral Maturation 59
2.7 Cellular receptors interacting with HIV env. 61
2.7.1 CD4 receptor 62
2.7.2 Coreceptors 63
2.7.2.1 Chemokine coreceptor-5 (CCR5) 65
2.7.2.2 CXCR4 68
2.8 HIV-1 tropism determination 71
2.8.1 Phenotypic assays 71
2.8.1.1 The Trofile Assay 73
2.8.1.2 Virco Tropism Assay 74
2.8.1.3 Phenoscript Env Assay 74
2.8.1.4 MT-2 Assay 74
2.8.2 Genotypic *in-silico* prediction algorithms 76
2.8.2.1 Net charge rule 77
2.8.2.2 ‘11/25 or 11/24/25’ Rule 77
2.8.2.3 Position specific scoring matrices (PSSM) 78
2.8.2.4 Geno2Pheno 78
2.8.2.5 HIV CoR 79
Chapter 3 80
3.0 Methodology 80

3.1 Problem statement 81
3.2 Justification 82
3.3 Study objectives 83
3.3.1 Study Questions 83
3.3.2 Overall objective 83
3.3.3 Specific objectives 83
3.3.4 Research hypothesis 83
3.4 Materials and methods 84
3.4.1 Study Design, setting, and population. 84
3.4.2 Study setting and population 84
3.4.3 Study material 84
3.4.4 RNA extraction 85
3.4.5 Amplification of Env gene 86
3.4.5.1 Reverse transcription 86
3.4.5.2 Nested PCR 86
3.4.6 V3 loop sequencing 87
3.4.6.1 PCR product purification 87
3.4.6.2 Cycle sequencing 87
3.4.6.3 Purification of cycle sequencing products with Big Dye XTerminator™ 88
3.4.6.4 Sequence detection 88
3.4.7 HIV-1 Subtyping 88
3.4.8 Determination of Viral tropism 88
3.4.9 Phylogenetic tree construction 89
3.4.10 Statistical analyses 90
3.5 Results 91
3.5.1 Sample amplification 91
3.5.2 Patient characteristics 91
3.5.3 HIV-1 Subtypes 92
3.5.4 Concordance of the HIV-1 subtyping tools 92
3.5.5 Viral tropism prediction 93
3.5.6 Concordance of the genotypic tropism testing tools 95
3.5.7 Tropism in relation to host and viral characteristics 96

3.5.8 V3 loop sequence diversity 99
3.5.9 Phylogenetic tree analyses 102
Chapter 4 104
4.0 Discussion 104
4.1 Viral tropism prediction 105
4.2 Concordance of tropism prediction algorithms 106
4.3 HIV-1 subtypes 107
4.4 Concordance of HIV-1 subtyping tools 107
4.5 Predictors of coreceptor usage 108
4.5.1 HIV-1 subtype 108
4.5.2 CD4 T cell count 108
4.5.3 SWOP clinic location 109
4.5.4 Features of the env V3 loop 110
4.6 Degree of relatedness of the viruses 110
4.7 Study limitations and recommendations 111
4.7.1 Study limitations 111
4.7.2 Study recommendations 112
4.8 Conclusion 113
4.9 Appendices 114
4.9.1 Waiver of consent application 115
4.9.2 Ethical approval for nested study 121
4.9.3 Consent to futuristic use of collected sample 123
References 124

List of figures

Figure 1: Worldwide distribution of HIV-1 subtypes. 23
Figure 2: Classification of HIV-1 groups and subtypes 25
Figure 3: Structure and organization of the HIV-1 genome. 28
Figure 4 Linear arrangement of the HIV-1 Gag polyprotein precursor Pr55Gag 30
Figure 5: Structure of the V3 loop. 39
Figure 6: HIV-1 Tat protein. 41
Figure 7: An overview of HIV pathogenesis 52
Figure 8: Viral fusion and entry. 53
Figure 9: Importation of the PIC into the nucleus and the selection of the integration sites. 54
Figure 10: Mechanism of proviral DNA integration into host genome 56
Figure 11: Cryoelectron micrograph and graphic representation of immature and mature HIV-1 particles. 60
Figure 12: Schematic representation of CXCR4 64
Figure 13: Sample amplification rate 91
Figure 14: Comparisons of (a) median baseline CD4 + T cell counts and CD4 and (b) median baseline T cell percentages between CCR5 and CXCR4 tropisms. (c) Comparison of CD4 T count in subjects infected with different HIV-1 subtypes. 99
Figure 15: Sequence logos of CXC4 and CCR5 tropic HIV-1 V3 loop data sets. 100
Figure 16; Schema of the phylogenetic tree of HIV-1 viruses circulating in this cohort of treatment naïve commercial female sex workers. In red are CXCR4 tropic variants from different SWOP clinic locations, distributed throughout Nairobi. 103

List of tables

Table 1: Reverse transcription PCR reaction mix set up ... 86
Table 2: Nested PCR reaction mix set up ... 86
Table 3: PCR conditions for amplicon clean-up ... 87
Table 4: Cycle sequencing reaction mix set up ... 87
Table 5: PCR conditions for Cycle sequencing product cleanup ... 88
Table 6: HIV-1 subtypes assigned by the different subtyping tools and their reliability calculated using Cronbach's alpha coefficient. (n=76) ... 92
Table 7:The distribution by number and (proportions), of viral tropism across COMET based HIV-1 subtypes using Geno2Pheno at four cut off points. P-value is derived from cross-tabulation of coreceptor tropism and virus subtype using Pearson's Chi-square test. ... 93
Table 8: The distribution by number and (proportions), of viral tropism across COMET based HIV-1 subtypes using three WebPSSM matrices. P-value is derived from cross-tabulation of coreceptor tropism and virus subtype using Pearson's Chi-square test. [a]WebPSSM [SINSI -C] matrix. ... 94
Table 9: Distribution of viral tropism phenotypes and variation in tropism prediction reliability when an algorithm is dropped. ... 96
Table 10: HIV-1 subtype and viral tropism distribution ... 96
Table 11: Demographic and clinical characteristics of 76 study subjects included in the analysis. Results are median (IQR) or frequency (%), as appropriate. Correlations were calculated using Pearson's Chi-Square test. ... 97
Table 12: Factors associated with coreceptor usage in 76 HIV-1 treatment naïve subjects ... 101

Abbreviation list

APOBEC3- Apolipoprotein B Editing Complex 3

CPE- Cytopathic effect

CPSF6- Cleavage and adenylation specificity factor 6

CRM1- exportin 1

CTD- Carboxyl terminal domain

CTL- Cytotoxic T cell

CYPA -Cyclophilin A

DC-SIGN- Dendritic cell (**DC**)-specific **HIV**-1 receptor

DISF- Sensitivity inducing factor

ECL2- Extracellular loop 2

ESCRT- Cellular endosomal sorting complex required for transport

ESTA- Enhanced sensitivity Trofile assay

FPR- False positive rate

GALT- Gut associated lymphoid tissue

GPCR- G-protein coupled receptor

GTT- Genotypic tropism testing

LEDGF- Lens-epithelium-derived growth factor

LESTR- Leukocyte-derived seven transmembrane domain receptor

LTR- Long terminal repeat

Mre11- Double stranded break repair protein 11

MRN complex- Mre11, Rad50, NBSI complex

NAbs- Neutralizing antibodies

NACC- National AIDS control council of Kenya

NASCOP- National AIDS and STIs control program

NBSI- Nibrin

NC- Nucleocapsid protein

NFAT- Nuclear factor of activated T cells

NELF- Negative elongation factor

NES- Nuclear export signal

NSI- Non syncytium inducing NEF- Negative regulating factor

ORF- Open reading frame

PBMC- Peripheral blood mononuclear cells

PBS- Primer binding site

PIC- Pre-integration complex

PND- Principal neutralization determinant

PPT- Polypurine tract

Pro- Protease

PSSM- Position specific scoring matrices

PTEF-13- Positive elongation factor 13

PTT- Phenotypic tropism testing

RNAP II- RNA polymerase II

RANTES- Regulated upon Activation, Normal T Cell Expressed and Presumably Secreted chemokine

REV- Regulator of expression of the virion

RRE- Rev response element

RNase H- Ribonuclease H

RT- Reverse transcriptase

R5X4- Dual/mixed variants

SI- Syncytium inducing

SP1- Spacer protein 1

SP2- Spacer protein 2

SWOP-Sex worker outreach program

TAT- Transactivator of transcription

TGF-β – Transforming growth factor beta

TNF-α- Tumor necrosis factor alpha

TNPO3- Transportin 3

TRIM5- Retrovirus restriction factor

5’UTR- 5’-Untranslated region

Abstract

Background: Sub-Saharan Africa bears 67% of the global HIV-1 infection burden, with more than half of the infections occurring in women above 15 years. The risk of contracting HIV is 30 times higher for female sex workers compared to non-FSW. This group, therefore, bears a double burden of the HIV epidemic.

Methods: We assessed HIV tropism and the variables associated with it in recently infected, treatment-naïve female sex workers in Nairobi, Kenya. We analyzed 76 HIV-1 positive plasma samples obtained between November 2020 and April 2021 from female sex workers accessing clinical services at seven Sex Worker Outreach Programme (SWOP) clinics in Nairobi, Kenya. The primary outcome was viral tropism predicted by genotypic algorithms targeting the V3 loop of the env gene. Logistic regression was used to determine the effects of age, HIV-subtype, CD4 count, CD4%, and V3 loop genetic features on the primary outcome.

Results:The prevalence of CXCR4-tropic viruses was 26.3%. HIV-1 subtype A1 accounted for 89.5% of subtypes followed by subtype D (6.6%) and subtype C (3.9%). WebPSSM [WebPSSM:R5X4] and Geno2Pheno [G2P:10% FPR, G2P:15% FPR and G2P:20% FPR] were highly concordant at 88% (95% CI 82% –91%) .A unit increase in V3 loop's net-charge increased the odds of a virus being CXCR4-tropic 75% (OR = 1.75, 95% CI = 1.20-2.70, p = 0.006). The odds of Subjects from Kawangware harboring CXCR4 tropic strains were 256% higher (OR 3.56; 95% CI 0.97-14.7; $p = 0.034$) compared to the reference group (Thika Road). HIV-1 subtype, age, CD4 count, and CD4 % were not associated with viral tropism.

Conclusion

We found a relatively higher prevalence of X4 viruses in HIV-1 subtype A1 than previously thought to be present. The most notable finding was the significant association between CCR5 tropism A1 viruses and the presence of the amino acid Alanine at position 22 of the V3 loop. Additional studies with a large dataset are warranted to confirm our findings.

Chapter 1

1.0 Introduction

HIV-1 is still a major burden to public health and continues to impart devasting effects worldwide. 37.7 million people were reported to be living with HIV in 2021, with Sub-Saharan Africa bearing 67% of the infection burden. Subtype C is the most dominant group M lineage accounting for nearly half of the global HIV-1 infections. Subtype A and its CRFs(circulating recombinant forms) come in second at 25% and predominate in East Africa (Digban et al., 2020; Gounder et al., 2017). Subtypes A1, D, and AD are the most prevalent in Kenya, while subtypes A2, C, and G constitute the minority subtypes (Kitawi et al., 2017; Lihana et al., 2009; Lwembe et al., 2009; Mabeya et al., 2018).

A hallmark of the HIV-1 virus is the vast genetic variability attributed to reverse transcriptase-induced errors and host immune selection pressures. This diversity is manifested as a sequence variability, particularly within the *env* variable regions. Env is a glycoprotein on the virion surface that mediates viral entry into host target cells. HIV entry into the host target cell is a sequential multistep process facilitated by the viral env, and host CD4 receptor in concert with either the CCR5 or CXCR4 coreceptors (De Jong et al., 1992). The V3 loop is the principal determinant of viral tropism since it participates directly in viral entry by interacting with the host chemokine coreceptors (Hwang et al., 1991). Viruses that utilize the CCR5 coreceptor are termed R5 variants, those utilizing the CXCR4 coreceptor are known as X4 variants, while dual/mixed tropism (R5X4) variants can utilize either of the coreceptors for entry into the target cell (Berger et al., 1998). CCR5 is the principal coreceptor used for entry by a majority of HIV strains and predominates early in infection (De Jong et al., 1992), it is also the principal coreceptor expressed by sexually transmitted HIV-1 variants (Dragic T et al., 1996; Kawamura et al., 2003) fueling the development of entry inhibitors such as the CCR5 antagonist, Maraviroc among others (Dorr et al., 2005). The selective nature of this entry inhibitor requires viral tropism testing to establish coreceptor usage (Gilliam et al., 2010; Rao, 2009; Seto et al., 2006). Despite the approval of this salvage therapy, the scarcity of data on coreceptor tropism partly limits the use of this class of ART (antiretroviral therapy) in Kenya. HIV-1 tropism can be determined using phenotypic tropism testing (PTT) assays that rely on sequencing of the envelope gp120 or genotypic techniques that rely on the amplification of the V3 loop (Cabral et al., 2012; Jensen A. Mark, Wout Angelique, 2003) PTT assays such as the Trofile, Toulouse test, MT-2 assay, and Virco amongst others, utilize plasma or peripheral blood mononuclear cells peripheral blood mononuclear cells (PBMC)-derived viruses (Sing et al., 2007; Skrabal et al., 2007). Pseudoviruses or recombinant viruses are then

constructed using the full or partial *env* (gp160) amplicons. Modified CD4/CCR5 or CD4/CXCR4 expressing reporter cell lines -mainly U87 and U373- are then transfected with viruses and a signal is subsequently measured following entry into one or either cell type (Reeves et al., 2002).

Although PTT assays are considered the gold standard in tropism determination, limitations such as the cost implications, turnaround time, and technical demanding nature of these assays prompted the development of the less costly, faster, and user-friendly *in silico* prediction methods (Garrido et al., 2008; Poveda et al., 2009; Recordon-Pinson et al., 2010). These algorithms utilize the 11/25 rule, the sum of positively charged amino acids within the V3 loop, the universal charge of the V3 loop to more composite statistical models such as the position-specific scoring matrices (PSSM), support vector machines, or a blend of the different algorithms, which has been shown to yield more accurate results (Low et al., 2007). There is currently no consensus on the tool of choice for tropism prediction. Other than Geno2Pheno 10% which is well validated and recommended by European virologists for clinical tropism testing (Vandekerckhove et al., 2011), several widely validated subtype-specific tools have been developed. Some of these algorithms such as WebPSSM$_{[SINSI-C]}$, HIV-1 subtype C specific V3 sequence based coreceptor usage prediction algorithm-CoRSeqV3-C (Cashin et al., 2013) were trained for HIV-1C. Other GTT tools such as Subtype A coreceptor tropism classification in HIV-1-SCOTCH (Lö Chel et al., 2018), Toulouse HIV Extended Tropism Algorithm-THETA (DImeglio et al., 2020), PhenoSeq (Cashin et al., 2015), and HIVCoR (Hongjaisee et al., 2019) were optimized for non-B subtypes. Since PTT is considered the gold standard for assigning coreceptor tropism, the accuracy and efficiency of the algorithms are usually validated against PTT assays.

A hallmark of HIV-1 transmission is the existence of numerous transmission clusters that are crucial in the sustenance of the epidemic, and should thus be prioritized for HIV prevention (Kouyos et al., 2010a; Petersen et al., 2018). In Kenya, the HIV infection rate is disproportionately high in marginalized populations such as female sex workers (FSWs), men who have sex with men (MSM) and injecting drug users (IDUs). Approximately 5.6% of the Kenyan population are living with HIV-1, and the prevalence is 3X higher among so-called high-risk groups ((NASCOP), 2018). FSWs are the most vulnerable of the key population groups. Prevalence estimates are 13.5-fold higher in FSWs compared to their non-sex worker counterparts (Bitty-Anderson et al., 2022). In Kenya, the estimated risk of transmission is 9.8 times higher in this group compared to non-sex

worker females (Tago et al., 2021). This highlights the burden of HIV in this key population and their central role in sustaining population-level infections. This population is also more likely to harbor multiple HIV-1 variants and recombinants due to their social vulnerability and the risks associated with their occupation including multiple sexual partners, inconsistent condom use, use of injecting and non-injecting drugs, and co-infections with other STIs (Baral et al., 2012; Prakash et al., 2018; Tago et al., 2021). It is still not clear how virally encoded factors and host immune status influence HIV transmission in such a key population. Monitoring HIV-1 in this key population is therefore crucial to understanding the pandemic's trajectory and formulating effective preventative and control measures. We sought to evaluate subtype diversity and coreceptor tropism in a cohort of treatment naïve FSWs from seven SWOP clinics located throughout Nairobi. We additionally evaluated the performance of the HIV subtyping, and tropism prediction algorithms, and identified tropism determinants using a multivariate logistic regression model.

Chapter 2

2.0 Literature review

2.1 Origin of HIV

Discovered in 1981, Human Immunodeficiency Virus (HIV) is the root of one of humanity's most lethal and persistent epidemics. Acquired Immune Deficiency Syndrome (AIDS) was initially recognized as a new illness when a growing number of youthful, gay men died from uncommon malignancies and atypical opportunistic infections (CDC, 1981). Luc Montagnier in conjunction with a research team at the Paris Pasteur Institute led the discovery of the HIV-1 virus in 1983. The team used a lymph node biopsy obtained from a homosexual patient of French nationality who exhibited manifestations that preceded the onset of AIDS, such as lymphadenopathy. Observation of the virion morphology as well as RT activity in the supernatant culture implied that a retrovirus was present. The team's subsequent attempts to infect different types of cells including B cells, and fibroblast was futile. They however successfully infected T cells, implying that the patient was infected with a retrovirus with a preference for T cells, but a connection with AIDS was still speculative.

The discovery of a human lentivirus as the causative agent of AIDS occurred two years post AIDS discovery. In 1984, a team of scientists led by Robert Gallo at the National Cancer Institute in Bethesda, Maryland, isolated HIV-1 from a bigger batch of patients and proposed that the HIV virus was a causative agent of AIDS. In the same year, the team discovered a permissive T cell for HIV-1 that enabled continued propagation of patient-isolated viruses, allowing for the production of the virus in larger quantities to facilitate further studies. The association between AIDS and HIV-1 was further corroborated by a third group of scientists led by Levy and a team from the California Department of Health Sciences in Berkeley using similar techniques as the previous groups.

Analysis of the available phylogenetic data estimates the emergence of both viruses to 1930 (± 20 years) with the origin of the pandemic being traced to the Democratic Republic of Congo (Korber et al., 2000; Worobey et al., 2008). The emergence of both strains was attributed to numerous transmissions of the simian immunodeficiency virus (SIV) across species from non-human primates to humans (Korber et al., 2000). According to (Hebb et al., 1999), zoonotic transmission of the primate Lentiviruses that defines simian reservoirs for the human viruses has been authenticated by several lines of evidence such as; similarities in the organization of the viral genomes, phylogenetic relatedness to studied SIVs, coincidence in geographic distribution, the prevalence in the natural host range and cogent transmission routes.

HIV-1 is most analogous to the SIVcpz, a recombinant virus discovered in the *Pan troglodytes troglodytes* chimpanzee subspecies. The *gag* and *pol* genes of the SIVcpz are built up from the SIV of the red-capped mangabey monkeys (*Cercocebus torquatus).* Its *env* gene is a recombinant of the SIVgsn or SIVmon of the greater spot monkey (*Cercopithecus nicticans*) or mona monkey (*Cercopithecus mona*), respectively (Sharp and Hahn, 2011).

The most diverse forms of HIV-1 are all located in distinct geographic regions correlated to the range of *P.t troglodytes* in West Equatorial Africa. The interspersed distance between HIV clades and SIVcpz sequences in phylogenetic trees also suggests shared viral lineages (Robertson et al., 2000). The existence of a HIV-like retrovirus in 5% of *P.t troglodytes* suggests a non-human primate as the most likely reservoir of HIV-1 (Eberle and Gürtler, 2012). However, the origin of HIV-1 remains uncertain as three such SIVcpz infections have been recorded, of which one is so disparate that it is possibly representative of a distinct lineage of primate Lentiviruses (Groen et al., 2000; Janssens et al., 1994). HIV-2 is mostly restricted to West Africa and studies have identified the sooty mangabey, Cercocebus atys (green monkey) as its primate reservoir.

2.2 Classification of HIV

HIV is phylogenetically classified into HIV-type 1 and HIV-type 2 based on genetic variations in viral antigens (Clavel et al., 1986; Hebb et al., 1999). HIV-1 is most widespread and has wider genetic variability (Berger et al., 1998), which may have an impact on facets of its biology such as infectivity, transmissibility, and immunogenicity (Robertson et al., 2000). HIV-2 is relatively less common and predominantly confined to West Africa (Hirsch et al., 1989).

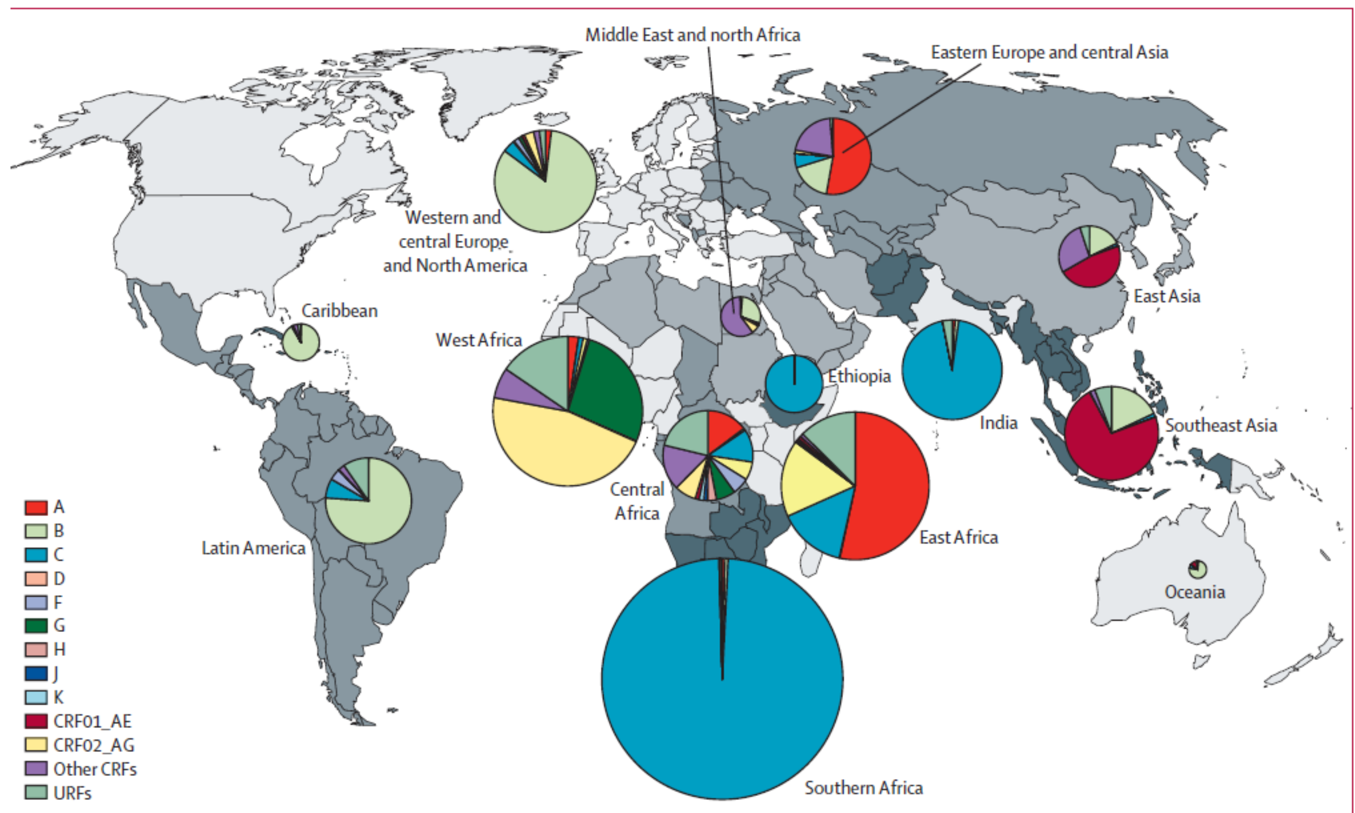


Figure 1: Worldwide distribution of HIV-1 subtypes.
The countries are color-coded based on the most prevalent M subtype. The pie charts show the percentage of each subtype in a selected geographical region. Pie chart sizes are representative of the number of infected individuals in a continent with respect to the total HIV-1 infections globally. *Adopted from (Hemelaar et al., 2019)***.**

HIV-2 is also relatively less efficient in vertical and horizontal transmission and exhibits a slower disease progression rate (Kanki et al., 1994; Marlink et al., 1994). Both HIV types share a common ancestor, and sequence homology of 58%, 59%, and 39% in the *gag, pol,* and *env* genes, respectively (Clavel et al., 1986). HIV-1 is made up of four groups namely M, N, O, and P with

group M being the most predominant worldwide, comprising ten subtypes, several sub-subtypes, and several recombinant forms (Désiré et al., 2018). The different groups correspond with discrete lineages individually introduced from non-human primates into the human population, while the subtypes and sub-subtypes stemmed from post-introduction by founder events, and additional diversification (Robertson et al., 2000).

The first proposed classification of HIV strains was geographic, with strains from Europe/North America being clustered into one group while those from Africa clustered into another. This system was however found to be limiting following analysis additional of sequences derived from other geographic areas that revealed the existence of multiple phylogenetic clusters. Phylogenetic analysis of these envelope sequences revealed clades that were roughly equidistant from each other. These clades were designated as subtypes A to F with the 'North American' strains being relabeled as subtype B (Myers et al., 1992). Subsequent inference from sequences of the gag region enabled the identification of five subtypes (A, B, C, D, and F, but not subtype E) of the six env-based clades (Louwagie et al., 1993). Four more subtypes, G to J, were characterized following analysis and comparison of partial sequences using phylogenetic tools (Leitner et al., 1995; Robbins et al., 1999). Phylogenetic comparisons based on gag and env showed that subtype F was further made up of sub-subtypes F1, F2, and F3. Sub-subtype F3 was subsequently renamed subtype K following analysis of the complete genome (Triques et al., 2000).

More recently, a new subtype denoted L has been reported (Yamaguchi et al., 2019). Two divergent strains collected in DRC in 1983 and 1990 had been putatively labeled as group M subtype L, but there was a lack of a third subtype L genome which was epidemiologically different to assign L as a true subtype, as stipulated by the current HIV nomenclature guidelines. A 13.2%–14.5% nucleotide variance of each of these genomes from all the other group M clades further supported the existence of subtype L (Carr et al., 2002).

The classification of HIV-1 viruses into different subtypes considers all regions of this genome although it was previously dependent on the *gag* and or *env* sequences. Genetic variation within subtypes ranges from 15-20% whereas inter-subtype variation stands at 25-35% depending on the subtype and region of the genome under consideration (Korber et al., 2001). All these HIV-1 subtype groups (A-D, F-H, and J-L) collectively form a clade or cluster now denoted group M for 'main' to differentiate it from the other HIV-1 groups namely group O (outlier/outgroup) and the

HIV-1 group N (non-M/non-O) and P (Eberle and Gürtler, 2012). Group N is not as widespread as group M and is mainly found in Cameroon. Group O accounts for 1% of the total HIV infections and is mostly confined to Cameroon and Gabon. Group P is rare but was identified in a pregnant woman of Cameroonian origin and France (Bhatti et al., 2016).

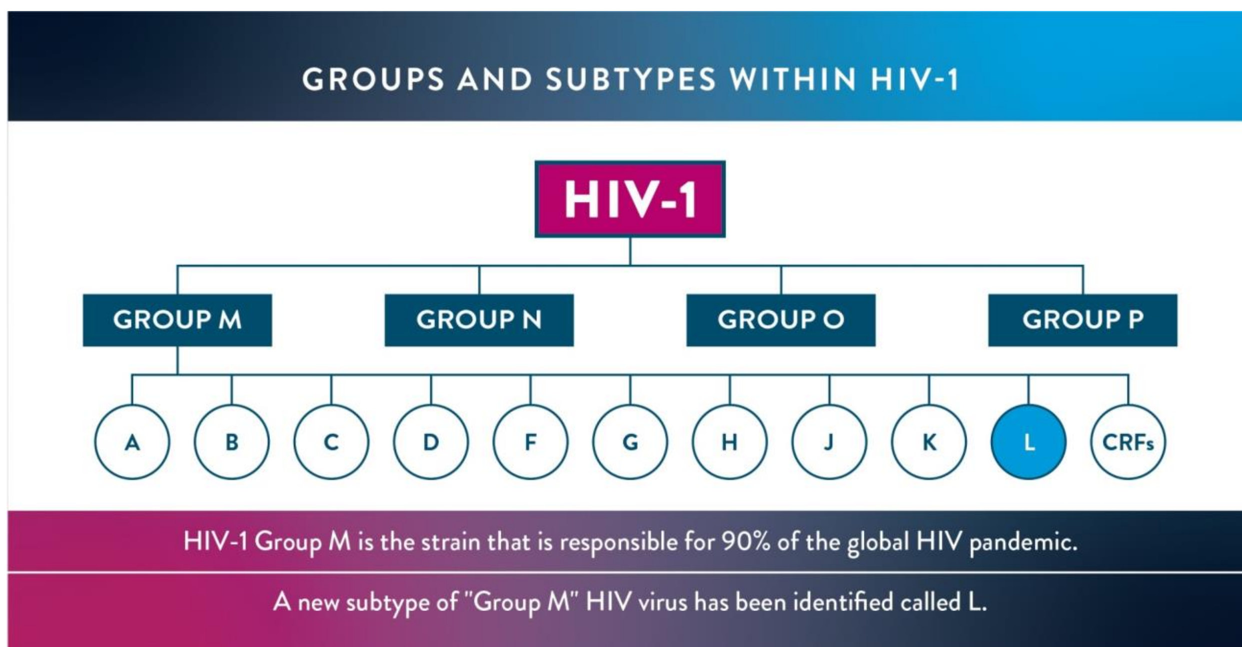


Figure 2: Classification of HIV-1 groups and subtypes
Adopted from Abbott Global Surveillance Program

Most HIV-1 strains consistently cluster into the same group or subtype irrespective of which part of their genome is analyzed. However, some of these strains were shown to exhibit discordant clustering following analysis of distinct genome parts (Li et al., 1998). This observation and the additional co-circulation of these divergent subtypes in the same geographic regions suggested that the discordant strains resulted from recombination events. Recombination is a relatively common phenomenon between groups (between HIV-1 groups M and O), inter-subtype (most common), and intra-strain within different HIV-1 group M, subtypes (Rousseau et al., 2007). Recombination occurs when there is swapping of whole gene sequences at unselected locations, following target cell infection with two discrete HIV species resulting in circulating recombinant forms (CRFs) (Robertson et al., 1995). This event occurs during plus or minus strand synthesis through three

major processes. The first mechanism involves the highly erroneous RT with reduced binding affinity for facilitating strand transfer producing extensively heterogenous, recombinant genomes. The second and third mechanisms involve the selection and functional and replicable forms, respectively to ensure survival and transmission from host to host (Lau and Wong, 2013). Recombination of HIV-1 and HIV-2 has not yet been detected but has been reported as feasible in vitro. Patients doubly infected with both types have been reported. However, infection with one type does not induce super immunity to prevent superinfection with another type (Eberle and Gürtler, 2012).

According to the HIV nomenclature rules by (Robertson et al., 2000), members of a CRF must possess a homogenous mosaic structure in that they are descendants of duplicate recombination event(s). In addition, the virus strain must be detected in at least three epidemiologically unrelated individuals and the strain should be able to initiate an epidemic on its own. According to the Los Alamos Laboratory guidelines on the nomenclature of HIV and SIV, CRFs are labeled by their number defined according to the order in which they were discovered, and the subtypes involved. For example, CRF_02 AG is the second identified recombinant of subtypes A and G. CRFs may also be named using their numbers and the letters 'cpx' which denotes complexity when more than two subtypes are involved e.g., CRF04_cpx or CFR06_cpx. CRFs that fail to meet these criteria are classified as unique recombinant forms (URFs). URFs are identified in a single individual or in a cluster of associated individuals and remain a localized phenomenon (Berger et al., 1998; Robertson et al., 2000). However, the onward transmission of URFs in the population can give rise to CRFs that subsequently spread to other regions, where they act as key drivers of the epidemic. Two CRFs -CRF01_AE and CRF02_AG- were initially found in Central Africa during the early stages of the epidemic, and subsequently spread to Southern China and Southeast Asia where they are currently key drivers of the regions' HIV-1 epidemic (An et al., 2020; Arthos et al., 2008; Giuliani et al., 2013; Li et al., 2017; Palm et al., 2014; Sun et al., 2020).

More than 96 epidemiologically stable CRFs and URFs have been recorded and their current 20% global incidence is predicted to increase as they gradually replace and phase out the initial predominant HIV-1 pure subtypes accounting for up to 20% of HIV-1 infections worldwide (Hemelaar et al., 2011). The most epidemiologically noteworthy example of recombinant strains is the 'subtype E' virus which is most rampant in Thailand and the neighboring South-East Asian

countries. Phylogenetic studies on gag and env carried out by (Louwagie et al., 1993) and others, and subsequent explorations of whole subtype E genomes by (Carr et al., 2002) designated this clade as a cluster of recombinant viruses CRF01_AE, rather than as an independent subtype. This was evidenced by the subtype A radiation observed within the gag and pol regions of all 'subtype E' genomes that is representative of a lineage of recombination between subtypes A and E (Robertson et al., 2000). It has been suggested that CRF01_AE belongs to clade E since a pure subtype E genome has never been recorded (Anderson et al., 2000).

HIV-1 viruses were also grouped based on the observed growth of some HIV-1 isolates in transformed T cells. As a result, isolates were grouped into two classes: T-tropic (T cell) or M-tropic (macrophage) (Asjö et al., 1986). Some viral isolates were also observed to induce syncytia while others did not cause syncytia resulting in the classification of variants as syncytium inducing (SI) and non-syncytium-inducing (NSI), representing the T-tropic and M-tropic variants, respectively. It was evident that the propensity of the HIV-1 virus to infect a particular target cell in vitro, could not be sufficiently explained by CD4 expression only. Coreceptor discovery by (Feng et al., 1996) and others prompted the classification of individual viruses based on their propensity to utilize the CCR5 (R5 variants), CXCR4 (R4 variants), and both (dual/mixed, R4X5 variants) to enter the target cells (Berger et al., 1998, 1999). The X4 variants became the equivalent of SI viruses and R5 variants the equivalent of NSI/M-tropic viruses, owing to the lack of CCR5 expression on most transformed CD4 T cell lines.

R5 viruses were initially assumed to be M-tropic and non-syncytium inducing (NSI). However, this grouping was flawed because it only accounted for a tiny, however important fraction of R5 viruses that were solely M-tropic. Since coreceptor utilization is not the exclusive viral tropism determinant, current groupings are as follows: R5 T-tropic, R5 M-tropic, X4, and R5X4 (Robertson et al., 2000). Well-characterized prototypic HIV-1 strains were used to describe the preliminary coreceptor activities. All primary HIV-1 isolates were observed to utilize either or both the coreceptors in the subsequent studies (Connor et al., 1997; Zhang et al., 1998).

2.3 Structure of the HIV-1 genome

HIV-1 is a retrovirus of the *Lentivirus* genus. Retroviruses alternate between two forms of genomic material, a ssRNA in the virion, and dsDNA that is integrated into the genome of the host (Coffin et al., 1997). Unlike other viruses, retroviruses such as HIV contain two genome molecules. The core of the virus encloses two identical, positive, ssRNA molecules that are capped at the 5' end, polyadenylated at the 3' end, and annealed to a host tRNA lysine molecule. Both these RNA molecules are employed in strand transfer mediated recombination reactions, that occur during reverse transcription. Thus, retroviruses are considered 'pseudodiploid' as only one allele is generated following recombination (Takeuchi et al., 1991). The dsDNA is generated by reverse transcriptase utilizing genomic RNA as a template for synthesis of the first strand, and the resulting cDNA as a template for synthesis of the second strand (van der Kuyl and Berkhout, 2012). The integrated, proviral genome of HIV-1 is approximately 9.8 kilobases in length (Watts et al., 2009).

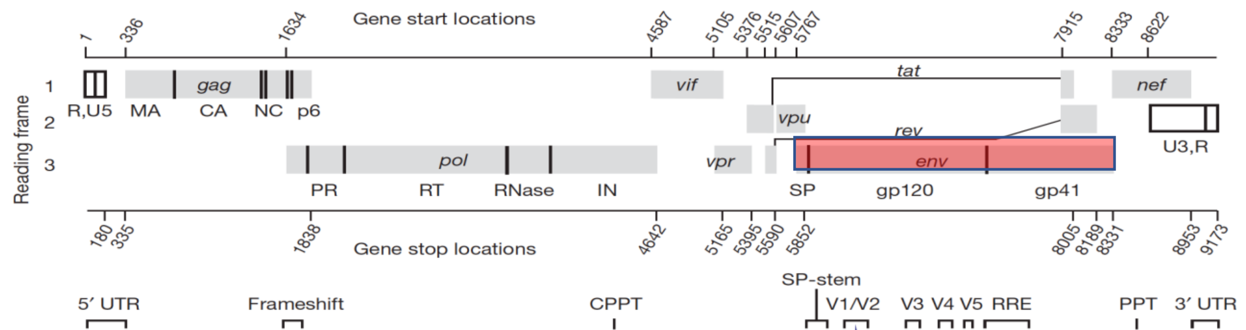


Figure 3: Structure and organization of the HIV-1 genome.
The grey boxes represent 9 exons. The env gene has been highlighted. The beginning and end sites of the genes are numbered according to NL4-3.***Adopted from (Watts et al., 2009)***

In essence, the genome of HIV-1 is a coding RNA made up of nine open reading frames that encode 15 proteins. Six of these ORFs are unique to HIV while three (5'-*gag-pol-env*-3') are common to all retroviruses. LTRs flank the genome at both ends, while the HIV-1 genes localize to the central region (Coffin et al., 1997). The promoter that facilitates viral gene expression is found at the 5' LTR region and precedes the Gag gene reading frame in the 5'-3' direction. The Gag polyprotein encoded by *Gag* undergoes processing by viral protease to yield the matrix protein (MA, p17), the capsid protein (CA, p24), the nucleocapsid protein (NC, p7), a smaller protein that stabilizes nucleic acids termed p6, as well as two spacer proteins (SP1 and SP2) (Bryant and Ratner, 1990). The *pol* ORF that encodes viral integrase, protease, and RT then follows. Adjacent

to this is the *env* ORF that codes for two envelope glycoproteins: the transmembrane protein gp41 and the surface protein gp120, and a 30 amino acid signal peptide (Coffin et al., 1997). Gene expression is achieved by three discrete mechanisms. Gag is a product of unspliced viral RNA, env a product of spliced RNA, while pol is expressed as a Gag-Pol precursor protein due to sporadic ribosomal frameshifting. This frameshifting subtly regulates gene expression by ensuring gag gene encoded structural proteins are in excess relative to enzymatic proteins encoded by the pol gene (Scarlata and Carter, 2003).

The genome contains additional structures required for replication and evasion of immune responses mounted by the host. These regulatory motifs have been identified and include packaging signals, ribosomal frameshift motifs, and cis-regulatory elements, among others. They modulate processes such as transcription activation, reverse transcription initiation, formation of the genomic RNA dimers, mediating packaging of HIV-1 particles, regulating the export of RNA from the nucleus as well as viral and host proteins interaction (Damgaard et al., 2004; Wilkinson et al., 2008). Other regulatory proteins encoded by the HIV-1 genome include Tat; a transactivator protein and Rev; an RNA splicing factor. Both these factors are required to initiate the HIV replication cycle. Accessory proteins encoded by the proviral genome include *Nef* (negative regulating factor), *Vif* (viral infectivity factor), *Vp*r (virus protein), and *Vpu* (virus protein unique). These proteins modulate viral replication, budding, and pathogenesis.

2.3.1 Structural proteins

2.3.1.1 Gag

Group-specific antigen (gag) is a 55kD protein that coordinates the assembly and subsequent release of virion particles from HIV-1 infected cells. This protein can also form virus-like particles (VLPs) *in vitro* when viral and cellular components are unavailable. All major structural proteins in retroviruses including HIV-1 are derived from the Gag precursor polyprotein (Pr55Gag), a product of unspliced viral RNA. Gag accounts for up to 50% of the viral mass, with each HIV-1 virion containing up to ~2500 gag copies (Campbell and Rein, 1999; Campbell et al., 2001; Scarlata and Carter, 2003).

Figure 4 Linear arrangement of the HIV-1 Gag polyprotein precursor Pr55Gag.
The constituent domains namely matrix (MA), capsid (CA), nucleocapsid (NC), and p6 domains are shown. The spacer peptides SP1 and SP2 are also shown. *Adopted from (Balasubramaniam and Freed, 2011).*

HIV-1 buds as a non-infectious, immature virion. Gag processing, which occurs concurrent with or shortly after budding, is the hallmark of structural viral maturation and infectivity. Gag undergoes co-translational myristoylation at the N-terminal after which viral protease-mediated Gag processing yields four functional domains; matrix (MA), capsid (CA), nucleocapsid (NC), spacer protein 1 and 2 (SP1&2), and P6, which are rearranged to form mature and infectious virions cell surface membrane during viral assembly. The myristic acid moiety fosters the anchoring of Gag to the cytoplasmic aspect of the cell surface membrane, fortifying the Gag-membrane interaction (Bryant and Ratner, 1990). This myristic acid moiety functions in concert with an extremely basic motif in the MA that binds the phosphatidyl inositol 4,5-bisphosphate phospholipid in the cytoplasmic aspect of the plasma membrane (Balasubramaniam and Freed, 2011). In this way, Gag steers the assembly of HIV-1 virus particle into a hexameric protein lattice at the cell surface membrane (Briggs et al., 2009), or in intracellular compartments such as the late endosome or multivesicular bodies as observed in macrophages (Perlman and Resh, 2006). The membrane-attached Gag polyprotein then assembles two copies of the viral genomic RNA, several other RNAs as well as cellular and viral proteins to form a complex which triggers the budding of the viral particle (Bryant and Ratner, 1990; Saad et al., 2006).

Gag is also involved in nuclear shuttling. A subgroup of these proteins is recruited into deeper layers of the virion particle, where they become members of an intricate translocation multiplex that ushers viral DNA into the nucleus. Cellular import machinery recognizes a karyophilic signal on the MA, enabling transient Gag shuttling through the nucleus before the assembly of a HIV-1 particle (Dupont et al., 1999; Marchetti et al., 2013)

The capsid plays varied roles in the replication cycle of HIV-1. During viral assembly, the CA domains drive Gag-Gag multimerization at the virus assembly sites. In the mature virion, the

conical core that encloses the genomic viral RNA complexed with the NC proteins and viral enzymes is made up of CA monomers (Balasubramaniam and Freed, 2011). Investigation into the CA regions involved in HIV assembly and release mapped a Pro-rich loop that is crucial for immature Gag lattice generation. Non-infectious virions with abnormal core structures and defective reverse transcription initiation were observed when mutations were introduced to this region (Chen et al., 2008; Novikova et al., 2018; Tang et al., 2001). The capsid is also responsible for recruiting Cyclophilin A into virions. Cyclophilin A is a cytosolic host protein that catalyzes the conversion of proline residues from cis to trans isomers, regulating gene espression. This protein is also the drug target for the immunosuppressant-cyclosporin. Cyclophilin A enhances viral infectivity by interacting with a pro-rich region at the N-terminus of CA (Bosco et al., 2002; Dietrich et al., 2001; Scarlata and Carter, 2003).

The NC is a 55aa protein that functions as a chaperone during the folding of genomic RNA. The main function of this protein is to encapsulate and protect viral RNA. NC modulates the inclusion of a heterologous RNA into the HIV-1 virion through recognition of the HIV packaging signal known as Psi (Ψ)- a hairpin loop proximal to the 5' UTR end of the viral RNA. Contact between NC and psi is modulated by two zinc-finger motifs on NC (Lapadat-tapolsky et al., 1993). Psi is 110 nucleotides long and composed of four stem-loops (SL1-Sl4) essential for DNA packaging (Parkin et al., 1992). SL1 mediates the initial steps of RNA dimerization through the dimerization Initiation Site (DIS), a palindromic sequence in its apical loop made up of six nucleotides. Studies have shown that this palindrome forms an intermolecular kissing loop-loop interaction, and it is now generally accepted that the SL1 participates in various stages of the HIV life cycle such as packaging, reverse transcription, and recombination (Lever et al., 1989; Paillart et al., 1996; Skripkin et al., 1994). The principal splice donor site is found in SL2, SL3 is historically considered the major element of specific gRNA packaging, and SL4 contains the AUG start codon and participates in alternative base-pairing with U5, which may regulate the final stages of the packaging process (Abbink and Berkhout, 2003; D'Souza and Summers, 2005; Keane et al., 2015).

The P6 protein, 52 amino acids long, located at the carboxyl terminus of gag, is required for the efficient budding of nascent virus particles (Dubois et al., 2018; Freed, 2015). P6 contains the 'late domain' which mediates Gag interaction with cellular factors to facilitate the release of the assembled virus particle. This occurs through two short peptide motifs known as PTAP and

YPXnL that recruit host factors namely TSG101 (Tumor Susceptibility Gene 101) and ALIX (ALG-2 interaction protein X) associated with the ESCRT machinery, which mediates budding and vesicle release from the late endosome (Balasubramaniam and Freed, 2011; Garrus et al., 2001; Martin-Serrano et al., 2001, 2003; Strack et al., 2003; VonSchwedler et al., 2003). P6 also mediates Gag-Vpr interaction through the LXXLF p6 domain, enabling the assimilation of Vpr into assembling virions (Dubois et al., 2018; Nabatov et al., 2004; Paxton et al., 1993).

Spacer protein 1 also known as P2, 14 amino acids long, regulates the formation of the mature capsid and is therefore crucial in the assembage of the immature HIV-1 particle. The function of the SP2 also known as P1 remains unclear (Datta et al., 2011; De Marco et al., 2012).

2.3.3.2 Gag-Pol Precursor

The gag ORF overlaps with the 5' end of the Pol precursor ORF. Thus, Pol is therefore translated fused to the gag protein, yielding the Gag-Pol precursor protein (Parkin et al., 1992). This precursor is the product of a ribosomal -1 frameshifting event that involves a cis-acting element made up of a sequence of seven nucleotides, and a downstream short stem-loop structure that causes ribosome stalling during Gag synthesis (Parkin et al., 1992). The shift in Pol ORF remains uninterrupted on encounter with this motif approximately 5-10% of the time. In the event of a frameshift, a conserved 20:1 (Gag:Gag-Pol precursor) ratio is usually observed (Shehu-Xhilaga et al., 2001). The incorporation of Gag-Pol polyprotein or the constituent mature Pol domains is an indispensable requirement for infectivity of the virus particle, as the constituent domains of these polyproteins mediate viral cDNA synthesis and subsequent integration of the virus into the genome of the host (Cimarelli and Darlix, 2014; Craigie and Bushman, 2012; Louis et al., 1999; Sarafianos et al., 2009; Silliciano and Lusic, 2017).

2.3.3.3 Protease (PR)

HIV-1 protease is by far the most studied retropepsin due to its essential role in viral replication. PR is responsible for Gag, Gag-Pol as well as further Pol processing shortly after or concomitant with virion release, leading to morphological maturation of the virus (Coffin et al., 1997; Konvalinka et al., 2015; Pettit et al., 2004; Turner and Summers, 1999). HIV-1 PR is translated as part of the Gag-Pol polyprotein precursor. The arrangement of the *gag, pol,* and *pro* ORFs varies among retroviruses, resulting in variation in the frameshift event that modulates Gag:Pol ratio. As a result, PR is synthesized in equimolar amounts to Gag, RT, or IN, or in lower amounts compared to the structural proteins but in higher amounts than the other enzymes (Konvalinka et al., 2015).

HIV-1 protease is classified as an aspartic protease due to the presence of conserved Asp-Thr/Ser-Gly residues in the active site. Protease activation is initiated by the dimerization of two Gag-Pol precursors. During dimerization, each polyprotein monomer contributes an aspartic acid residue to the binding cleft. The mature enzyme is active in its homodimeric form comprised of two 99-amino acid subunits, each with a copy of the catalytic triplet. (Navia et al., 1988; Wlodawer et al., 1989). Since the protease is embedded in the Gag-Pol polyprotein, the maturation process is set in motion by auto-processing of the polyprotein precursor. The resultant PR molecules then mediate additional cleavage events during maturation which occur in concomitant with or shortly after viral budding (Louis et al., 1999). The assembly of fully infectious viral particles is dependent on orderly protease-mediated processing. Studies have shown that mutations that alter cleavage sites modify the systemic way in which these sites are cleaved, resulting in anomalous virion assembly which ultimately affects viral infectivity (Pettit et al., 2005).

2.2.3.4 Reverse Transcriptase (RT)

RT is a multifunctional enzyme that modulates the synthesis of proviral DNA. It is a heterodimeric protein made up of two distinct yet related chains: a large 66kD, 560 amino acids long subunit known as p66, and a 51kD, 440 amino acid chain termed p51 (Lightfoote,' et al., 1986).

RT contains two enzymatic domains that corporate to convert genomic viral RNA into double-stranded proviral DNA. The first is a DNA polymerase domain that can act on either an RNA or a DNA template, while the second is a ribonuclease (RNase H) domain that catalyzes simultaneous RNA degradation during cDNA synthesis, defining proviral DNA terminals for host integration (Dean et al., 1996; Rosenthal, 1994). Both subunits contain a common N-terminus but the larger p66 subunit contains enzymatic sites for both enzymatic activities of RT, while the smaller chain lacks an RNase H domain at the C-terminus and plays a structural role (Sarafianos et al., 2001). It is estimated that additional RT processing catalyzed by viral protease occurs in about 50% of Gag-Pol molecules. The result is a shorter RT (55kDa) lacking the carboxy-terminal RNase H domain (Zennou et al., 2000). A HIV-1 virion may contain up to 50 different RTs. It is not clear whether the same RT that synthesizes DNA also degrades RNA. This is not a prerequisite since retroviruses can replicate, albeit with lower productivity, with an assortment of RTs, with varying enzymatic activities of reverse transcriptase (Hu and Hughes, 2012).

RT is activated in the cytoplasm and requires a primer and template just like other DNA polymerases. Host tRNA (tRNAlys3) primer initiates the synthesis of proviral DNA by hybridizing to the primer binding site (PBS), a complimentary 18-nucleotide, sequence located at the 5' end of the viral genome (Klarmann et al., 1997; Zack et al., 1992; Zhuang et al., 2002). During replication, RT generates both strands of proviral DNA by utilizing the genomic RNA (plus strand) as a template for the synthesis of the first strand yielding an RNA-DNA hybrid that functions as a substrate for RNase H. The resulting cDNA (first minus strand) is then used as a template for the synthesis of the second strand. The result is a double-stranded proviral DNA that functions as a substrate for integrase (Swanstrom and Coffin, 2012; Zack et al., 1992).

Viral RNA terminals contain direct repeats known as R that serve as a bridge for the transfer of the nascent DNA minus-strand to the 3' end of the viral RNA (Yu et al., 1998). Synthesis continues along the length of the genome following strand transfer. Although RNase H is not site-specific, a purine-rich sequence in s known as the polypurine tract (PPT) is relatively impervious to RNase H action, and functions as an initiation site for the synthesis of the second plus strand. All retroviral genomes harbor at least one PPT site. HIV-1 has two PPT sites located in the middle and proximal to the 3' end of the viral RNA (Sarafianos et al., 2009). Viral DNA is detectable within hours of infection since reverse transcription occurs shortly following entry of the virus(Butler et al., 2001; Mbisa et al., 2009; Thomas et al., 2006). Minus strand synthesis is estimated to occur at a rate of 70 nucleotides per minute (Mbisa et al., 2009), while synthesis of the plus strand is relatively faster (Klarmann et al., 1997; Miller et al., 1995; Thomas et al., 2007). However, the rate of DNA synthesis varies based on the target cell type. Quiescent cells exhibit slow rates due to the low levels of dNTPs, and synthesis may be derailed in resting T cells (Zack et al., 1992).

HIV-1 sequences are extremely diverse not only between infected individuals but also within an infected individual. This is somewhat surprising given that only one founder virus is required for the initiation of infection, suggesting that the diversity arises post-infection. The large population and rapid turnover of the infected cells are major contributing factors to this diversity. However, the ultimate source of viral diversity is the RT-induced mutations that arise during the virus life cycle (Coffin et al., 1997). Several enzymatic anomalies have been observed for the RT domain of thepolymerase. First, an intrinsic error-prone nature and lack of proofreading activity in combination with a likelihood of incorporating G in favor of A results in the generation of related,

yet distinct viral variants during replication (Simon et al., 2006). Secondly, RT tends to jump between two co-packaged RNAs, employing segments of the individual RNAs as a template to facilitate deletions and insertions, generating chimeric DNA (Eberle and Gürtler, 2012; Hu and Temin, 1990). Switching can transpire between two genetically identical RNAs or genetically different RNAs (when a cell is infected by more than one virus) co-packaged together in a virion. The latter results in recombinant progenies with distinct genotypes compared to the genitors (Hu and Temin, 1990).

2.2.4.5 Integrase (IN)

IN mediates proviral DNA insertion into the genome of the host target cell (Pruss et al., 1994). This enzyme contains a central catalytic domain that is conserved in retroviruses. The catalytic activity of integrase is dependent on a zinc-finger motif made up of three alpha helices, situated at the amino-terminal of this protein. The mature enzyme is made up of three distinct structural and functional domains. First, two nucleotides are trimmed from each of the 3' ends of the linear viral DNA duplex by exonuclease activity. Double-stranded endonuclease activity then cleaves the DNA of the host at the site of integration, and finally, ligase mediates the formation of covalent linkages on either end of the proviral DNA (Brown et al., 1989; Fujiwara and Mizuuchi, 1988).

Areas with high gene density and transcriptionally active domains within the host chromosomes have been identified as hotspots of proviral DNA integration, as they promote efficient viral gene expression after integration (Shinn et al., 2002; Wiskerchen and Muesing, 1995). The route of entry into the nucleus, cell cycle phase, high content of GC, high CpG island density, the structure of chromatin (easy access of PIC to euchromatin) as well as sequence specificities which include specific histone marks and integrase interaction with host components such as the lens-epithelium-derived growth factor (LEDGF), also regulate the preference of transcriptional units as integration sites (Albanese et al., 2008; Singh et al., 2015). Nuclear architecture has been recently implicated as a key mediator in the targeting of the PIC towards nuclear environment permissive for efficient viral gene expression (Silliciano and Lusic, 2017).

2.2.3.6 Env

The outermost envelope of the HIV-1 virus is made up of a lipid bilayer with spikes of glycoprotein known as Env that solely mediate viral binding and entry into target cells (Bernstein et al., 1995). Env is a 160kD (gp160) protein expressed from a singly spliced mRNA, organized as a trimer of

three gp120-gp41 heterodimers that protrudes from the viral surface as a spike (Coakley et al., 2009; Zhu et al., 2003). Each virion contains at least ten Env trimers, but the precise number of these surface trimers varies depending on the strain. The passage history of a virus during culture may also influence the number of env trimers present on the viral envelope (Hartley et al., 2005).

Env synthesis occurs in the ER after which it undergoes post-translational modifications such as glycosylation while migrating through the Golgi complex. Glycosylation, required for infectivity, involves the addition of 25-30 complex N-linked and perhaps O-linked glycans at an asparagine residue (Novikova et al., 2018; Seitz, 2016). These glycan chains are observed on the exposed surface of the spike and function in masking the surface of the env glycoprotein from host immune responses (Hartley et al., 2005; Stanfield et al., 1999). Glycosylation density determined by the number of potential N-glycosylation sites (PNGs), accounts for 50% of the mass of gp120 and is significantly associated with subtype diversity, viral coreceptor tropism, and viral suppression. R5 isolates have relatively fewer PNGs than X4 isolates at specific positions. Some subtypes such as subtype D contain a relatively higher glycosylation density. Furthermore, viruses isolated from treatment-experienced patients with virologic success, are more glycosylated in comparison to isolates from treatment-experienced patients experiencing virologic treatment failure (Kitawi et al., 2017; Korber et al., 2001; Pollakis et al., 2001).

After translation, cleavage of gp160 the furin family of endoproteases yields the two non-covalently associated subunits of Env: gp41 and gp120 (Wyatt and Sodroski, 1998). Gp41 is the C-terminal subunit comprised of cytoplasmic, transmembrane, and extracellular domains that moderate conformational changes during fusion. Besides catalyzing the merging of the cellular and viral membranes, gp41 additionally contains a fusogenic moiety at its N-terminal whose key role is anchoring Env to the viral host-derived membrane (Coakley et al., 2009). Gp120, the N-terminal moiety of the Env glycoprotein mediates the interaction between CD4 and the virus (Arrildt et al., 2012). This moiety additionally interacts with the DC-SIGN receptor on the surface of dendritic cells, increasing the infection efficiency of CD4 positive cells. DC-SIGN also transports HIV to lymphoid tissues, facilitating mucosal transmission (Geijtenbeek et al., 2000; Turville et al., 2002).

Each gp120 subunit is comprised of five hypervariable regions: V1, V2, V3, V4, and V5, which variable amino acid sequences among HIV-1 isolates and five highly conserved regions termed C1-C5 (Rosenthal, 1994). Highly conserved regions such as the gp41 binding surface are made up of residues from the C1 and C5 regions. The CD4 and coreceptor binding sites found in the inner domain of the gp120 core play a role in gp120 folding. The CD4 interacts with conserved sequences in gp120 which consist of sequences from either side of V4 (Arrildt et al., 2012). The variable loops localize to the surface of gp120 and are essential in mediating infection via the V3 loop and partial occlusion from immune recognition by the host (Hartley et al., 2005). Host chemokine coreceptors CCR5 and CXR4 interact with a GPCR motif at the crest and base of the V3 loop. Therefore, the V3 loop determines coreceptor usage and thus accounts for viral tropism (Pastore et al., 2006a). Since the susceptibility of different cell types to different strains of the virus is partially determined by these coreceptors, isolates are labeled as R5 when using the CCR5 coreceptor, X4 when using CXR4, and dual tropic/mixed (R5X4) when utilizing both coreceptors (Berger et al., 1998).

Additional env regions such as V1/V2 and C4 have been observed to influence coreceptor tropism (Berger et al., 1998; Bryant and Ratner, 1990; Hwang et al., 1991). The format of glycosylation of the V1/V2 regions also modulates susceptibility of the virus to NAbs that recognize various epitopes. Studies by (Saunders et al., 2005) revealed that the V1/V2 loops may have differing and sometimes opposite effects on the neutralization susceptibility profile of HIV-1. While V2 loop deletion increased the virus' susceptibility to NAbs targeting various epitopes, particularly specific ones within the CD4 binding site and the V3 loop, V1 deletion made the virus obstinate to NAbs targeting epitopes induced by CD4.

The env domain that encodes gp120 is believed to evolve faster as compared to any other region of the HIV-1 genome, with an estimated evolution rate of 1-2% per year. Since env is the principal NAbs target, the selection pressure on this gene is unique. The formation of escape mutants is thus driven by antibody and cell mediated immune response of the host (Arrildt et al., 2012). However, the change is limited to discrete regions of the protein to ensure that the variability does not alter structures required for infection and ultimately their function. These changes may nevertheless result in altered interaction between gp120 and host receptors. As the virus evolves over the course of an infection, the env protein may alter receptor and coreceptor usage to facilitate access to a

broader pool of target host cells. Coreceptor switching occurs in up to 50% of infections in advanced disease stages and usually correlates with poor prognosis (Connor et al., 1997).

2.2.3.6.1 Variable loop 3 (V3 loop)

The V3 loop is about 40 amino acids long and forms a loop-like structure that protrudes on the outermost domain of gp120. It is replete with basic amino acids and exhibits high sequence diversity among different isolates of the virus (Stanfield et al., 1999). It also contains aromatic amino acids that participate in stacking protein-protein interactions (Huang et al., 2005). Early studies on Env showed that the V3 loop contained both variable and conserved regions (LaRosa et al., 1991). These sequences were further classified into three structural regions namely a β-hairpin tip, a conserved base that constitutes an intrinsic chunk of the core, and a pliable stem that extends away from this core (Huang et al., 2005).

A 35 amino acid subregion of the V3 has been mapped as the major modulator of biological and immunological HIV-1 phenotypes. It directly binds to coreceptors during viral entry and thus contains the most important decisive factor of coreceptor usage (Hwang et al., 1991). Specific V3 loop sequences are believed to give rise to several conformational variations in gp120 and V3 which in turn modulate the function of gp120 (Huang et al., 2005; Wyatt et al., 1992). Research findings have highlighted the evolution of some cell types to not requiring CD4 for infectivity purposes, but chemokine binding remains crucial as demonstrated by infection abrogation in isolates with a V3 deletion (Saunders et al., 2005). It is also an important determinant of preferential HIV-1 tropism for either T cells or macrophages (Bryant and Ratner, 1990). The V3 sequences of T-tropic viruses have been observed to be more basic in terms of charge as compared to M-tropic viruses, a characteristic attributed to an accumulation of positively charged residues on either side of the GPCR peak of the loop (Fouchier et al., 1992; De Jong et al., 1992). However, there are no detectable patterns that can be attributed to changes in viral phenotype. The V3 loop is a critical determining factor of chemokine receptor -CXCR4 and CCR5- binding. Since the susceptibility of different cell types to different strains of the virus is partially determined by these coreceptors isolates are labeled as R5 when using the CCR5 coreceptor, X4 when using CXR4, and dual tropic/mixed (R5X4) when utilizing both coreceptors (Berger et al., 1998)**.** In this way, the V3 loop affects sensitivity to, and susceptibility of HIV-1 to a class of ARTs known as entry inhibitors particularly the CCR5 antagonist, Maraviroc (Westby et al., 2006).

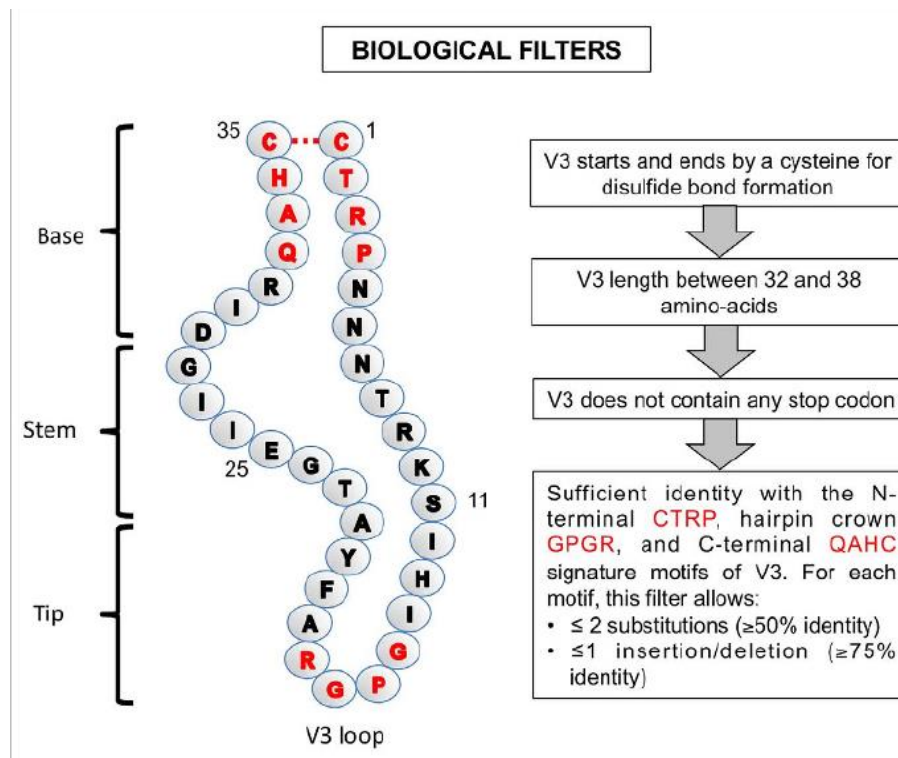


***Figure 5: Structure of the V3 loop**.*
Adopted from (Jeanne et al., 2015)

When V3 amino acid sequences and their corresponding tropism profiles were initially statistically analyzed, the results gave the impression that the occurrence of a positively charged amino acid at position 11 and or 25 of the V3 loop, conferred an X4 phenotype, while the lack thereof conferred the default R5 phenotype (Fouchier et al., 1992). This rule, commonly termed the '11/25 rule' has been used in conjunction with other properties of the V3 loop to develop in silico genotypic tropism testing tools. The V3 loop is also believed to be an electrostatic modulator of viral phenotype. Discrete disparities in the amino acid sequence of the V3 loop, particularly dispersal of the charged amino acids have been shown to have a bearing on viral phenotype (Chesebro et al., 1992; Fouchier et al., 1992; De Jong et al., 1992). CCR5 variants have been observed to be less positively charged as compared to their CXCR4 counterparts (Chesebro et al., 1996; Naganawa et al., 2008). Moreover, a net charge increment in CCR5 variants, and loss of an N-linked glycosylation event in the V3 loop have been linked to tropism switching to CXCR4 and loss of antibody resistance (Kato et al., 1999; Milich et al., 1997; Pollakis et al., 2001; Shioda et al., 1994). (Yokoyama et al.,

2012) combined computational and experimental approaches to examine the influence of the V3 loop universal charge on the surface of the outer region of gp120. Two V3 recombinant HIV-1 infectious clones with a net charge of +3 and +7 were used. Fluctuating the V3 net charge impelled global changes on gp120 that fluctuated the configuration of V3 loop regions involved in CD4, coreceptor, and antibody binding. The clone with a +3-net charge was also observed to be resistant to CD4 binding site-MAbs because of reduced exposure of the loops binding the CD4 receptor.

The crown of the V3 is particularly immunogenic and was for a long time considered a Principal Neutralization Determinant (PND) since it is a prominent target for neutralizing NAbs (Javaherian et al., 1989). Emerging data however showed that the efficacy of the NAbs was particularly contingent on the virus under testing, the V3 Ab epitope specificity, and the neutralization assay employed (Zolla-pazner et al., 1999). Therefore, this region is very prone to changes in amino acid sequences that directly alter NAbs epitopes, resulting in neutralization escape mutants that are not efficiently recognized by the antibodies (Hartley et al., 2005). These escape mutants usually arise from the expansion of a minor, naturally resistant variant, or a de novo mutation that alters how the V3 loop is packed within the Env trimer. Due to this conformational plasticity, V3 alternates between an occluded conformation accessible to chemokine receptors resulting in infection, or an open conformation accessible to V3 Abs which neutralizes the virus (Zolla-Pazner et al., 2016). Studies have shown that substituting amino acids at positions 313-314 of the V3 loop affects its sensitivity to V3-directed antibodies as well as the sensitivity of other antibodies directed to other regions of gp120 (Takeuchi et al., 1991). In contrast, mutations introduced to other regions of the V3 can also activate an open conformation that is more accessible to V3 NAbs (Zolla-Pazner et al., 2016). Experimental removal of the V3 region has also resulted in improved accessibility to other gp120 epitopes, by antibodies (Binley et al., 2000; Wyatt et al., 1992), implying that the V3 loop overlies the more conserved receptor binding sites.

2.3.2 Regulatory proteins

2.3.2.1 Tat (Transactivator of transcription)

Tat is an 86-102 amino acid long, 14-16kDa product of the *Tat* gene that regulates HIV-1 genome transcription. Tar binds the transactivation response element (TAR), and enhancer-promoter-binding factors, accelerating availability of viral RNA for virus production x100 (Isaacman-Beck et al., 2009; Jeang, 1996; Seitz, 2016).

Two exons spliced together encode the tat protein. The first exon is conserved within the HIV-1 subtypes while the second exon varies (Perkins et al., 1989). Tat is made up of two domains termed the transcription activation and basic domains. The former consists of a region rich in cysteine residues that function in the dimerization and stabilization of the protein structure as well as the binding of metals and a hydrophobic core. The latter is an arginine-rich motif that binds RNA conferring specificity of the protein to a triplet base in the bulge of the stem-loop TAR, localizes Tat to the nucleus, and interacts with surface proteins such as heparan sulfate proteoglycans to facilitate the internalization of this protein into bystander cells (Clark et al., 2017; Ruiz et al., 2019; Tyagi et al., 2001)

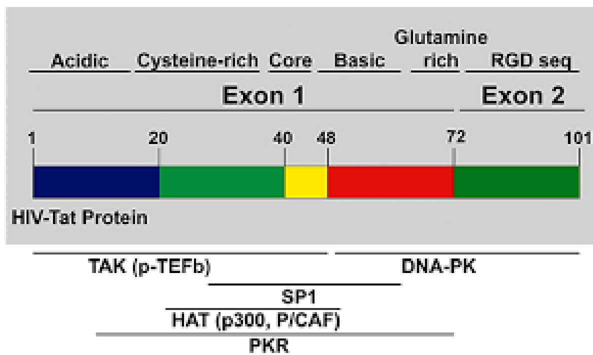


Figure 6: HIV-1 Tat protein.
Encoded by two exons (1-72) and (72-101). The structural and functional domains are also shown. *Adopted from (Tsukamoto et al., 2020)*

Tat principally acts as a promoter of the initiation and elongation phases of HIV-1 transcription, promoting full-length transcript production (Feinberg et al., 1991; Perkins et al., 1989). To stimulate the elongation phase, Tat recruits an RNAP II elongation factor known as P-TEF13 (Zhu et al., 1997). P-TEF13 (positive elongation factor 13) is comprised of catalytic and regulatory subunits termed CDK9 and Cyclin T1, respectively (Majello et al., 1999). CDK9 is a serine kinase that catalyzes the phosphorylation of the carboxyl-terminal (CTD) of RNAP II, while Cyclin T

enables the identification of the TAR loop region by the Cyclin T-Tat complex (Herrmann and Rice, 1995; Majello et al., 1999; Peng et al., 1998). Further studies revealed Tat's mechanism of action was not as simple as suggested above. Studies showed that the association of RNAP II with two multiprotein complexes known as the NELF (negative elongation factor) and DSIF (sensitivity-inducing factor) inhibits transcription elongation (Yamada et al., 2006). Recruitment of NELF to the HIV-1 promoter is followed by direct binding of its RD subunit to HIV-1 TAR and P-TEF13 whether Tat is present or absent. The kinase activity of P-TEF13 then phosphorylates RD, abolishing its interaction with TAR. This results in short, promoter-proximal transcripts, a factor that may contribute to low transcription levels in HIV-1 latent cells (Fujinaga et al., 2004). CDK9 was found to additionally phosphorylate subunits of both multiprotein complexes potentially activating the elongation phase of transcription (Fujinaga et al., 2004; Zhou et al., 2012).

Tat has also been observed to exert pleiotropic effects on cells, resulting in the upregulation of genes that modulate inflammation and activation of cytotoxic pathways. Tat promotes angiogenesis by stimulating the growth of Kaposi-like cells, by acting as a secreted growth factor. It also up-regulates the expression of cytokines such as IL-2, IL-2, and TGF, and downregulates Bcl-2 expression. Activation of cellular transduction pathways that utilize PI-3 kinase, and NF-KB are also affected by this viral protein (Tsukamoto et al., 2020). Neurotoxicity has been observed at low Tat concentrations in addition to progressive deregulation of neurons, and astrocytes that eventually results in the occurrence of HIV-Associated Neurocognitive Disorders(HAND), and accelerated brain aging (Bagashev and Sawaya, 2013).

2.3.2.2 Rev (Regulator of Expression of the Virion)

Rev is a 13-kD phosphoprotein roughly 116 amino acids long that regulates the splicing and transport of newly formed RNA to the cytoplasm (Perkins et al., 1989; Seitz, 2016). It binds to a specific RNA sequence known as the Rev response element (RRE). This is a 240 bp region located in the second intron of HIV-1 that forms a complex RNA secondary structure, including a 'bubble' within this stem-loop structure to which Rev binds with high affinity (Micheal et al., 1989). The Rev protein is made up of three domains; an arginine-rich RNA binding domain that moderates sequence-specific interaction with the RRE and localization of the nucleus/nucleolar (Cochrane et al., 1990; Kjems et al., 1992), a multimerization domain critical for its function (Hope et al., 1992;

Zapp et al., 1991) and an effector domain that doubles up as specific nuclear export signal(NES) (Fischer et al., 1995).

Viral gene expression is comprised of two phases: Rev-independent early phase, and Rev-dependent late phase. Rev modulates early to late phase transition during replication (Micheal et al., 1989). Within the cytoplasm, a GTPase known as Ran facilitates the interaction of multimerized Rev with importin β to form a complex that transports Rev into the nucleus of infected cells (Izaurralde et al., 1997). Once in the nucleus, Rev disassembles from this complex and binds to the RRE (Hope et al., 1990; Zapp et al., 1991). Accumulation of the Rev protein to a certain threshold triggers the cellular splicing machinery to halt the production of multiply spliced transcripts and shift to producing singly spliced transcripts, which typically occurs in the late phase. The leucine-rich Rev NES facilitates binding of Rev with its RNA load to exportin-1(CRM1) (Fornerod et al., 1997; Fukuda et al., 1997; Stade et al., 1997) and its ultimate transport into the cytoplasm where Ran catalyzes the disassembly of this complex, releasing the RNA cargo (Nielsen et al., 2005).

Additional effects of Rev include increased stability and translation of HIV-1 mRNA (Felber et al., 1989). Studies by (Malim and Cullen', 1993) showed that the half-life of RNAs within the nucleus of a HIV-infected T-cell line increased significantly in the presence of Rev.

2.3.3 Accessory proteins

2.3.3.1 Viral Protein U (Vpu)

Vpu is a 9KDa membrane protein that stimulates virion release from the surface of the cell and particularly targets the CD4 protein to an ER-degradation course (Aloja et al., 1998; Miller et al., 2000). CD4 employs a domain in its cytoplasmic tail that is proximal to the membrane to attach to h-βTrCP, which in turn facilitates the interaction of CD4 and ER expressed Vpu (Margottin et al., 1988). h-βTrCP is belongs to the F-box protein family that were initially identified as elements of ubiquitin-ligase complexes (Kipreos and Pagano, 2000). This ternary complex then engages an additional component of the ubiquitination machinery known as SKP1 that ubiquitinates CD4 and consequently targets it for proteasome degradation (West et al., 2004).

Increased secretion of virus progeny from infected cells over expressing Vpu was initially attributed to Vpu-induced pore formation within ion conductive membranes (Bour and Strebel, 2003). However, studies later showed that the Vpu requirement is dependent on the host cell

suggesting that this protein may negate the effects of inhibitory factors such TASK-1(an acid-sensitive K^+ channel) in some cell types but not in others (Hsu et al., 2004).

2.3.3.2 Viral Protein R (Vpr)

Vpr is a basic, 96 amino acid long protein highly conserved within H1V-1 and 2, and SIV (Tristem et al., 1992). The biological function of Vpr is exerted on many targets as indicated by the presence of this protein in several compartments including the virion, cerebrospinal fluid, and sera of infected subjects (Paxton et al., 1993). It is an element of the reverse transcription complex (RTC) and possibly interacts with uracil DNA glycosylase of HIV-infected individuals, modifying newly synthesized viral DNA (Chen et al., 2004; Mansky et al., 2000). Vpr also mediates the importation of the pre-integration complex(PIC) into the nucleus, anchors this complex to the envelope of the nucleus, and translocates viral DNA into the nucleus (Bukrinsky and Adzhubei, 1999; Cullen, 2001; Fouchier and Malim, 1999).

Furthermore, Vpr impedes the progress of the host cell cycle progression by inducing arrest in the G2 phase. Progress from G2 to mitosis during the cell cycle is highly governed by a kinase known as Cdc2. Vpr induces the hyperphosphorylation of Cdc2 rendering it inactive (Re et al., 1995; Zhao et al., 1996). Vpr also stimulates apoptosis in cells infected with HIV (Alimonti et al., 2003). The mechanism of action Vpr in inducing apoptosis is elusive. It is debated whether the G2-arrest triggers Vpr-induced apoptosis as indicated by a decreased apoptosis rate in the presence of a cell cycle regulator known as Wee-1 kinase (Yuan et al., 2003). The most favored explanation suggests that apoptosis is induced through the mitochondrial pathway that ultimately results in the release of apoptosis-mediating factors following mitochondrial swelling (Brenner and Kroemer, 2003; Jacotot et al., 2000, 2001). Vpr has also been shown to activate Caspase-9, setting in motion caspases of the intrinsic pathway of apoptosis (Muthumani et al., 2002). Contrary studies have suggested that Vpr may be a negative regulator of T cell apoptosis (Ayyavoo et al., 1997; Conti et al., 1998).

Moreover, Vpr transactivates HIV-LTR and host cellular genes. Host responses to Vpr include CD8+ T-lymphocytes-Vpr targeting in acute HIV infection (Altfeld et al., 2001; Mothé et al., 2002) and expression of heat shock proteins such as HSP70 protecting cells from the effects of Vpr (Benko et al., 2004; Iordanskiy et al., 2004b, 2004a). This protein alters cell proliferation and

cytokine production (e.g. TNF α and IL-12), and chemokines such as RANTES which in turn down-regulate humoral and cellular immune response (Ayyavoo et al., 1997; Refaeli et al., 1995).

2.3.3.3 Virus infectivity factor (Vif)

Vif is a 192 amino acid long protein concentrated in the cytoplasm of infected cells (Strebel et al., 1987). Vif is thought to influence the latter phases of virion assembly and also blocks premature Gag-precursor processing by HIV protease. This ensures that the products of gag processing namely CA, MA and NC and are available during virion assembly at the surface of the cell (Aloja et al., 1998; Miller et al., 2000). Research conducted by (Sheehy et al., 2002) implicated Vif in overcoming host antiviral elements such as APOBEC3G (apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G), a cytidine deaminase that induces hypermutations in newly synthesized HIV-1 DNA, via deamination of deoxycytidines in the minus strand of the viral cDNA replication intermediate (Harris et al., 2003; Lecossier et al., 2003a; Zhang et al., 2003). G-to-A transitions and stop codons are consequently generated, prompting the degradation of the newly synthesized DNA by host glycosidases and repair enzymes. Thus, the end product is hypermutated viral DNA (Lecossier et al., 2003b). Vif complexes with APOBEC3G, blocking its encapsidation by the virus (Schäfer et al., 2004). Moreover, Vif mediates the breakdown of APOBEC3G by targeting it to the ubiquitin-proteasome pathway (Conticello et al., 2003; Marin et al., 2003). Furthermore, Vif interacts with genomic HIV-1 RNA at the 5’ region to stabilize the viral nucleoprotein complex (Gabuzda, 2000; Henriet et al., 2005; Höglund et al., 1994). This protein also modulates RT activity by enhancing the binding of RT and primer or upregulating RT polymerization (Cancio et al., 2004).

2.3.3.4 Negative regulating factor (Nef)

Nef is a 27-kD myristoylated protein produced in abundance in the initial stages of the replication cycle. It is conserved to a great extent in lentiviruses that infect primates, and is thus crucial for the survivability of these viruses (Ahmad and Venkatesan, 1988; Cheng-Mayer et al., 1989) in four major ways: Nef alters expression of several cellular proteins at the cell surface. MHC class I & II, CD4, CD8, and CD28 expression are downregulated by Nef (Garcia and Miller, 1991; Stove et al., 2005; Swigut et al., 2001). Owing to the importance of these receptors in the unmitigated performance of the immune system, modulation of their expression at the cell surface greatly impairs anti-HIV immune responses. This shields HIV-infected cells from CTL response and decreases the adhesion of these cells to APCs, promoting their movement and thus, that of the

virus. Reduced CD4 expression also abolishes interaction between this receptor and Env, increasing the rate of virion release from the infected cells. Conversely, Nef upregulates the expression of CD74 (Schindler et al., 2003).

Secondly, Nef has been shown to interfere with transduction signals that impart the activation of infected cells, by interacting with lipid rafts and cholesterol-rich domains of the membrane that cluster puissant signal mediators (Wang et al., 2000). The interruption of these signaling pathways induces the release of chemokines and T cells attracting factors and enhances infection of these cells by the virus (Schmidtmayerova et al., 1996; Swingler et al., 2003).

Thirdly, the processing of Nef by proteases during replication has been observed to alter virion infectivity and replication of the virus (Chowers et al., 1994). Virions generated in the presence of Nef are highly infectious (up to 10X) as compared to those produced in its absence (Seitz, 2016). This is attributed to the Nef-induced remodeling of actin and facilitation of transport of the viral core post a presumably obstructive cortical actin barrier (Campbell et al., 2004). Lastly, this protein modulates the trafficking of cholesterol in infected cells. Since cholesterol is crucial in the viral life cycle, a decrease in cellular levels of cholesterol noticeably decreases HIV-1 virion production (Maziere et al., 1994). Nef contains a cholesterol recognition motif at the C-terminal that binds to and traffics cholesterol to viral budding sites (Zheng et al., 2003).

2.4 HIV Transmission

Although there has been a significant global decline in new HIV-1 infections, some regions and risk groups are still disproportionally affected, underlining the importance of different routes of transmission and behavioral risks as facilitators of transmission. HIV is transmitted through sexual contact across mucosal surfaces, percutaneous inoculation, and maternal-infant exposure. Different exposure routes are associated with varied infection risks (Cohen et al., 2011a).

Sexual contact-either homosexual or heterosexual, through cervicovaginal, penile and rectal route-is the major route through which HIV is transmitted (Ames et al., 1998). This is attributed to the presence of the virus in semen in both infected mononuclear cells and in the seminal fluid, as a cell-free virus as well as in cervical smears and vaginal fluid (De Vincenzi, 1994). Despite reduced transmission probability per coital act (1 in 200-1 in 3000), heterosexual transmission is said to contribute to up to 70% of global HIV infections, while men who have sex with men (MSM), mother-to-child transmission (MTC) and injection drug use (IDU) contributing to the remainder (Hladik and McElrath, 2008; Shaw and Hunter, 2012). However, the transmission probability for heterosexual exposure is attributed to other confounding risk factors such as male circumcision, STD/STI induced genital inflammation, and ulcers genital that increase the shedding of the virus into the genital tract, increasing susceptibility to infection by X2-X11, clinical stage of HIV disease, and route of exposure (Hladik and McElrath, 2008; McElrath et al., 2008; Quinn, 2007).

The penile-vaginal route bears a transmission probability of 1 in 10 (Powers et al., 2008). Transmission is correlated to viral load (VL) and clinical stage of the disease-acute vs middle vs late (Shaw and Hunter, 2012; De Vincenzi, 1994). The risk of transmission is determined by the viral load of the transmitting partner and the clinical stage of infection. This is highlighted in sero-discordant couples where the risk of transmission was observed to increase by X2.5 for every 10-fold increment in viral load, and a 96% decrease in transmission was observed when the VL of the infected partner is <1000copies/ml (Cohen et al., 2011b; Quinn et al., 2000). Transmission is also most efficient during acute infection, reflecting high VLs in this infection phase, absence of neutralizing antibodies to inactivate the circulating virus, and clonal expansion of fit viruses (Brenner et al., 2007a; Miller et al., 2010).

There is a compelling relationship between the transmission of the virus with receptive anal intercourse with a reported transmission risk of 1 in 3 (Powers et al., 2008). This has been

associated with the thin and fragile rectal mucosal membrane that acts as a barrier between the semen deposit and the potential target cells within and beneath the mucosa of the anus, and anal intercourse-induced trauma. In contrast, the vaginal mucosa is relatively thicker and less susceptible to intercourse-induced trauma, but reciprocal viral transmission can occur during vaginal intercourse (De Vincenzi, 1994). Transmission of the virus through oral sex is less efficient. The hypotonicity of saliva was shown to disrupt >90% of mononuclear cells that host, produce and transmit the virus as well as other cultured cells with concomitant multiplication inhibition (>10,000 fold) of HIV and other surrogate viruses (Baron et al., 1999). However, cases of receptive transmission through receptive fellatio and insertive cunnilingus, albeit rare, have been reported (Weller, 1993).

Percutaneous transmission of HIV is attributed to blood transfusions and by sharing of needles among injection drug users (IDUs). Hemophiliacs and blood or blood products recipients represent approximately 2% of total HIV infections. Nowadays, the risk of transfusing contaminated blood and blood products has been reduced by screening of the blood supply (Coffin et al., 1997)

Vertical transmission from a mother who is infected to a child may occur either in utero, perinatally, or through breast feeding and represents at least 90% of all HIV infections in children and infants. Mother-to-child transmission in late pregnancy occurs following entry of the mother's blood into the fetal circulation and perinatally due to exposure of the mucosal membrane to the virus during labor and delivery (Dunn et al., 1992).

2.5 Clinical staging of HIV infection

Irrespective of the transmission route, the blood appearance of viral and host markers generally follows an orderly and reproducible pattern following a clinically productive transmission event.

2.5.1 Acute phase

This is the earliest stage and generally develops two to four weeks following infection with HIV. The acute phase is accompanied by non-specific, mild symptoms including symptoms similar to those of flu such as fever, headaches, rash, adenopathy, nausea, and fatigue in 40-90% of infected individuals (Ames et al., 1998).

Acute infection is characterized by intensive viral replication which translates into a viral peak (VL> 100,000 copies per ml of blood) and by a transient drop in circulating CD4 lymphocytes in the peripheral blood and lymphatic tissues of the lymph node and gut-associated lymphocytes (GALT), where 94% of CD4+ T cells reside (Cohen et al., 2011a). This phase is characterized by increased mucosal, submucosal, and lymphoreticular viral replication (McMichael et al., 2010). A decrease in T cell number in the gut coupled with immense growth in the size of the HIV DNA reservoir usually occurs within the first 25 days following the acquisition of the virus. During this stage, viral markers are not detectable in plasma but are observed in the affected tissues only (Estes et al., 2008). This stage is termed the 'eclipse' stage and often lasts between 7 and 21 days post virus acquisition (Cohen et al., 2011a). Once the HIV RNA concentration increases by 1-5 copies per ml in plasma, HIV can be qualitatively detected by techniques that utilize nucleic acid amplification (Palmer et al., 2003). Following this, viral load can be monitored by employing different quantitative clinical assays to the various viral markers that consequently appear (Kumi et al., 2013).

The initial response by the immune system which includes a 'cytokine storm' that coincides with the increasing viremia (ramp-up viremia), may result in acute retroviral syndrome in a significant number of individuals who are infected, and may therefore be used as an indicator of acute HIV infection (Cohen et al., 2011a). This primary rise in virus numbers within the plasma, often beyond levels higher than 1 million RNA copies per ml, is usually followed by a significant reduction in VL to a steady-state level of viral replication. This viral load decrease has been correlated with specific antiviral immune responses that include the cytotoxic T cells (Usey et al., 1997), as well as soluble factors produced by CD8+ cells that hinder viral replication (Martinez-Mariño et al., 2007).

The acute phase is of epidemiological importance since an individual is highly contagious toward their sexual partners because of the very high viral load (Kumi et al., 2013). Moreover, this phase is clinically critical since an individual is rendered incurable from infection by HIV due to the reservoir of infected cells formed, and the resulting viremia at the axe of this phase dictates the disease's natural progression since high viral load and rapid progression of the disease are correlated (Ames et al., 1998). The use of immediate ART to successfully manage early HIV infection (EHI) has been endorsed by many studies where outcomes were a healthy CD4 count and functionality, a HIV reservoir limited in size, and a reduced risk of onward viral transmission (Brenner et al., 2007b).

2.5.2 Asymptomatic phase

Also known as the clinical latency phase, this phase is characterized by HIV inactivity or dormancy. After seroconversion, the viral load settles and the asymptomatic phase which varies in length from one infected individual to the other, follows (Ford et al., 2009). This phase may last up to 20 years with the administration of improved ART since ART minimizes viral replication and thus immune status deterioration. Intensive viral replication continues despite the absence of physical symptoms. Levels of the virus drop in peripheral blood although antibodies against HIV can be detected on a regular basis (Rastogi et al., 2011). The reduction in viremia is associated with antiviral responses of the adaptive and innate immune systems, which include components such as complement and mannose-binding lectins (MBLs) as well as anti-HIV antibodies, T cells, and natural killer (NK). In an unusual group known as 'elite controllers', undetectable viremia may indicate efficient HIV control (Levy, 2009).

HIV latency during this phase has been associated with various host and viral factors such as the site of proviral DNA integration within the genome of the host, the state of the proviral DNA following integration (e.g. methylation), and insufficient expression of viral Tat and Rev which affects viral fitness (Eberle and Gürtler, 2012). Host factors that may suppress viral infection include cellular proteins such as histone deacetylases, YY1 (a transcription repressor protein), and the CD8þ antiviral factor. Intrinsic cellular antiviral factors may be involved, and examples include APOBEC3 and 3F, a cytosine deaminase that interferes with reverse transcription by altering single DNA synthesis, and TRIM5a which prevents uncoating of the virus by interacting with the viral capsid (Levy, 2009).

CD4+ T cells are massively damaged by the virus through head-on and indirect mechanisms including viral cytopathic effects, virus-induced apoptosis, and immunological senescence. Another additional mechanism responsible for the inflammatory damage that results in the architectural destruction of lymphoid organs has been described by (Estes et al., 2008). It is referred to as the ‘damaged niche’ hypothesis and ascribes the depletion of CD4+ T cell numbers to cytokine induced-fibrosis (TGF-β and other cytokines) which may hamper interactions of dendritic T cells and their access to growth factors and cytokines necessary for proliferation and survival of CD4+T cells (Ford et al., 2009).

Immune activation also contributes to pathogenesis during this phase. Epithelial barrier breach and reduction in gut CD4 T cell numbers may result in the release of microbial products that propel immune activation. Studies have also shown that patients infected with HIV, and *Rhesus macaques* infected with SIV exhibit elevated circulating lipopolysaccharide (LPS), and soluble CD14 (Brenchley et al., 2006). As a result, the innate immune system is continuously stimulated to produce type I interferons as well as pro-inflammatory cytokines that further intensify the immune activation and give rise to targets of HIV replication (Brenchley and Douek, 2008; Ford et al., 2009).

2.5.3 AIDS phase

The outcome of HIV infection is the AIDS phase which may present itself anywhere between 1-2 years post-infection. This phase is also referred to as immunodeficiency syndrome and is usually fatal and terminal for infected patients. About 50% of infected individuals progress to AIDS without ART (Grossman et al., 2006; Mosier, 2009; Schuitemaker et al., 1992). The AIDS phase is characterized by extremely low CD4 counts below 35cells/ml and impaired immune functions specifically HIV-specific immune responses mounted by CD4+ and CD8+ cells (Levy, 2009). Severe destruction of the lymphoid tissue orchestrated by the virus, which serves as an indicator of infection progress, also occurs. This severe immune suppression favors the emergence of numerous opportunistic infections and finally encephalopathy and wasting syndrome. It is also during this phase that a tropism switch may take place. The appearance of X4 or mixed/dual tropic variants is usually an indicator of poor prognosis and pronounced reduction in CD4+ T cell number. Such a tropism switch expands the viral targets and occurs after several years of infection (Zaunders et al., 2002).

2.6 HIV pathogenesis

HIV replication occurs in two phases: the early and late phases. Early events include binding and entry of the virus into the target cells to host genome integration of proviral DNA. The latter phase includes events occurring from the expression of the viral genes to the packaging, release, and maturation of the new virion (Freed, 2015).

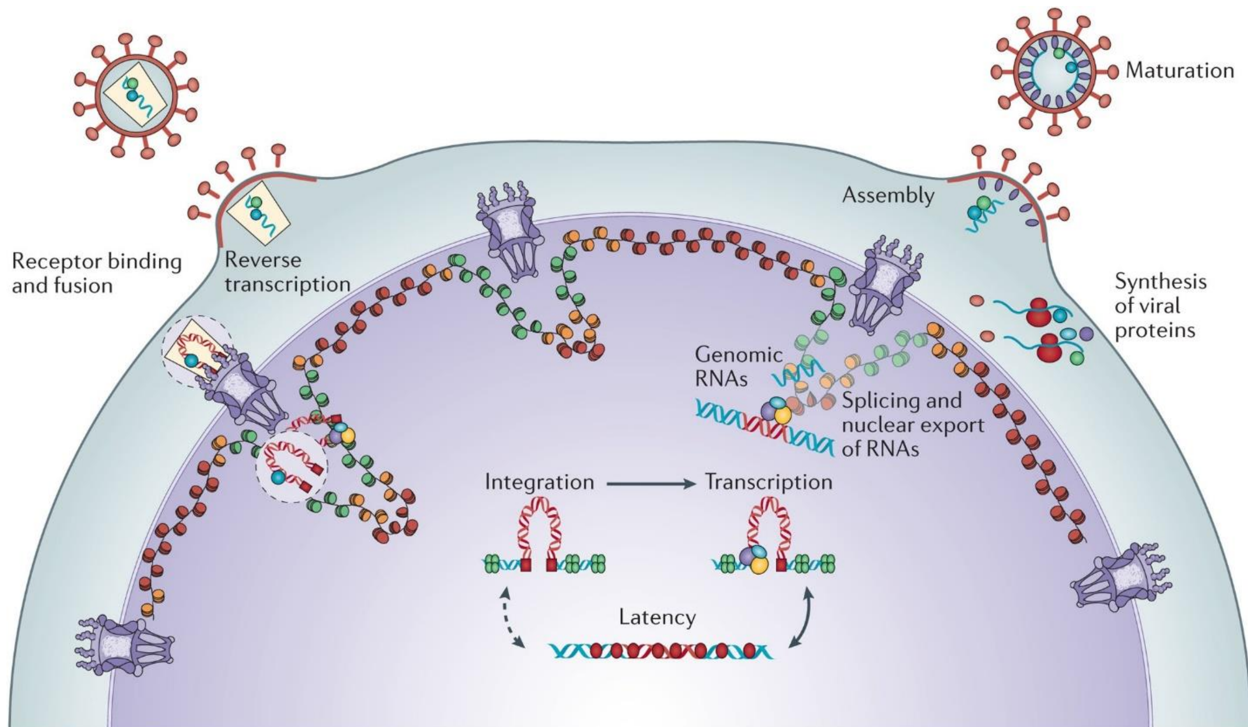


Figure 7: An overview of HIV pathogenesis
Adopted from (Silliciano and Lusic, 2017)

2.6.1 Fusion and entry

Viral entry is facilitated by complex protein-protein interactions that involve both receptors found on the target cells as well structural proteins on the viral envelope. Virus adhesion to the host cell surface, facilitated by viral env kick starts the early phase of the HIV replication cycle (Bernstein et al., 1995). The functional unit of env is gp160, a glycoprotein that appears as spikes on the lipid bilayer. Gp160 is a trimer made up of gp120 (surface moiety) and gp41(transmembrane moiety) heterodimers (Coakley et al., 2009; Zhu et al., 2003). Viral entry begins when gp120 binds to the CD4 expressed on various host cells such as DCs, macrophages, and astrocytes (Bour et al., 1995).

The CD4-binding domain of gp120 binds to the CD4 receptor inducing conformational changes in both proteins that expose the viral V3 loop, which then binds the CCR5 or CXCR4 (fusin) coreceptors on the plasma membrane of target cells (Dean et al., 1996; Feng et al., 1996). The complex protein-protein interactions ultimately result in a 'fusion-active' state necessary for the merging of the cellular and viral membranes (Mulinge et al., 2013).

Additional conformational changes triggered by the binding of gp120 to the coreceptors, that occur in gp120 and subsequently gp41, lead to the presentation of the N-terminus of gp41 on the viral coat where it creates a channel that inserts into the cell surface membrane due to its hydrophobicity (Seitz, 2016), completing marriage of the cellular and viral membranes. This fusion occurs within minutes by pore formation, with subsequent emptying of the viral core into the cytoplasm. Liberation of the capsid contents into the cytoplasm is triggered by pH changes within the phagosome after the capsid is taken up by a late endosome (Simon et al., 2006).

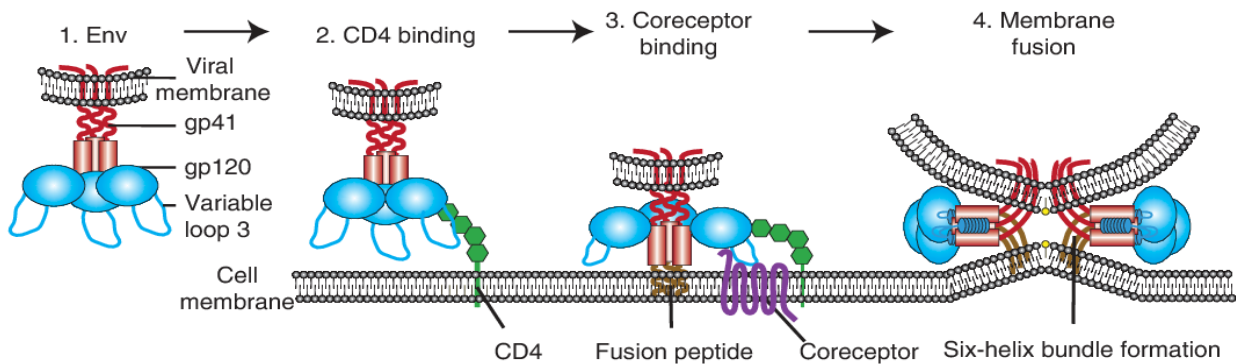


Figure 8: Viral fusion and entry.
Adopted from (Wilen et al., 2012a)

2.6.2 Reverse transcription and nuclear import

Viral RNA is reverse transcribed following the disassembly of the virus core by the RT enzyme. Single-stranded proviral RNA is reverse transcribed into viral cDNA by RT activated in the cytoplasm, while RNase H simultaneously degrades the RNA. This DNA is then used a template for second strand synthesis of viral DNA by the DNA-dependent polymerase activity of RT, forming a double-stranded proviral DNA (Dean et al., 1996; Rosenthal, 1994). The intrinsic lack of proofreading and error-prone nature of RT combined with high viral replication rates generate

homologous yet discrete variants of the virus (Simon et al., 2006) that may differ by up to 10% within an individual (Korber et al., 2001). The proviral DNA is then packaged in a complex made of viral DNA, integrase and capsid proteins, and cellular proteins such as high Mw nucleoprotein complex, forming a PIC that is imported into the nucleus via nucleopores, for subsequent integration (Craigie and Bushman, 2012; Silliciano and Lusic, 2017).

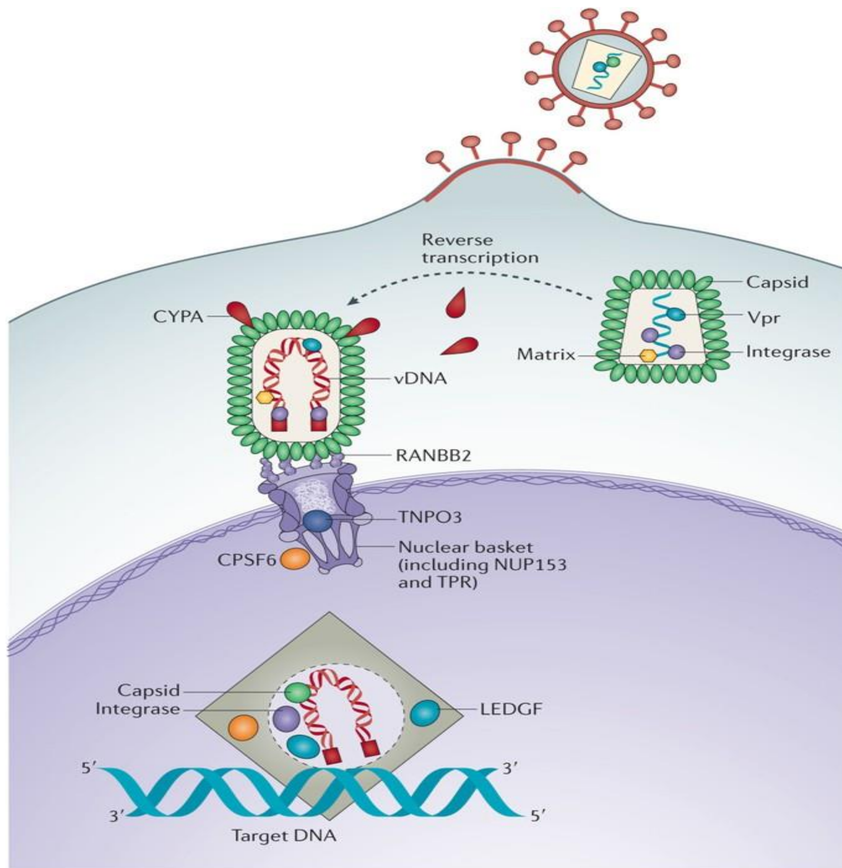


Figure 9: Importation of the PIC into the nucleus and the selection of the integration sites.
Viral and host factors that influence nuclear import are shown. (PIC-pre-integration complex. *Adopted from (Silliciano and Lusic, 2017)*

Nuclear import of PIC is an energy-requiring process under the influence of both host and viral factors. Viral components in this process include the capsid(major determinant)t, matrix, vpr, integrase, and the central polypurine tract (Silliciano and Lusic, 2017). The capsid contains a nuclear localization signal, and it also interacts with RANBP2, a nucleoporin, docking the virion to the cytoplasmic leaflet during the final stages of reverse transcription (Bukrinsky et al., 1993; Lee et al., 2010). Cellular host factors influencing nuclear PIC importation include RANBP2, transportin 3(TNPO3: also known as TRN-SR), Cyclophilin A(CYPA), and CPSF6. RANBP2 and CYPA are localized on the cytoplasmic side, and both serve as binding sites for molecules

traversing the nuclear pore complex (NPC). TNPO3 shuttles molecules through the NPC. The cleavage and polyadenylation specificity factor 6(CPSF6) also participates in the importation of the virus. Both NUP153 and CPSF6 are involved in the selection of target sites, in conjunction with LEDGF(lens-epithelium-derived growth factor) (Ciuffi et al., 2005; Koh et al., 2013; Pruss et al., 1994; Schaller et al., 2011).

2.6.3 Integration

Host genome proviral DNA integration is facilitated by viral integrase in collaboration with enzymes that function in DNA repair within the host (Pruss et al., 1994). The termini of viral DNA is the precise integration site but integration may occur at various sites within the genome of the host cells (Craigie and Bushman, 2012). Areas with high gene density and transcriptionally active domains within the host chromosomes are usually the hotspots of proviral DNA integration, as they promote efficient viral gene expression after integration (Shinn et al., 2002; Wiskerchen and Muesing, 1995). The route of entry into the nucleus, cell cycle phase, high GC content, high CpG island density, the structure of chromatin (easy access of PIC to euchromatin) as well as sequence specificities which include specific histone marks and interaction of viral integrase with cellular elements of the host such as LEDGF are some of the factors influencing the preference of transcriptional units as integration sites (Albanese et al., 2008; Lucic and Lusic, 2016; Singh et al., 2015). Recent studies have also implicated the nuclear architecture as a key mediator in the targeting of the PIC towards nuclear environment permissive for efficient viral gene expression (Silliciano and Lusic, 2017).

LEDGF/P75 mediates the targeting of the highly spliced transcriptional units by the proviral DNA, by interacting with various splicing factors to boost integration efficiency, through a tethering interaction (Ciuffi et al., 2005; Singh et al., 2015). Proviral DNA integration into the host genome finalizes the establishment of a persistent infection by irreversibly turning the host cell under transformation into a potential virus producer (Bryant and Ratner, 1990; Dean et al., 1996)

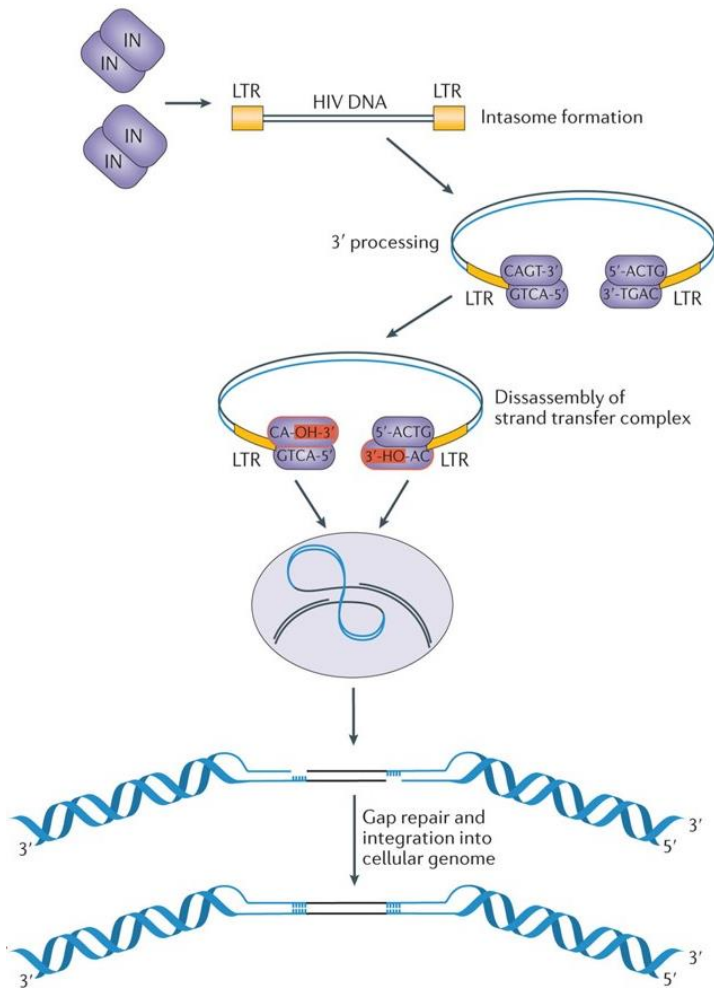


Figure 10: Mechanism of proviral DNA integration into host genome
Integrase mediates two major reactions: 3' end processing and strand transfer reactions. Integration is completed following repair of the DNA recombination intermediate by cellular and proviral enzymes. Adopted from (Silliciano and Lusic, 2017)

Analysis of the intermediate integration structures conducted by (Brown et al., 1989; Fujiwara and Mizuuchi, 1988), showed that the integration process is made up of two steps. The first involves multimerization of integrase on the viral LTR ends followed by 3' end processing. Here, two nucleotides are clipped from all 3'ends of the linear, blunt-ended viral DNA giving rise to a chemically reactive hydroxyl group. The second step, which occurs after nuclear PIC importation, involves the binding of integrase to the target DNA. In this step, phosphodiester bond pairs on opposite strands of the target DNA are attacked by the reactive 3' end hydroxyl groups. Disassembly of the integration complex follows. Cellular enzymes complete integration by repairing gaps within the single strands, and the two nucleotide overhangs at the 5' ends of the viral DNA (Freed and Mouland, 2006; Nisole and Saïb, 2004).

This stage is critical for productive infection. However, viral genetic material may have other fates. It can undergo some circularization reactions that represent dead ends for the virus. After dissociation of the PIC proteins, two ends of the viral DNA are ligated yielding 2-long terminal repeat (2-LTR) circles. Circles containing one LRT copy have also been detected and their formation has been attributed to recombination mediated by the cellular MRN complex(Mre11, Rad50, and NBS1), which functions in the sensing, targeting, and repair of double-strand breaks (Kilzer et al., 2003). These one LTR circles may also be stalled products of reverse transcription that did not undergo displacement synthesis final steps (Hu and Hughes, 2012). Furthermore, circles with inverted segments may be formed if the virus uses itself acts as an integration target (Silliciano and Lusic, 2017).

2.6.4 Transcription and translation

Tat facilitates assembly of transcriptionally active complexes following integration. It binds to the cyclin component (Cyclin T1) of (P-TEFb), increasing its affinity for the transactivation response element (TAR); a step-loom HIV-1 RNA element found at the viral genome 5' terminus (Ott et al., 2011). This resulting tri-partite complex is crucial as it dictates the assemblage of RNA Pol II machinery at the LTR. Various host transcription elements are also required for optimum LTR activation. These include NFAT (nuclear factor of activated T cells) and Rel/NF-j, whose accessibility and attachment to the promoter at LTR, dictate the result of viral transcription within host cells (Karn et al., 2012).

The CDK9, a cyclin-dependent kinase component of P-TEFb, stimulates transcription efficiency. P-TEFb phosphorylates the CTD of RNA pol II, converting it into an assembly platform. Phosphorylation regulates numerous transcription factors involved in RNA processing that modulates elongation, termination, and co-transcriptional processing of pre-mRNAs (Bieniasz et al., 1999). Moreover, negative elongation factors such as DSIF and NELF are phosphorylated by P-TEFb, inhibiting their antagonistic action and releasing Pol II from pausing at the promoter (Peterlin and Price, 2006). P-TEFb further stimulates transcription efficiency through histone phosphorylation, promoting Tat-mediated cellular transcription (O' Brien et al., 2010).

Transcription of the viral genome by cellular transcription machinery results in spliced and genomic forms of viral RNA. The portion of these pre-mRNAs that subvert the normal processing and retain their introns serve a dual role as the genomic material and precursors of the Gag and

Gag-Pol polyproteins. Splicing of the newly transcribed pre-mRNAs by the cellular splicing machinery yields two additional groups of mRNAs; singly spliced mRNAs that encode Env, Vif, and Vpu, and multiple/fully-spliced RNAs that yield Rev, Tat, Vpr, and Nef upon translation (Craigie and Bushman, 2012; Nielsen et al., 2005). The early phase of viral replication is Rev-independent while the late phase is Rev-dependent. Accumulation of the Rev protein to a certain threshold triggers the cellular splicing machinery to halt production of multiply spliced transcripts and shift to producing singly spliced transcripts, which typically occurs in the late phase. An RNA binding motif interacts with the RRE located within the *env* of all incompletely spliced viral mRNAs. The Rev-RRE complex stimulates unspliced and singly spliced mRNA export into the cytoplasm. Rev also contains an NES whose interaction with cellular cofactors such as exportin 1 and Ran-GTP facilitates nuclear export of the processed mRNAs (Nielsen et al., 2005).

2.6.5 Assembly, budding, and maturation

2.6.5.1 Viral Assembly

Assembly involves the packaging of essential components for viral infectivity, together. Virion assembly occurs at the plasma membrane. All viruses including lentiviruses such as HIV-1 utilize the C assembly pathway, where the assembly of virus particles predominantly occurs at a subcellular site. However, it was recognized that this may not always be the case since virion assembly and budding have been reported to occur in intracellular compartments such as the multivesicular body, and the late endosome in macrophages infected by HIV (Perlman and Resh, 2006).

During assembly, Gag and Gag-Pol polyprotein interact via the capsid domain. The formation of the HIV-1 virus particle is driven by the self-assembly of Gag into a curved hexameric protein lattice- whose structure is not fully understood- at the cell surface membrane (Briggs et al., 2009). There is a poor understanding of the mechanisms that direct Gag molecules to the assembly sites. However, studies have implicated An extremely basic section at the amino terminus of the MA domain that binds to phosphatidyl inositol bisphosphate in the inner leaflet cell surface phospholipid bilayer, stabilizing the interaction between Gag and the plasma membrane (Balasubramaniam and Freed, 2011). In addition, a myristic acid moiety, which is simultaneously added to terminal Gly of this N terminus during translation, has been implicated in membrane binding (Bryant and Ratner, 1990; Saad et al., 2006).

The capsid plays a crucial role during the assembly of the immature virion and virion maturation as well. The genomic RNA packaged in the assembling virion consists of two copies of unspliced, 5'capped, and 3'-polyadenylated positive, strand RNA molecules. Assembly occurs in such a way that the nucleocapsid (NC) recognizes a packaging signal known as Psi, a highly conserved RNA element within the 5'-UTR. NC additionally facilitates the formation of an RNA dimer through the palindromic dimer linkage sequence, located within the Psi sequence (Nielsen et al., 2005) The mechanistic coupling of dimerization and packaging observed in several studies favors packaging of the genome in a dimeric form, discriminating against the other forms such as a monomeric genome, spliced viral mRNAs that give rise to accessory proteins as well as cellular mRNAs(Lu et al., 2012).

Several RNAs are additionally packaged within the virion. The assembling virion also incorporates the Gag-Pol polyprotein that yields viral RT, integrase, and protease upon processing (Bryant and Ratner, 1990). The Vpu and Vif proteins partly regulate virion packaging. Virion release from the cell surface is moderated by Vpu. Vif is thought to influence the final stages of virion assembly and additionally blocks premature protease-mediated Gag processing (Nielsen et al., 2005).

2.6.5.2 Viral Budding

Viral budding describes the acquisition of a lipid envelope by the virus through fusion with the cellular membrane. The budding virus is enveloped by a host-derived membrane that bears viral envelope glycoproteins in trimeric complexes that consist of gp120 and gp41. The exit of virions from the host cells utilizes the cellular ESCRT machinery that usually mediates endosome budding into multivesicular bodies (Martin-Serrano et al., 2003). A motif in the p6 domain of Gag facilitates the virus' access to this protein sorting pathway. This motif is referred to as the 'late domain' since it aids vesicle budding and release from the lumen of late endosomes (Hurley, 2008).

2.6.5.3 Viral Maturation

Viral maturation involves the virion becoming infectious. It occurs concurrently or shortly after the budding of the immature virion from the cell. It involves cleaving of the Gag and Gag-Pol polyprotein mediated by the Gag-Pol precursor protease domain, as well as structural rearrangements in the interior of the HIV-1 particle. Gag is processed to yield MA, CA, NC, p6, and two spacer proteins namely SP1 and SP2 (Bryant and Ratner, 1990). Gag-Pol processing yields enzymatic proteins namely integrase, RT, and protease (Parkin et al., 1992). The immature virion

gag shell is predominantly made up of radially arranged Gag molecules. The most noteworthy interior structural rearrangement during maturation is the disassembly of the Gag shell.

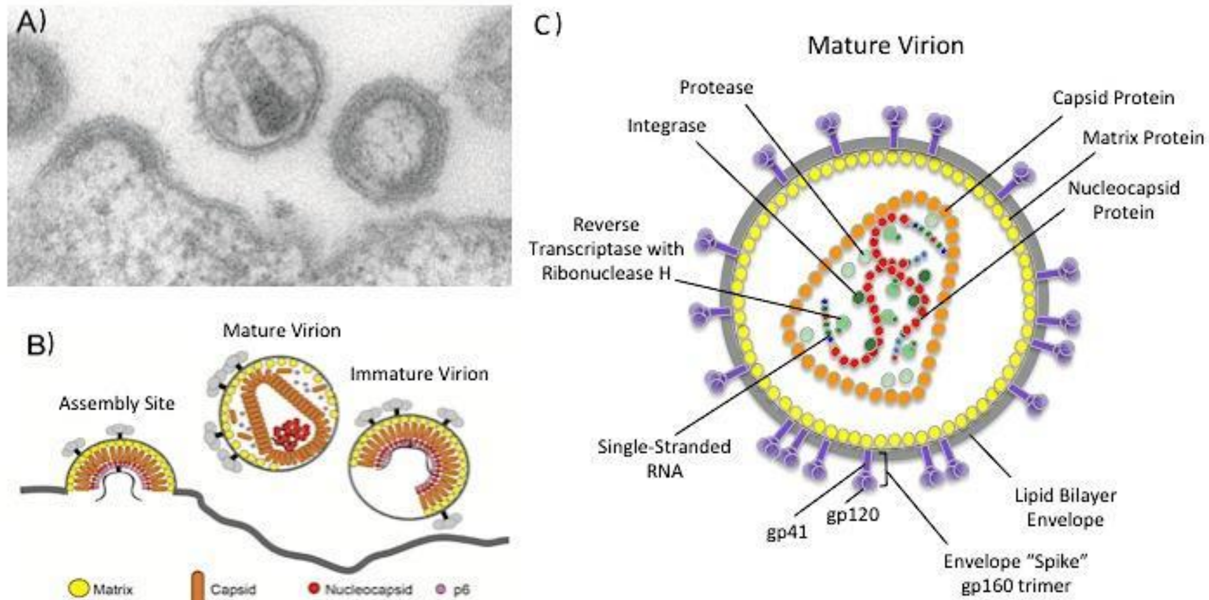


Figure 11: Cryoelectron micrograph and graphic representation of immature and mature HIV-1 particles.
(a) and (b) are cryoelectron micrographs and graphic illustrations of the assembly, budding, and maturation of a HIV-1 particle, respectively. *Adopted from: (Balasubramaniam and Freed, 2011)*

Subsequent reassembly of the CA molecules generates the conical core, while the MA remains membrane-associated and forms a matrix underneath the viral envelope (Freed et al., 1994). The viral core encloses a ribonucleoprotein complex made up of dimeric viral RNA complexed with the NC, reverse transcriptase and integrase. These processes result in a mature virus ready for another round of infection (Nielsen et al., 2005).

2.7 Cellular receptors interacting with HIV env.

A thorough understanding of the molecules utilized by HIV-1 for binding and access into the target cells is pertinent to the development of effective prophylaxes. The CD4 receptor is most crucial for HIV infection. MABs studies by (Dalgeish et al., 1984; Klatzmann et al., 1984) showed that transfecting CD4 deficient human cell lines with the CD4 gene, conferred susceptibility to infection by the HIV virus. Nevertheless, induced CD4 receptor expression in other mammalian cell lines such as mice resulted in non-productive viral infections. The incompatibility of the results indicated the possible presence of other specific proteins in the human cell line that propagated the viral infection. Further studies led to the discovery of secondary receptors utilized by HIV-1 to gain access into the target cells. The secondary receptor used for entry of the HIV virus into target cells was discovered by (Feng et al., 1996). This was a 7-transmembrane G-coupled receptor that was initially termed fusin because it was crucial in the merging of the viral and target cell membranes. Later studies by (Berger et al., 1999) and others, used the wider chemokine nomenclature to label these coreceptors CXCR4 and CCR5. Other entry cofactors include CCR2b, CCR3, CCR8, and US28-also members of the seven-span transmembrane chemokine receptor family and chemokine receptor-like orphan molecules such as STRL33 or BONZO or TYMSTR, GPR15 or BOB, and V28 (Agrawal et al., 2004; Dragic et al., 2000).

The CD26 also referred to as dipeptidyl peptidase IV, has also been implicated as a cofactor that interacts with HIV env. This protein is expressed on memory and helper CD4+ T cell subsets and is said to exert its protease activity on a particular motif within a highly conserved region of the V3 of gp120 (Kameoka et al., 1993). Studies by (Callebaut et al., 1993) also showed that CD26 MAbs inhibit viral entry and murine NIH 3T3 fibroblasts are rendered permissive to HIV infection when CD4 and CD26 are co-expressed. Further studies on CD26 revealed a positive association between the HIV infection rate and expression levels of CD26 CD4+ T cells, making CD26 a possible antiretroviral target (Callebaut et al., 1997). Contrary investigations by (Mattern et al., 1991; Ohtsuki et al., 2000) demonstrated conflicting evidence in which the rate of HIV infection and expression levels of CD26 were negatively correlated, making CD26 a controversial drug target.

Various integrins also mediate HIV infection. Research has demonstrated that HIV attaches to an integrin termed $\alpha_4\beta_7$, expressed on CD4+ T helper cells in the gut (Alvarez et al., 2013; Arthos et al., 2008). This attachment is said to increase susceptibility to infection but conflicting evidence

by (Ding et al., 2015) suggested a mechanism of action that is independent of attachment. Another integrin known as Lymphocyte function-associated antigen 1(LFA-1) binds with high affinity to its ligand (ICAM-1), which may be integrated into the envelope of the virus during budding (Fortin et al., 1998; Kondo and Melikyan, 2012). Furthermore, studies have shown that this cofactor can be sufficiently activated by gp120 (Hioe et al., 2011). Using an in vitro and an in vivo mouse model, monoclonal and single domain antibodies against the LFA-1 beta subunit have been demonstrated to effectively block transmission of the virus (Guedon et al., 2015).

2.7.1 CD4 receptor

The CD4 receptor was identified as a HIV-1 receptor is 1984. (McDougal et al., 1986) carried out experiments in which radiolabeled, HIV-1 treated CD4 T cells were exposed to two MABs that bind to different CD4 epitopes. One of the two MABs was incapable of binding to the HIV-1 treated cells and further antibody-antigen analyses confirmed that the CD4 receptor rendered cells permissive to env-mediated (gp120) HIV fusion, entry, and ultimately infection by the HIV virus (Fauci et al., 1996). Thus, CD4 expressing T cells -mostly immune cells such as different subsets of CD4+ T cells including the naïve, central, and effector memory subtypes, macrophages and dendritic cells exhibit robust infection (Bour et al., 1995). This receptor is a commencement point of a signal cascade that activates these cells with respect to immunological stimuli. In HIV infection, binding of env gp120 to CD4 initiates entry of the virus into the target cells (Seitz, 2016). Gp120 binds to the CD4 receptor via the CD4 domain of gp120, inducing a conformational change in both proteins that ultimately exposes the V3 loop. HIV infection down-regulates CD4+ T cells-CD4 expression (Wilen et al., 2012b). Decreased expression of this receptor has been correlated with the emergence of opportunistic infections and rapid disease progression ultimately resulting in AIDS (Langford et al., 2007). CD4 also functions as a receptor for simian immunodeficiency viruses which share close homology with HIV (Bour et al., 1995).

CD4 is a member of the immunoglobulin superfamily. It is a 58kDa protein made up of an extracellular region 370 amino acids long, a transmembrane region made up of 25 amino acid residues, and a cytoplasmic tail 38 amino acids long. The extracellular region folds into four domains (D1-D4). The D1 domain localizes to the N-terminal and is homologous in structure and sequence to the variable region of immunoglobulin (Ig) light chains. The primary structures of D1, D2, and D3 are less homologous to that of Ig molecules but fold in a similar way to Ig family domains. CD4 also undergoes various post-translational modifications such as the addition of two

N-linked glycans between D3 and D4, and disulfide bonds to reinforce the stability of D1, D2, and D4 domains (Bour et al., 1995).

The recognition of CD4 as the principal HIV receptor was the driving force behind a class of ART known as entry inhibitors (EIs). Initial approaches proposed the use of MAbs to block viral binding. This strategy was however flawed due to the critical role of CD4 in basic immunological functions. In the following years, soluble CD4s (sCD4) capable of inhibiting HIV-1 replication both in vitro and in vivo, as well as a soluble CD4-immunoglobulin fusion protein were developed. Primary HIV-1 isolates exhibited significantly lowered sensitivity to sCD4 neutralization compared to lab isolates, with the sCD4 treatment enhancing infection in some cases (Schooley et al., 1990). The ultimate observation was that the administration of therapeutic sCD4 did not affect viremia or disease (Collier et al., 1995).

2.7.2 Coreceptors

The discovery of the coreceptors mediating HIV-1 entry was aided by studies demonstrating that the replication of the virus could be inhibited by previously unknown, leukocyte-derived, soluble suppressor factors (Brinchmann et al., 1990). These soluble factors derived from CD8+ T cells were identified as the C-C chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) (Cocchi et al., 1995). Chemokines are small paracrine signaling molecules associated with the inflammatory response. Four main chemokine groups have been identified thus far. The nomenclature of the chemokine classes is based on the number and orientation of the N-terminal cysteine motifs. C chemokines have one cysteine residue. C-C chemokines, C-X-C chemokines, and C-X3-C chemokines each have two cysteine residues, separated by 0, 1, or 3 other residues, respectively (Bachelerie et al., 2014). Only the C-C chemokines and C-X-C chemokines have been recognized as additional factors critical for HIV-1 binding, and entry into the host target cell. This process occurs in three steps and involves sequential interaction of the viral envelope glycoprotein trimer Env with the CD4 receptor, and either of the two host chemokine coreceptors CCR5 or CXCR4 (Dragic T et al., 1996).

The CCR5 and CXCR4 chemokine coreceptors are structurally related members of the 7-transmembrane G-protein coupled receptors superfamily. These receptors contain an N-terminus, and 3-extracellular loops both located in the extracellular leaflet of the cell, and C-terminus located on the cytoplasmic aspect of the cell (Coakley et al., 2009).

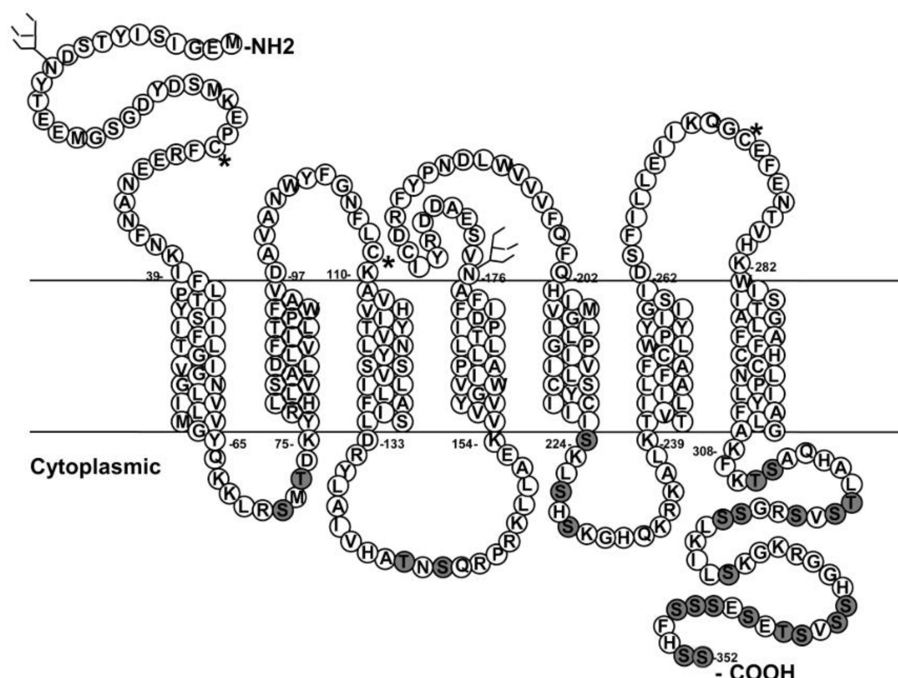


Figure 12: Schematic representation of CXCR4
Potential phosphorylation sites (a total of 25) are indicated by the shaded residues. Potential N-linked glycosylation sites are marked. Cysteine residues involved in the formation of disulfide bonds are designated by asterisks**.** *Adopted from* (Coakley et al., 2009)

GPCRs activated by agonists rapidly phosphorylate at serine and threonine residues within the cytoplasmic tail and the third intracellular loop (Berger et al., 1999). The sulfated N-terminus and the second extracellular loop of CCR5 are examples of host factors implicated in tropism determination (Coakley et al., 2009). The V3 loop of gp120 principally associates with the second extracellular loop (ECL2) region. The bridging sheet which interacts with the amino terminus of the coreceptor following CD4 binding is formed by the C1, C2, and C4 domains of g-120 (Sterjovski et al., 2010)

Replication kinetics and cytopathology *in vitro* of a particular HIV-1 strain are largely defined by the coreceptor utilized by the specific strain. Coreceptor tropism is additionally the key indicator of pathogenicity, tissue tropism, and transmissibility of the virus *in vivo* (Pillai et al., 2003). Therefore, understanding coreceptor tropism is crucial for disease progression monitoring and for ART support that involves a new class of drugs termed coreceptor antagonists (Sander et al., 2007). CCR5 is the major coreceptor used for entry by a majority of HIV strains and predominates early in infection (De Jong et al., 1992), it is also the principal coreceptor expressed by sexually

transmitted HIV-1 variants (Dragic T et al., 1996; Kawamura et al., 2003) fueling the development of entry inhibitors such as the CCR5 antagonist, Maraviroc among others (Dorr et al., 2005). X4 and dual tropic R5X4 variants, emerge in approximately 50% of individuals progressing towards a high level of immune suppression (Berger et al., 1998; Coakley et al., 2009; Hartley et al., 2005). These viruses are dominant during the advanced stages of infection and this coreceptor switch is generally correlated with poor clinical prognosis (Connor et al., 1997; Pastore et al., 2006a), and has been suggested as a likely cause of resistance to CCR5 inhibitors (Pastore et al., 2006a). Tropism switching by HIV-1 isolates from CCR5 to CXCR4 is correlated with poor clinical prognosis (Connor et al., 1997) Coreceptor utilization and switching has been studied broadly for clade B isolates which are prevalent in North America and Western Europe. However, viruses of non-clade B nature now comprise most of the new HIV-1 infections and should therefore be the focal point of vaccine and drug development efforts. Unfortunately, the biological, and immunological properties of the other types of isolates remain largely unknown.

2.7.2.1 Chemokine coreceptor-5 (CCR5)

CCR5 was initially isolated as a three CC antagonized, GPCR (Blanpain et al., 1999). This receptor is commonly expressed in several immune cells implicated in the inflammatory response since protective cellular and humoral responses, and intracellular trafficking are regulated by the binding of chemokines to the CCR5 coreceptor (Berger et al., 1999; Zhong et al., 2018).

CCR5 is the prime co-receptor utilized for entry by the majority of HIV strains that predominate early in infection (Schuitemaker et al., 1992). HIV-1 viruses utilizing this coreceptor are termed R5 viruses. One of the initial designations of HIV viruses was based on their in vitro cellular host range and placed the variants into two subgroups; T-tropic and M-tropic (Hwang et al., 1991). R5 viruses were initially assumed to be M-tropic, non-syncytium inducing owing to the lack of CCR5 expression on most transformed CD4 T cell lines. However, this grouping was flawed because SI phenotype and coreceptor usage are not synonymous. Most primary R5 viruses that could replicate in CD4+T lymphocytes, could not replicate in macrophages (Sterjovski et al., 2010). Since viral coreceptor utilization is not a singular determining factor of viral tropism, R5 variants are now classified as R5 M-tropic or R5 T-tropic (Berger et al., 1998; Robertson et al., 2000). The expression of the CCR5 coreceptor is heterogenous in CD4+ T cells with different subsets of these cells showing differential expression of this co-receptor. CCR5 is preferentially expressed by memory T cells-central effector memory CD4+T cells (Zaunders et al., 2002), gut-associated

lymphoid tissue (GALT), microglia, macrophages, and dendritic cells (Zhang et al., 1998). It is also expressed on T follicular helper cells(T_{FH}) (Allam et al., 2015) and effector memory T cells(T_{EM}) (Groot et al., 2006).

In addition to predominating early in infection, CCR5 viruses are preferentially transmitted by all routes for reasons still not clearly understood. Studies have attributed the preferential expression of this coreceptor by sexually transmitted HIV-1 variants to the presence of $CCR5^+$ macrophages and DCs on mucous membranes of the genital and gastrointestinal tract (GALT), that transport HIV to the regional lymph nodes and mediate contact with, and infection of activated T cells (Dragic, 2001; Kawamura et al., 2003). R5 dominance in the initial stages of HIV infection is thus attributed to infection of macrophages at the inoculation site which then spread the virus to other body parts and, the relatively higher turnover rate of memory CD4+ T cells, the target cells of R5 viruses, as compared to naïve CD4+T cells (Zaunders et al., 2002).

Coreceptor switching from R5 to X4 or dual R5X4 is observed in up to 50% of patients progressing to AIDS and is linked to faster progression of the disease. Studies carried out by (Gorry and Ancuta, 2011) and others revealed that R5 viruses isolated from subjects in the AIDS stage exhibited increased tropism to macrophages, better viral fitness, enhanced ability to scavenge CD4 and CCR5, decreased sensitivity to CCR5 antagonist, and higher efficiency in CCR5 usage as well as increased ability to induce CD+ 4 cell apoptosis, compared to isolates form subjects in chronic stages of HIV-1 infection (Sterjovski et al., 2007; Wade et al., 2010).

One of the most important factors in disease susceptibility and resistance is host genetic diversity. Studies have shown that CCR5 polymorphism affects HIV-1transmission and disease progress. The locus of one such polymorphism is in the CCR5 promoter region at position 59029. This polymorphism influences disease progression by regulating CCR5 expression. Here, an A/G transition upregulates CCR5 expression affecting the pathogenesis of AIDS, while GG individuals progress slower to AIDS than the AA homozygotes (Masankwa et al., 2016).

Perhaps the most important polymorphism with respect to CCR5 and its role in HIV-1 pathophysiology is the 32-bp deletion in the CCR5 gene (CCR5Δ32). The occurrence of this mutation is quite common in some Western Europe populations (Agrawal et al., 2004). CCR5Δ32 homozygotes are estimated at 1% and are extraordinarily resistant to HIV-1 infection because the mutation results in the expression of a dysfunctional CCR5 chemokine receptor, hindering the

productive entry of the virus into target cells (Agrawal et al., 2004; Naif, 2013; Nasir et al., 2015). Resistance in CCR5Δ32 homozygotes has been attributed to the genetic loss of cell surface CCR5 expression and active mutant CCR5 protein-mediated down-regulation of CXCR4 expression (Agrawal et al., 2004, 2007; Jin et al., 2008). This genetic defect, however, does not confer absolute protection to HIV-1 infection as there have been reports of infection in hemophiliac patients and some CCR5Δ32 homosexuals (Balotta et al., 1997; Sheppard et al., 2002). An alternative theory suggests that CD4+ T cells in these individuals are protected from infection by the autocrine production of CCR5 ligands (Casazza et al., 2009; Samson et al., 1996). These proteins act as HIV-suppressive factors by inducing changes in the stoichiometry of molecules that mediate HIV-1 entry into cells. The progression of aids is also slower in CCR5Δ32 heterozygotes. An obvious phenotype is not conferred by the mutant allele (Coakley et al., 2009), however, the expression and stability of the defective, prune CCR5Δ32 is crucial for the resistant phenotype (Agrawal et al., 2007).

CCR5 has become a principal target of therapeutic and preventative solutions for the HIV-1 pandemic owing to its critical role in HIV-1 entry. Biochemical blockade of the CCR5 coreceptor by CCR5 antagonists is the principal mechanism of action behind most available therapeutic options but there is growing interest in gene editing therapies. A radical therapeutic approach is a gene-editing knockout to resemble the naturally occurring CCR5Δ32 deletion. The success of this strategy is evident in the 'Berlin-Patient', a seropositive individual who was rendered HIV-1 free, after a double CCR5Δ32 stem cell transplant (Brown, 2015; Hütter et al., 2009). Other gene editing techniques targeting the CCR5 receptor include the CRISPR/Cas9 nuclease system and transcription activator-like effector nuclease, short hairpin RNAs, and ribozymes as therapeutic targets for HIV-1 (Scarborough and Gatignol, 2015; Swamy et al., 2016)

Biochemical blockade therapeutic options exploiting the CCR5 coreceptor include entry inhibitors such as Maraviroc. Maraviroc is a CCR5 antagonist approved by the FDA in 2007, that competitively inhibits entry of HIV-1 into target cells by acting as a negative allosteric modulator of the CCR5 coreceptor (Dorr et al., 2005). Two other CCR5 entry inhibitors were evaluated in clinical trials in vain. Viral rebound was observed in treatment naïve patients administered with the non-competitive allosteric inhibitor, Vicriviroc (Tagat et al., 2004; Westby and Van Der Ryst, 2010), while treatment-experienced patients exhibited increased malignancies (Wilkin and Gulick,

2012). Aplaviroc resulted in severe hepatotoxicity in seropositive patients in phase II of the clinical trials (Nichols et al., 2008). Ceniciviroc, an inhibitor of both CCR5 and CCR2, exhibited good oral absorption (Seto et al., 2006). Protein-based CCR5 antagonists that resemble the natural CCR5 ligands have also been successfully developed (Vangelista et al., 2008). Other protein-based approaches include mAbs against CCR5 such as PRO14-a murine-derived humanized IgG4 mAb currently in phase III trials, that efficiently inhibits the binding of gp120 to CCR5 (Reichert, 2017), and RoAb13 (Ji et al., 2007)

2.7.2.2 CXCR4

CXCR4 was initially identified as an orphan receptor known as leukocyte-derived seven-transmembrane domain receptor(LESTR) (Loetscher et al., 1994) but did not come into the limelight until it was isolated as HIV-1 coreceptor and its natural ligand SDF-1/CXCL12 was discovered too (Dragic, 2001; Dragic T et al., 1996).

Compared to CCR5, CXCR4 is broadly expressed on many hematopoietic cells including hematopoietic progenitor cells, T cells (naïve and resting CD4 lymphocytes, and CD8 cells), B cells, neutrophils, and eosinophils (Berkowitz et al., 1998; Zhang et al., 1998). HIV-1 isolates capable of utilizing CXCR4 either alone (X4 HIV) or in combination with CCR5 (R5X4) emerge in at least 40%-50% of individuals progressing towards a high level of immune suppression (Berger et al., 1998; Coakley et al., 2009; Hartley et al., 2005). Such a coreceptor switch expands the pool of target cells and is associated with a poor prognosis (Pastore et al., 2006a).

The exact mechanisms governing coreceptor tropism switching remain elusive. However, three main hypotheses that consider the selection pressure exerted on the population of HIV viruses within the host and how the virus interacts with the host's defense system have been suggested. The first attributes the emergence of the X4 and R5/X4 viruses to reduced viral fitness of R5 viruses (Coetzer et al., 2008). This theory cements the importance of R5 variants in transmitting HIV, attributing the appearance of X4 variants to gradual mutations in the founder R5 strains. In addition, X4 viruses are dominant during later stages of infection. This is attributed to the low intrinsic turnover rate of naïve CD4+T cells (Fenyö et al., 2011; Mclean and Michie, 1995) and is usually indicative of disease progression to AIDS. Tropism switching usually occurs after several years of infection. This is unexpectedly slow since this switch requires only a limited number of mutations such as changing a few residues in gp120, to occur (Zaunders et al., 2002). Mutations

in the V1/V2 and V3 regions of gp120 are imperative for coreceptor switching in a majority of subtypes: A, B, D, CRF01_AE, and CRFO2_AG (Palm et al., 2014). However, the genetic barrier in subtype C is almost impenetrable, requiring substantial changes in regions outside of the V3 (Coetzer et al., 2011). These changes are essential because they make up for the loss of fitness changes associated with tropism switching in the V3 loop, and may explain the predominance of CCR5 viruses in HIV-1 subtype C infections (Jakobsen et al., 2010). Moreover, (Kato et al., 1999) studied a chimeric panel of HIV-1 subtype E viruses in which naturally occurring mutations were systematically introduced at positions 8, 11, and 18 of the V3 loop of an X4 tropic strain belonging to the subtype E clade. Following infectivity assays on an array of cell lines that naturally function as targets of the virus, position 11 was found to be critical for CCR5 utilization since a substitution here resulted in a uniform loss of infectivity in the CCR5 cell lines utilized. Conversely, a minimum of two arginine substitutions, irrespective of the combination, conferred CXCR4 tropism.

The second mechanism is contingent on the host immune system and suggests a relatively higher susceptibility of R5 viruses to immune responses as compared to their X4 counterparts (Fenyö et al., 2011). The appearance of X4 or R5X4 isolates indicates a pronounced reduction in CD4+ T cell numbers and faster development of the disease. Studies have suggested the rapid disease progression may be due to elevated cytopathogenicity or increased ability to infect and induce CD4+ T cell depletion (Zaunders et al., 2002).

The third mechanism is the target cell hypothesis that correlates the phenotype of the virus to the host's pool of target cells (Fenyö et al., 2011). R5 and X4 variants exhibit distinct pathogenicity in vivo and this has been attributed to different patterns of coreceptor expression within the host cells (Berkowitz et al., 1998; Zaunders et al., 2002). The CXCR4-expression pattern on many hematopoietic cells increases viral access to a crucial pool of target cells belonging to the T cell ontogeny, contributing to the rapid drop in CD4 T cell number (Blaak et al., 2000). A reduced CD4+T cell population may be both stimulus and response to the dominance of X4 viruses, since a low CD4+ cell count results in the reduction of CCR5 targets and or abolished immune suppression on CRCX4 or X4 variants (Zaunders et al., 2002).

Another emerging hypothesis has proposed recombination as a driving force of coreceptor usage switch (Mild et al., 2007). Here, co-existing R5 and X4 viruses undergo recombination in which

and R5-derived *env* V1-V2, suggested to confer resistance to NAbs, and an X4-derived *env* V3 loop are exchanged (Pinter et al., 2004).

2.8 HIV-1 tropism determination

Viral tropism refers to the propensity of a virus to infect a particular host species, tissue, and cell type (Cardozo et al., 2007). Coreceptor usage is a principal viral tropism determinant in HIV infection and is mediated by the V3 loop of gp120 (env). Changes in the V3 loop sequence can therefore modify viral cell tropism, neutralization by antibodies, neutralization of soluble CD4, formation of syncytium, and usage of chemokine receptors (Stanfield et al., 1999).

Coreceptor usage by HIV-1 may be assessed by genotypic and phenotypic methods. Phenotypic assays rely on culturing infected cells from the host or engineering recombinant viruses using patient-derived env sequences (Raymond et al., 2010; Resch et al., 2001). Examples include the Trofile Assay, PhenoScript assay, and many others (Whitcomb et al., 2007a). Limitations of phenotypic tests such us cost and long turn-around time prompted the development of relatively cheaper and faster genotypic tropism testing methods. Since the V3 has been implicated as the major determinant of viral tropism, *in silico* genotypic tropism determination methods rely on the amplification of env gene, from both viral RNA and DNA and subsequent interpretation of these sequences by one or more bioinformatics algorithms and tools such as the 11/25 rule, support vector machine (SVM), position specific scoring matrices (PSSM) among others (Beerenwinkel et al., 2003a; Chiappetta et al., 2014; Mortier et al., 2013). Phenotypic tropism determination techniques are normally used to validate genotypic predictions (Cabral et al., 2012).

The NIH provides a set of guidelines that dictates the criteria for the determination of coreceptor usage under clinical conditions. Antiretroviral support therapy that involves coreceptor antagonists has a prerequisite for tropism determination, irrespective of prior tropism testing that showed CCR5 usage. This eliminates coreceptor switching as one of the pathways leading to resistance to CCR5 antagonists, resulting in virologic failure.

2.8.1 Phenotypic assays

The pioneer phenotypic tests relied on the identification of syncytium formation by HIV-1 on MT-2 cells while the modern assays employ the use of CD4/CCR5 or CD4/CXCR4 expressing cell lines to detect viral infection (Tremblay et al., 2013). The latter assays utilize viral genetic material derived from the patient's plasma or PBMCs. The assays utilizing plasma samples require a plasma HIV RNA level of about >1000copies/ml. The patient-derived envelope sequence(s) are amplified and cloned into vectors to create replication competent or replication defective recombinant viruses or pseudoviruses. CD4+ T cells engineered to express CCR5 or CXCR4 such as U87-CD4-

CCR5, U87-CD4-CXCR4 respectively, are then transfected using these constructs. Viral tropism is consequently deduced by the expression of a reporter gene which may be delivered by the pseudovirus or one that functions as a responsive cellular gene such as luciferase or B-galactosidase. Infection is then quantified by bioluminescent or colorimetric signals (Van Baelen et al., 2007).

Sample failure during phenotypic tropism testing has been attributed to amplification difficulties due to sample degradation as well as low viremia, reduced viral fitness, and reduced infectivity of the recombinant virus. In addition, studies have shown that pseudo-typing efficiency can be affected by interactions between the gag and env proteins of the viral vector acquired from the patient and using varying reporter cell lines. Studies by (Reeves et al., 2002) implicated the use of dissimilar reporter cells as a cause of disparity between two phenotypic assays. This was attributed to differential coreceptor expression as well as different receptor ratios in engineered cell lines such as the human glioma cell lines.

There is an unending debate as to which methods are suitable for routine clinical use. Phenotypic assays have several limitations in that the assays are expensive, have long turn-around times, and require huge amounts of fresh specimen. Moreover, these techniques are only feasible in sophisticated labs of BSL-3 (Skrabal et al., 2007; Vandenbroucke et al., 2010). Furthermore, substantial discordance is observed between the assays when they are applied to clinical samples, which are generally more heterogeneous. Population-based phenotypic assays are also limited in their ability to distinguish populations of viruses comprised of mixtures of coreceptors with different phenotypes from exclusively dual/mixed tropism viral populations (Skrabal et al., 2007). These assays also have a minimum viral load (>1000 copies/mL) requirement, excluding low viral load patients with recent virological failure and patients that may require alternative therapy due to tolerability factors despite having a viral load that is below the threshold (Tremblay et al., 2013). To address the limitation brought about by viremia, (Baumann et al., 2015) and others explored the possibility of using HIV DNA in the event of undetectable viral loads, but the assays have not been validated for clinical use.

Discordance observed between phenotypic assays and genotypic prediction tools may hamper the usefulness of the latter tools since clinical decisions on coreceptor antagonist use are based on the detection of X4 viruses (Garrido et al., 2008; Skrabal et al., 2007; Vandenbroucke et al., 2010).

Variation in HIV-1 quasispecies has also been partially implicated in this reduced concordance (Poveda et al., 2009).

2.8.1.1 The Trofile Assay

The Trofile HIV Coreceptor Tropism Assay previously known as the PhenoSense HIV-Entry assay from Monogram Biosciences(South San Francisco, USA) is a CLIA-validated(Clinical Laboratory Improvement Amendments-CDC), plasma-based, phenotypic HIV tropism determination assay widely used in clinical settings (Skrabal et al., 2007; Vandenbroucke et al., 2010).

Here, patient-derived envelope sequences (i.e. gp120 and gp41) of HIV are amplified by RT-PCR and incorporated into expression vectors. An env-deficient HIV vector harboring a reporter luciferase gene is then mixed with or used in the transfection of expression vectors. The resulting pseudoviruses and the patient-derived env genes are deficient in terms of replication. These pseudoviruses are used to infect cells $CD4^+$ cell lines expressing CCR5 or CXCR4. These cells are subsequently exposed to a luciferase substrate and coreceptor usage determined by luminescence or lack thereof in the different cell types. If luminescence is observed in $CD4^+$ CCR5 expressing cells, the patient harbors R5 variants, if only $CD4^+$ CXCR4 expressing cells emit a signal the patient harbors X4 variants, and finally, the patient is said to harbor dual/mixed tropic viruses of both $CD4^+$ $CCR5^+$ and $CD4^+$ $CXCR4^+$ exhibit luminescence (Trouplin et al., 2001; Whitcomb et al., 2007b).

The Trofile Assay was vastly used during the MOTIVATE clinical trial to validate the efficacy of Maraviroc (Vandenbroucke et al., 2010; Westby et al., 2006). Rapid virologic failure was observed in some patients following enrollment and initiation of a CCR5 antagonist. This was attributed to the presence of X4 variants present in amounts lower than the detection threshold of the assay (Westby et al., 2006).

The Trofile Assay has now been superseded by a better technique known as the Enhanced Sensitivity Trofile Assay(ESTA) which has an improved, approximately x30 increase in sensitivity to X4 viruses in clinical isolates (Cabral et al., 2012). ESTA was validated by several studies carried out by (Prosperi et al., 2010; Saag et al., 2008) amongst others which demonstrated this assay's superiority to the Trofile assay with regard to the accuracy, precision, reproducibility, and improved sensitivity to X4 variants; 100% sensitivity when applied to samples containing as low as 0.3% X4 variants. ESTA was implemented as a clinical substitute for the original Trofile assay

in June 2008.

2.8.1.2 Virco Tropism Assay

Virco Tropism Assay is a population tropism phenotypic test used to verify genotypic tropism predictions. Briefly, one step RT-PCR is used to generate V4 gp120 amino terminal amplicons. The amplicons are then purified and cloned into pHXB2D-ΔNH_2-V4-eGFP vectors by in vitro recombination. Recombination mixes are then used to transform cells. Recombinant plasmids are purified from the transformed cells and used to transfect 293T cells. These cells are then cultured for 48 hours and recombinant virus stocks are subsequently harvested. An aliquot of recombinant virus stock is then added to U87-CD4, U87-CD4-CXCR4, and U87-CD4-CCR5 cells. After five days, an argon laser-scanning microscope is used to evaluate infection through the expression of EGFP (enhanced green fluorescent protein) (Van Baelen et al., 2007).

2.8.1.3 Phenoscript Env Assay

This assay has been validated with respect to sensitivity to minority X4 variants and is particularly specific for samples containing variable viral loads and distinct subtypes of HIV-1. In this assay, a patient-derived env gene from plasma samples is amplified. Env amplicons and a deleted NL-43 plasmid undergo homologous recombination to produce recombinant plasmids, which are then used to transform producer cells resulting in the generation of recombinant viruses. CD4+ indicator cell lines expressing either the CCR5 or CXR4 coreceptors, and an additional lacZ reporter gene under the control of a viral LTR are infected with the recombinant viruses. Infection specificity is further assessed by incubating indicator cells with or without co-receptor antagonists (Seclé et al., 2010).

2.8.1.4 MT-2 Assay

Tropism determination by MT-2 assay relies on PBMCs and expression of the CXCR4 coreceptor at the cell surface (Coakley et al., 2009). This assay requires freshly obtained or cryopreserved viable patient cells and must be conducted in a BSL-3 laboratory. This assay assesses the likelihood of HIV-1 isolates to prompt syncytium formation in an *in vitro* model.

There are two commonly used adaptations of the MT-2 assay. The first assay entails direct coculturing of patient cells with MT-2 cells. In the alternative version, patient-derived cells are first cocultured with seronegative peripheral mononuclear blood cells (PBMCs) stimulated phytohemagglutinin to generate virus stocks (Japour et al., 1994). Studies by (Koot et al., 1992)

showed that the second approach limits the assay's sensitivity since PBMCs are less sensitive to infection by X4 variants than MT-2 cells. Furthermore, a change in the relative portions of viral subpopulations was observed in studies by (Voronin et al., 2007) when the patients' samples were passaged in PMBC prior to inoculation. The inoculated MT-2 cell cultures are then monitored every 2-3 days up to 2 weeks, for the development of typical cytopathic effect (CPE) indicated by the formation of large ballooning syncytia. This assay is normally performed in a 96-well, flat bottom culture plate. The major drawback of this assay is the imperfect correlation between syncytium induction phenotype and CXCR4 coreceptor use.

2.8.2 Genotypic *in-silico* prediction algorithms

Tropism determination using algorithms and bioinformatics programs are practical surrogates to phenotypic tropism assays in routine clinical practice. They are used as HIV-1 tropism inference tools in clinical settings lacking the resources to carry out phenotypic tests (Poveda et al., 2009). These genotypic predictive models are potentially cheaper and circumvent intrinsic confounders of phenotypic assays such as inadequate standardization among the labs carrying out these assays, differences in target cell activities, non-objective interpretation of cytopathic changes, and a requirement for stimulated or pseudotyped viruses. These tools also exhibit relatively high specificity but modest sensitivity to X4 variants when compared to phenotypic assays (Jensen A. Mark, Wout Angelique, 2003). Some authors argue that these in-silico methods may be able to predict phenotype shifts prior to the appearance of biological markers since they differentiate levels of CXCR4 usage by continuously scoring V3 loop sequences (Jensen et al., 2006). The reduced sensitivity of these tools to X4 viruses has been attributed to the possibility of other regions beside the V3 being co-determinants of coreceptor usage, and the limited detection of minority species even with the expansion of sequences into innumerous feasible permutations (Pastore et al., 2006a; Suphaphiphat et al., 2007).

Data sets consisting of HIV-1 subtype B and C clonal sequences were used in the development and validation of bioinformatics coreceptor predictors. Thus, the high degree of sequence heterogeneity observed in clinical isolates is a major confounding effect when utilizing these tools as tropism predictors. In addition, the lack of Gold standards for reference and validation purposes which results in misleading predictions, as well as their reduced ability to consistently detect minority variants in data derived from clinical sequences, hinders their adoption for regular clinical use (Skrabal et al., 2007). The sensitivity of genotypic tools to X4 viruses was found to be only 50% when these tools were applied to one of the largest studies conducted by (Low et al., 2007), that examined 903 V3 sequences from drug-naïve individuals whose phenotypic coreceptor usage information was already available. In conclusion, it was postulated that other regions outside of the V3 were minor tropism determinants, and expanding the analysis to these regions might considerably improve the sensitivity of these tools to X4 viruses (Poveda et al., 2009). Despite these shortcomings, enhanced prediction accuracy and a substantial increase in sensitivity and specificity were observed in a number of study models (Garrido et al., 2008; Poveda et al., 2009; Raymond et al., 2017), that combined three or more of these algorithms. It was suggested that

these tools can be improved further by supplementary data such as clinical status, CD4 counts (Low et al., 2007).

Examples of these algorithms and bioinformatic programs are discussed below.

2.8.2.1 Net charge rule

This tool is based on the global net charge of the V3 loop. If the overall charge of the V3 sequences is ≥ 5; otherwise R5 (Seclé et al., 2010). Coreceptor usage prediction using the charge rule does not always align with experimental tropism determination. This inaccuracy has been attributed to the relatively scant and unsubstantiated data that was utilized in the development of this assay (Pillai et al., 2003). The net charge of the V3 loop has been explored in conjunction with other properties of the V3 loop to develop algorithms for coreceptor usage prediction. This tool will be therefore utilized for genotypic in silico tropism prediction in this data to determine how this characteristic of the V3 loop affects the viral tropism of the isolates in our data set.

2.8.2.2 '11/25 or 11/24/25' Rule

This method is based on Sanger sequencing. It relies on the charge rule which associates an increase in the net positive charge with CXR4 usage, since the presence of a positively charged amino acid at positions 11 and or 25 conferred docking ability on CXCR4 coreceptors, with CCR5 binding occurring as the default (Fouchier et al., 1992; De Jong et al., 1992).

Over the years, studies have revealed that the $\beta 2$–$\beta 3$ loop in the CXC and CC chemokines- the natural ligands of CXCR4 and CCR5 respectively- and the crest of the V3 loop are structurally homologous. This served as the basis of 3D modeling of the V3 loops from primary HIV-1 isolates whose tropism had been phenotypically determined. Amino acid residues at positions 11 and 25 have been previously shown to impact HIV-1 tropism and thus the charge of these particular is often used to infer tropism (Cardozo et al., 2007). When these V3 amino acid sequences and their corresponding coreceptor profiles were initially analyzed, the results suggested that the occurrence of positively charged amino acid occurred at positions 11 and or 25 of the V3 loop conferred an X4 phenotype, while the default phenotype was R5 (Fouchier et al., 1992). Variants are thus categorized as syncytium inducing (SI) or X4, if a positively charged amino acid such as lysine or arginine, is present at positions 11 and 25 of the V3 loop, and non-syncytium inducing (NSI) if not (Fouchier et al., 1995; De Jong et al., 1992).

Coreceptor usage prediction using the charge rule did not always align with experimental tropism

determination. Comparative studies with the Trofile assay have only revealed a moderate correlation (Low et al., 2007). This inaccuracy was attributed to the relatively scant and unsubstantiated training set data that was utilized in the development of this assay (Pillai et al., 2003). The consideration of only two out of the 35 available amino acid positions of the V3 loop, was also a shortfall when using this predictive tool. The predictive value of this in-silico assay was enhanced when the charge of the residues at positions 11, 24, and 25 were considered as tropism determinants rather than the charge at positions 11 and 25 only. Therefore, according to this new data, a new rule "11/24/25" was suggested, in which a positively charged amino acid at position 11, 24, or 25 defines X4 tropic variants; otherwise R5 (Cardozo et al., 2007). This prediction method is remarkably accurate in sensitivity to R5 variants (>90%) than R4 variants (<50%) (Jensen and Wout, 2003; Low et al., 2007). However, the 11/25 rule remains a top choice due to its increased sensitivity to the minority X4 variants.

2.8.2.3 Position specific scoring matrices (PSSM)

PSSM is a very simple and reliable bioinformatics method of determining CXCR4 coreceptor usage with high sensitivity (84%) and specificity (96%) (Poveda et al., 2009). In this technique, the likelihood that a sequence is derived from a CXCR4 tropic virus is calculated for ever single amino acid at every individual position. The allocated score translates to a predicted tropism as follows: known R5 viruses have low scores while X4 viruses have the highest score. A value of -6.96 or less usually denotes CCR5 tropism, whereas CXRR4 tropic sequences have values above -2.88. The 11/25 rule is used to interpret intermediate scores (Jensen et al., 2003). A modified version of the PSSM [sinsi] matrix, known as PSSM [r5x4] with increased sensitivity to X4 variants has been developed (Poveda et al., 2009). The training data set for the initial PSSM [sinsi] matrix was based on HIV-1B variants. Currently, the PSSM [sinsi-C] optimized for tropism prediction in HIV-1 subtype C variants is also available (Riemenschneider et al., 2016). All PSSM matrices can be accessed online via WebPSSM *(indra.mullins.microbiol.washington.edu/webpssm/).* These matrices been evaluated in several cohort studies and retrospective analyses of clinical trials (Garrido et al., 2008; Siddik et al., 2018; Tremblay et al., 2013) to name a few, and will thus be utilized in this study.

2.8.2.4 Geno2Pheno

This tool bases prediction on the statistical method known as support vector machine (SVM) technology, where the algorithm is trained with over 1100 V3 loop sequences with pre-determined

coreceptor tropism (Seclé et al., 2010). The overall structure of the query V3 loop sequence is evaluated and assigned R5 or X4, depending on the false positive rate chosen. A false positive rate (FPR) is defined as the likelihood of a sequence being mistakenly classified as CXCR. This prediction tool allows for the configuration of the settings to an individual's preference by varying the significance level at FPR 1%, 2.5%, 5%, 10%, 15%, and 20% (Sing et al., 2007). Varying the threshold value for the classification of false-positive rates changes the sensitivity and specificity for X4 prediction. Increasing the FPR usually increases the likelihood of detecting X4 variants, but also of falsely assigning CXCR4 tropism to V3 loop sequences, while decreasing the FPR increases discordance (Verhofstede et al., 2011). Generally, FPR 5-20% can be utilized, but FPR 10% or 15% is recommended for clinical use by the European Consensus Group of clinical management for tor tropism testing (Vandekerckhove et al., 2011).

The Geno2Pheno algorithm has two variants: a clonal variant which is routinely employed in cohort studies and retrospective analyses of clinical trials, and another variant in which clinical data such as the nadir of CD4 and baseline viral load are taken into consideration. The latter has not been studied extensively. Both variations are available online at (www.geno2pheno.org). The Geno2Pheno clonal variant will thus be utilized in this study since this tool has also demonstrated comparable performance to the Original Trofile Assay (OTA) and the Enhanced Sensitivity Trofile Assay (ESTA) (Judicate et al., 2021; Seclé et al., 2010; Soulié et al., 2016).

2.8.2.5 HIV CoR

This prediction model was developed for predicting CRF01_AE coreceptor usage. it utilizes Random Forest and SVM as prediction models with input features comprising of the composition of amino acids, pseudo amino acid compositions, and the relative frequencies of synonymous codon usage. When validated externally against the objective benchmark dataset, a 93.79% accuracy was observed indicating the superiority of this predictor over other in silico coreceptor tropism prediction tools (Hongjaisee et al., 2019), informing the use of this tool to predict coreceptor usage of the isolates within this data set.

Chapter 3

3.0 Methodology

3.1 Problem statement

HIV infection rate is disproportionately high in marginalized populations such as female sex workers (FSWs), men who have sex with men (MSM), and injecting drug users (IDUs). Approximately 5.6% of the Kenyan population are infected by HIV-1, with >3-fold higher HIV-1 prevalence among so-called high-risk groups ((NASCOP), 2018). FSWs are amongst the most vulnerable of the key population groups. Prevalence estimates are 13.5-fold higher in FSWs compared to their non-sex worker counterparts (Bitty-Anderson et al., 2022). In Kenya, the estimated risk of transmission is 9.8 times higher in this group compared to non-sex worker females (Tago et al., 2021). This highlights the burden of HIV in this key population and their central role in sustaining population-level infections. This population is also more likely to harbor multiple HIV-1 variants and recombinants due to their social vulnerability and the risks associated with their occupation including multiple sexual partners, inconsistent condom use, use of injecting and non-injecting drugs, and co-infections with other STIs (Baral et al., 2012; Prakash et al., 2018; Tago et al., 2021). However, it is not clear how virally encoded factors and host immune status influence HIV transmission in such a key population. Monitoring HIV-1 in this key population is therefore crucial to understanding the pandemic's trajectory and formulating effective preventative and control measures.

3.2 Justification

Recent literature has highlighted changes in the geographical patterns of subtype distribution due to migration and the mixing of populations. Subtypes A, D, and C were reported to be the most prevalent subtypes in Kenya by ((NASCOP), 2018). Due to migrations and the central location of the study cohort within Nairobi, such a study to identify all the subtypes in circulation and their corresponding coreceptor tropism profiles is timely.

Different clades of HIV-1 exhibit major differences in pathogenesis and resistance pathways which in turn affect clinical outcomes, diagnosis, viral load, and vaccine development. Moreover, tropism testing now forms a crucial reference point for initiation of antiretroviral therapy, especially following the dawn of entry inhibitors such as Maraviroc, that solely inhibit entry at the CCR5 coreceptor. Due to their selective effect on the viral populations, determining viral tropism is critical for monitoring disease progression and for antiretroviral support therapy that involves these entry inhibitors.

3.3 Study objectives

3.3.1 Study Questions

1. What is the genetic diversity of HIV-1 subtypes and their corresponding coreceptor tropism profiles in Nairobi female sex workers?
2. What are the predictors (determinants) of coreceptor usage in this key population?
3. What is the degree of relatedness of the viruses within this cohort of female sex workers?

3.3.2 Overall objective

- To determine the genetic diversity of HIV-1 subtypes and their corresponding coreceptor tropism profiles in Nairobi female sex workers

3.3.3 Specific objectives

i. To identify HIV-1 subtypes circulating in this cohort of female sex workers using COMET (Context based modeling for ultrafast HIV-1 subtype identification), REGA (resistance interpretation algorithm), and NCBI (national resource for molecular biology information) viral subtyping tools), and the concordance of these subtyping tools &

   To identify the coreceptor usage based on the V3-loop of HIV-1 gp120 using genotypic in silico prediction tools ($\text{Geno2Pheno}_{[\text{coreceptor}]}$, WebPSSM (Web-position specific matrices), HIVCoR, 11/25 rule, net charge rule and concordance of these tropism prediction tools.

ii. To determine predictors of coreceptor usage.

iii. To determine the degree of relatedness of the viruses in this key population.

3.3.4 Research hypothesis

H_0 = HIV-1 subtype diversity and coreceptor usage in this key population is higher compared to the general population.

Due to the high-risk sexual nature of the cohort under study, with a reported high prevalence of 29.3%, there will be a likely identification of novel HIV-1 subtypes circulation which have gone unreported by previous studies, their corresponding tropism profiles as well as HIV superinfection.

3.4 Materials and methods

3.4.1 Study Design, setting, and population.

This was a cross-sectional study nested within the ongoing HIV Prevention and Care study, involving an open cohort of female sex workers from Nairobi, Kenya. Ethical approval was obtained from the KNH-UON ethical review committee under the approval number **P556/07/2019.** A waiver of consent was sought for this study since there was minimal risk of harm to the participants (see appendix). This study did not adversely affect the welfare and privacy of the study participants since the focus on other immunological and molecular aspects of the HIV virus.

During recruitment into the parent study, recruitment personnel thoroughly explained the research procedures in English and Swahili in line with the guidelines of the approved informed consent obtained from KNH-UON ERC. The informed consent was obtained in two steps; at the screening visit during screening procedures and at the enrollment visit during which eligible participants who met the inclusion criteria agreed to study participation. Since this study utilized samples that had been collected for the parent study, it was important to obtain the consent of the participants to the storage remaining samples after the trial had ended, and the utilization of these samples in futuristic HIV studies. Page 6 of parent study protocol: a statement of the consent document states, "A portion of my blood, cervical and vaginal specimens will be store for future studies of the genes involved in resistance and susceptibility to HIV and other infections." (see appendix)

3.4.2 Study setting and population

Following approval by KNH-UON ERC to utilize human subjects for this study, recruitment of treatment naïve, HIV-1 seropositive participants occurred at several clinics that provide comprehensive clinical care to female sex workers. These facilities are based in Nairobi city center, Donholm, Kawangware, Korogogocho, Lang'ata, Majengo, and Thika road in Nairobi County.

3.4.3 Study material

Blood samples were collected from consenting study participants at the clinics and transported in cooler boxes from point of collection to the PHDA laboratories where they were stored at-80°C until RNA extraction. Each sample was uniquely coded without any personal identification details to ensure the protection of privacy for each participant. The inclusion criteria were consenting, HIV-1 seropositive participants based on HIV ELISA assay, ≥ 18 years of age with a viral load of >1000 copies/ml, and where viral load data was unavailable, an amplifiable PCR band.

3.4.4 RNA extraction

Viral RNA was extracted from plasma using PureLink™ Viral RNA/DNA Mini Kit (Thermofisher) per the manufacturer's instructions. Briefly, frozen plasma samples, and in-house positive (1000RNA copies/ml) and negative controls were thawed at room temperature and thereafter placed on ice in the biosafety cabinet.

As the samples were thawing, buffer AW2 was freshly prepared by adding 60ml of ethanol (96-100%) to the buffer. Sterile, 2ml centrifuge tunes were labeled with the corresponding sample IDs and controls. 62.5 µl of Proteinase K was added to each of the labeled tubes. 500 µl of each plasma sample and in-house control was added to the appropriately labeled tubes followed by 500 µl of the viral lysis buffer. The mixture was vortexed briefly and incubated at 56°C for minutes to ensure complete viral lysis. 625µl of ethanol (96–100%) was then added to the lysed sample, vortexed briefly, and incubated at room temperature for 5 min.

The Purelink viral spin columns were set up in 2ml collection tubes and labeled appropriately. 825 µl of the lysed sample was carefully applied to the corresponding appropriately labeled spin columns and centrifuged at 6800g for 1 minute. The collection tubes containing the filtrate were discarded and the viral spin columns were placed in new collection tubes. The remaining lysate (approximately 825 µl) was then carefully added to the appropriate spin columns and centrifuged at 6800g for 1 min. The collection tubes were discarded, and the spin columns were placed in new collection tubes. 500 µl of AW2 buffer was added to each spin column and centrifuged at 6800g for 1 min. Collection tubes containing the filtrate were discarded and the viral spin columns were placed in new collection tubes. The spin columns were washed again with 500 µl of AW2 solution, centrifuged at 6800g for 1 minute, and the filtrate was discarded. The spin columns were then placed in fresh collection tubes and centrifuged at 17000g for 1 minute. The collection tubes were discarded.

The spin columns were then placed in 1.5ml microcentrifuge tubes appropriately labeled to depict the contents of the tube. RNA was eluted by adding 20 µl of RNAse-free water to the column, incubating at room temperature for 1 minute, and centrifuging at 17000g for 1 minute. This step was repeated to elute the RNA to a total volume of 40 µl. The spin columns were carefully disposed after centrifugation. The RNA was aliquoted for storage at -60°C to -80°C, and 10 µl chilled on ice for the downstream PCR reactions.

3.4.5 Amplification of Env gene

3.4.5.1 Reverse transcription

10µl of RNA sample was denatured and amplified in the first round using One*Taq*® One-Step RT-PCR Kit (New England Biolabs) and primers M5 5'-CCAATTCCCATACATTATTGTGCCCCAGCTGG-3' (forward) and M10 5'-CCAATTGTCCCTCATATCTCCTCCTCCAGG-3' (reverse) corresponding to the 2) of the HIV-1 group M, env gene. PCR conditions were reverse transcription at 48^0C for 30 minutes, denaturation at 94^0C and 40 PCR cycles of denaturation 94^0C at for 15s, annealing 55^0C at for 30s and extension at 68^0C for 1min and a final extension at 68^0C for 5min (Lwembe et al., 2009). The reaction mix was set up as shown below.

Component	Volume
One Taq One step reaction mix	12.5µl
One Taq one step enzyme mix	2µl
Forward primer	2µl
Reverse primer	2µl
RNAse free water	2.5µl
Template RNA	4µl
Total	25µl

Table 1: Reverse transcription PCR reaction mix set up

3.4.5.2 Nested PCR

The reverse transcribed cDNA was further amplified using GoTaq® G2 Hot Start Green Master Mix (Promega) and primers M3 5'-GTCAGCACAGTACAATGCACACATGG-3' (forward) and M8 5'-TCCTTGGATGGGAGGGGCATACATTGC -3'(reverse). PCR conditions were reverse transcription at 48^0C for 30 minutes, denaturation at 95^0C and 35 PCR cycles of denaturation 95^0C at for 30s, annealing 55^0C at for 30s and extension at 72^0C for 1min, and a final extension at 72^0C for 10min (Lwembe et al., 2009). The reaction mix was set up as follows:

Component	Volume
Go Taq green master mix	10µl
Forward primer	0.4µl
Reverse primer	0.4µl
RNAse free water	5.2µl
Template cDNA	4µl
Total	20µl

Table 2: Nested PCR reaction mix set up

The PCR products were separated using 1% agarose gel stained with SYBR™ safe DNA gel stain (Thermo Fisher) and visualized under UV light.

3.4.6 V3 loop sequencing

3.4.6.1 PCR product purification

The amplicons were purified using EXOSAP-IT™ (Applied Biosystems) to remove residual primers, enzymes, nucleotides, and salts. Briefly, PCR tubes were labeled appropriately and 2µl of EXOSAP-IT™ added to each tube. 5µl of the sample was then slowly added to the corresponding tube. The PCR conditions for purification were:

Cycle step	Temperature	Time	Cycles
Initial Denaturation	95^0C	10min	1
Denaturation	95^0C	30s	
Annealing	55^0C	30s	35
Extension	72^0C	1min	
Final Extension	72^0C	10min	1
Hold	4^0C	∞	1

Table 3: PCR conditions for amplicon clean-up

3.4.6.2 Cycle sequencing

Purified products were sequenced using standard Big-Dye Terminator v3.1 chemistry (Thermofisher). The sequencing primers used were sequencing forward primer 1 (SP1) 5'-AGYRCAGTACAATGYACACATGG-3' and forward sequencing primer 2 (SP2) 5'-TCAACHCAAYTRCTGTTAAATGG-3' (Mulinge et al., 2013) for the samples that failed to amplify with SP1. 20 µl of pGEM was added to at least one well per run as a sequencing control. The table below shows the reaction set up for each sample on a 96-well plate. The organization of the 96-well plate was charted before adding the reaction components.

Component	Volume
Big Dye Terminator	1.5µl
5X Buffer	2µl
RNAse free water	3µl
Forward primer	1µl
DNA	3µl
Total	**10.5µl**

Table 4: Cycle sequencing reaction mix set up

The 96 well plate was carefully sealed with clear adhesive film, placed in the thermocycler, and the BDT sequencing program outlined in the table below, initiated.

Cycle step	Temperature	Time	Cycles
Denaturation	96^0C	10s	
Annealing	50^0C	5s	25
Extension	60^0C	4min	
Hold	4^0C	∞	

Table 5: PCR conditions for Cycle sequencing product cleanup

3.4.6.3 Purification of cycle sequencing products with Big Dye XTerminator™

The Xterminator (Applied Biosystems) solution was removed from the freezer and allowed to equilibrate at room temperature and vortexed briefly. A SAM/Big Dyer Xterminator working solution was set up by mixing 45 of SAM solution with 10 of XTerminator for a 10 µl reaction. The mixture was vortexed and transferred into the 96-well plate containing the samples using a multichannel pipette. The solution was pipetted down the walls of the well to avoid touching the samples. The plate was then sealed with an adhesive seal, wrapped in gauze, and placed on the plate mixer, secured with rubber bands as needed for 30 minutes. The plate was then briefly centrifuged to collect unincorporated reaction components at the bottom of the wells.

3.4.6.4 Sequence detection

Sequence detection was performed by automated capillary electrophoresis on the ABI 3730xl DNA sequencer (Applied Biosystems, Foster City, CA). A septum was placed on the plate following the removal of the adhesive film and the plate was inspected to ensure no blockage by the septum. The reaction plate was then loaded on the instrument and the appropriate BDX run module was chosen. Sequencing files were interpreted using BioEdit (version 7.0.5.3) and Mega11 software. The *env* sequences were archived in the DDBJ Nucleotide Database with the accession numbers: **LC722376-LC722451**.

3.4.7 HIV-1 Subtyping

HIV-1 subtyping was carried out by submitting raw env sequences to three HIV-1 subtyping tools; (i)COMET software (https://comet.lih.lu/index.php?cat=hiv1), (ii) REGA HIV subtyping tool (www.bioafrica.net/rega- genotype/html/subtypinghiv.htm) (De Oliveira et al., 2005), (iii)NCBI Viral Genotyping tools www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi.

3.4.8 Determination of Viral tropism

The sample nucleotide sequences were aligned within MEGA11 using the Clustal W algorithm against the V3 loop reference sequence **Z15191.1.** The original sequences were trimmed following multiple sequence alignments against the reference V3 loop sequence to obtain sample V3 loop

sequences. Aligned, and edited sequences covering the entire V3 region were analyzed for coreceptor usage using various in silico genotypic tools: Geno2Pheno https://coreceptor.geno2pheno.org/, HIVCoR (Hongjaisee et al., 2019), and

For Geno2Pheno a 10% FPR, the standard recommended by the European Consensus Group on clinical management of HIV-1 tropism testing was first applied. The analysis was then repeated at 15% and 20% FPR to improve the confidence for assigning the X4-phenotype correctly. Given recent findings that have highlighted the emergence of subtype C variants in Kenya, a 5% FPR was also applied to the V3 loop sequences due to increased fidelity in identifying CXCR4 usage in both CXCR4 elusive and dual tropic variants at this cut-off, with respect to subtype C tropism prediction.

The V3 nucleotide sequences were translated to the corresponding amino acid sequences using the translate tool at (https://www.bioinformatics.org/sms2/translate.html). Translated V3 amino acid sequences were then analyzed for HIV tropism using $\text{WebPSSM}_{[sinsi]}$, $\text{WebPSSM}_{[R5X4]}$, $\text{WebPSSM}_{[sinsi\ C]}$, **(http://indra.mullins.microbiol.washington.edu/ PSSM/)**, 11/25 rule (Fouchier et al., 1992), and the net charge rule (Raymond et al., 2008). For the 11/25 rule, the presence of basic amino acid residues (Arg, Lys or His) at position 11 and or 25 of the V3 loop was predictive of CXCR4 use, while the default was CCR5. The net charge was calculated by subtracting the total number of negatively charged (D/E) residues from the total number of positively charged residues (R/H/K). Raymond's rule(Raymond et al., 2008) combines one of the following criteria; (i) Arg/ Lys at position 11 and/or Lys at position 25, (ii) Arg at position 25, and a net charge of ≥5, or (iii) a net charge of ≥6, was also used to predict viral tropism.

3.4.9 Phylogenetic tree construction

For the phylogenetic relationships, the V3-loop nucleotide sequences were aligned with MEGA11 (Stecher et al., 2020). Following the alignment, the optimal tree model was constructed using Maximum Likelihood method with 1000 bootstrap and visualized using FigTree (v1.4.4.) http://tree.bio.ed.ac.uk/software/figtree/. Clustering was assessed using Cluster Picker 1.2 with the following default settings: Initial threshold, 0.9; Main support threshold, 0.9; Genetic threshold distance threshold, 4.5; and large distance cluster threshold, 5.

3.4.10 Statistical analyses

Tropism coded (0 = CCR5; 1 = CXCR4) was the primary outcome in this study. The predictor variables were: virus-specific (HIV subtype and net charge); host-specific (CD4 count and CD4 %); and social demographic (age, sexual preference, and location). Mann-Whitney U test was used to compare the median values of continuous variables between the two groups (R5 and X4), respectively, and Fisher's exact test was used to compare categorical variables between the two groups as appropriate, with 2-sided p-values reported in all cases. Differences between proportions were tested by the χ^2 test. The consistency of the different algorithms to predict the coreceptor usage was calculated using the Kuder-Richardson Formula 20 (KR-20) with a cut-off of 0.70 deemed acceptable (Cronbach, 1951). Univariable and multivariable logistic regression models were used to estimate the predictors of tropism, with estimated odds ratios (ORs) and 95% confidence intervals reported. A two-sided p-value <0.05 was considered to be statistically significant and was included in the full model. HIV-1 subtype was included in the fitted models as a possible confounder regardless of statistical significance. All statistical analyses were conducted with the R statistical package (R version 4.1.0).

3.5 Results

3.5.1 Sample amplification

A total of 155 plasma samples collected between November 2020 and April 2021 from HIV-1 seropositive, treatment naïve FSWs were included in the study. Viral RNA was extracted from all the samples. Samples with no visible PCR band were excluded. Envelope (*env* C2V3 region) PCR amplicons and V3-loop sequences were obtained for 77 samples. One of the 77 samples had an erroneous nucleotide whose translated amino acid sequence was an atypical V3 loop sequence Seventy-six *env* C2V3 region (V3 loop) sequences were analyzed.

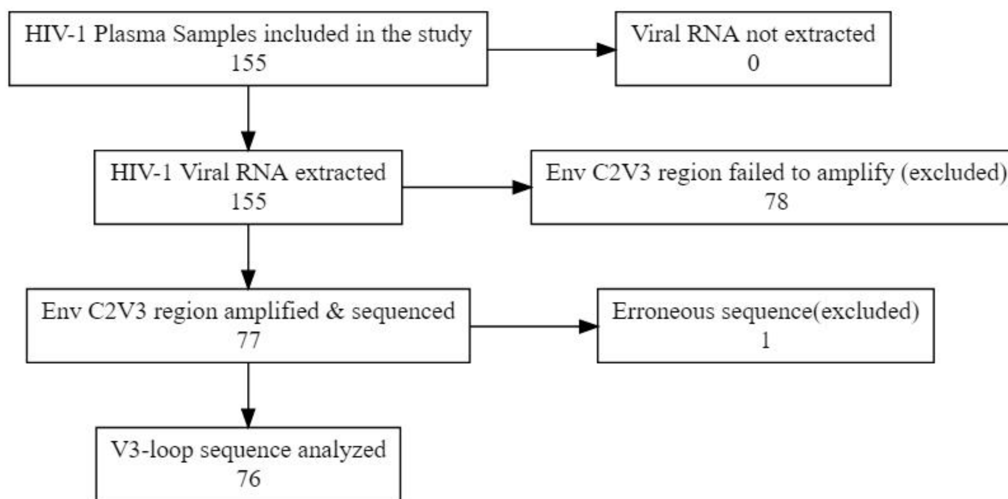


Figure 13: Sample amplification rate

3.5.2 Patient characteristics

The patients had a median age of 35.5 years (range 22-58 years). The majority of patients 35 (46.1%) were in the 30-40 age group while the rest, 20 (26.3%), 16 (21.1%), and 5 (6.6%) were in the 20-30, 30-40, and >50 years age categories, respectively. Seventy-one (93.4%) of the subjects were heterosexual, while 5 (6.6%) were homosexual. The median CD4 T cell counts, and CD 4% were 387.5 cells/µL (range 23-1113 cells/µL), and 22.5% (range 2-42%), respectively, with a majority of the subjects 50 (65.8%), having a CD4 count of less than 500 cells/µL, while 26 (34.2%) had a CD 4 T cell count of more than 500 cells/µL.

3.5.3 HIV-1 Subtypes

HIV subtype analysis was based on *env* C2V3 region sequences submitted to three subtyping tools: COMET, REGA, and the NCBI viral genotyping tool. COMET-assigned subtype distribution for the successfully amplified samples was A1 (65), D (5), C (3,) B (1), J (1), G (1) representing 85.5%, 6.6%, 3.9%, 1.3%, 1.3%, and 1.3% of the isolates, respectively. The most prevalent subtype was also A1 at 59 (77.6%) according to REGA followed by subtype D at 5 (6.6%), and subtype C to 2 (2.6%). Ten isolates (13.2%) were not assigned a HIV-1 subtype by the REGA tool. The NCBI Viral subtyping tool assigned subtype A1 to 59 (77.6%), subtype A2 to 1 (1.3%), subtype C to 2 (2.6%), subtype D to 3 (3.9%), CFR_02 to 7 (9.2%), CRF_10 to 2 (2.6%) and CRF_11 to 2 (2.6%) of the isolates (Table 6).

3.5.4 Concordance of the HIV-1 subtyping tools

The degree of agreement among the HIV subtyping tools was calculated using Cronbach's alpha coefficient. The raw alpha for the COMET, REGA, and NCBI viral subtyping tools was 55% (95% CI 34%-70%), which is lower than the acceptable reliability cutoff of 70% (0.7). Reliability, when a tool was dropped, is shown in the table below (Table 6). Raw alpha increased to 0.84 (95% CI 0.75-0.90,) when concordance was calculated for the COMET and REGA tools.

Subtyping tool	HIV-1 subtype n (%)				Reliability	
	A1	D	C	[a]OTHERS	[b]Raw alpha	[c]Raw alpha
COMET	65(85.5)	5(6.6)	3(3.9)	3(3.9)	0.32	0.60
REGA	59(77.6)	5(6.6)	2(2.6)	10(13.2)	0.42	0.94
NCBI TOOL	59(77.6)	3(3.9)	2(2.6)	12(15.7)	***0.82***	-

Table 6: HIV-1 subtypes assigned by the different subtyping tools and their reliability calculated using Cronbach's alpha coefficient. (n=76)

[a]OTHERS: COMET-B, G, J, A1/02_AG, Unassigned A1_C, REGA: Unassigned, NCBI: A2, CRF_02,10,11

[b]Raw alpha reliability when an item is dropped when comparing COMET, REGA and NCBI tool

[c]Raw alpha reliability when an item is dropped when comparing COMET and REGA tools.

Dropping the REGA subtyping tool significantly reduces the alpha on both occasions as shown in table 6. However, the REGA tool was unable to predict HIV-1 subtype for 10 (13.2%) of the 76 of isolates. In contrast, all isolates were assigned a subtype by the COMET tool, a commonly used HIV-1 subtyping tool. COMET-based HIV-1 subtypes were therefore selected for further comparative analysis.

3.5.5 Viral tropism prediction

A total of 76 aligned, trimmed, and edited V3 loop nucleotide sequences were analyzed for coreceptor tropism using Geno2Pheno [5%, 10%, 15% & 20% FPR] and the HIVCoR tools. All 76 sequences were assigned CCR5 tropism by the HIV-CoR tropism prediction tool.

Viral tropism at FPR cut off n (%)								
HIV-1	FPR, 5%		FPR, 10%		FPR, 15%		FPR, 20%	
Subtype	CCR5	CXCR4	CCR5	CXCR4	CCR5	CXCR4	CCR5	CXCR4
A1	59 (91)	6 (9)	50 (77)	15 (23)	49 (75)	16 (25)	48 (74)	17 (26)
B	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)
C	3 (100)	0 (0)	3 (100)	0 (0)	3 (100)	0 (0)	3 (100)	0 (0)
D	3 (60)	2 (40)	3 (60)	2 (40)	2 (40)	3 (60)	2 (40)	3 (60)
G	1 (100)	0 (0)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)
J	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	0 (0)
Total	68 (89)	8 (11)	58 (76)	18 (24)	56 (74)	20 (26)	55 (72)	21(28)
X^2 p value	0.37		0.36		0.18		0.20	

Table 7:The distribution by number and (proportions), of viral tropism across COMET based HIV-1 subtypes using Geno2Pheno at four cut off points. P-value is derived from cross-tabulation of coreceptor tropism and virus subtype using Pearson's Chi-square test.

For Geno2Pheno, setting the FPR at 5% resulted in 68 (89.5%) samples predicted as CCR5-tropic, and 8 (10.5%) predicted as CXCR4-tropic. There were no CXCR4 tropic subtype C variants at 5% FPR, despite the increased fidelity in identifying CXCR4 usage in both CXCR4 elusive and dual tropic variants at this cut-off for HIV-1 subtype C. Tropism of subtype C viruses appeared unaffected across different FPRs. Increasing the FPR to 10%, resulted in 58 (76.3%) R5 variants and 18 (23.7%) X4 variants. With Geno2Pheno, selecting greater FPRs increases the sensitivity in detecting X4 variants, but also for falsely identifying a sequence as X4. At 15% FPR, the number of X4 variants increased to 20 (26.3%), and the R5 variants reduced to 56 (73.7%). A slight

increase to 21 (27.6%) of CXCR4 tropic variants was observed at 20% FPR, with CCR5 tropic variants decreasing to 55 (72.4%) (Table 7).

Viral tropism clustered by viral subtype is shown in (Table 7) for the Geno2Pheno tropism prediction algorithm. There was no significant correlation between viral tropism and HIV-1 subtype for all four FPR cut-off algorithms. However, these analyses indicated a trend whereby subtype A1 variants were likely to be assigned CCR5 tropism at lower FPR, while slightly more of the subtype D isolates were inclined towards CXCR4 tropism at higher FPR. Tropism predicted by the Geno2Pheno 15% FPR algorithm was used in further comparative analyses for all the 73 V3-loop nucleotide sequences since Geno2Pheno 15%FPR is recommended for clinical use by the European Consensus Group of clinical management for tor tropism testing (Vandekerckhove et al., 2011).

Translated V3 amino acid sequences were analyzed for HIV tropism using $WebPSSM_{[sinsi]}$, $WebPSSM_{[R5X4]}$, $WebPSSM_{[sinsi\ C]}$, 11/25 rule, and the net charge rule. According to $WebPSSM_{[sinsi],}$ 67 (88.2%) of the isolates were CCR5 tropic, and 9 (11.8%) were CXCR4 tropic.

	Viral tropism n (%)					
HIV-1	[a]WebPSSM [C]		WebPSSM [SINSI]		WebPSSM [R5X4]	
Subtype	CCR5	CXCR4	CCR5	CXCR4	CCR5	CXCR4
A1	9 (14)	56 (86)	58 (89)	7(11)	52 (80)	13 (20)
B	0 (0)	1 (100)	1 (100)	0 (0)	1 (100)	0 (0)
C	2 (67)	1(33)	3 (100)	0 (0)	3 (100)	0 (0)
D	1 (20)	4 (80)	3 (60)	2 (40)	3 (60)	2 (40)
G	0 (0)	1 (100)	1 (100)	0 (0)	1 (100)	0 (0)
J	0 (0)	1 (100)	1 (100)	0 (0)	0 (0)	1 (100)
Total	12 (16)	64 (84)	67 (88)	9 (12)	60 (79)	16 (21)
X^2 *p* value	0.25		0.46		0.29	

Table 8: The distribution by number and (proportions), of viral tropism across COMET based HIV-1 subtypes using three WebPSSM matrices. P-value is derived from cross-tabulation of coreceptor tropism and virus subtype using Pearson's Chi-square test. [a]WebPSSM [SINSI -C] matrix.

WebPSSM$_{[R5X4]}$ assigned CCR5 tropism to 60 (78.9%) of the isolates, and CXCR4 tropism to 16 (21.1%) of the isolates. The WebPSSM$_{[sinsi\ C]}$ matrix has been reported to overestimate CXCR4 tropism. Overall, the sinsi-C matrix scored majority (84%) of the isolates as CXCR4. No significant associations were observed between COMET-based HIV-1 subtypes and viral tropism assigned by the WebPSSM matrices (Table 8).

Based on the 11/25 rule, 71(93.4%) of the isolates were R5 while 5 (6.6%) were CXCR4-tropic (Table 9). Stratification of net charge (NC) values using 4 and 5 breakpoints were based on previously published observations by (Clevestig et al., 2006), where sequences with NC values below 4.0 were classified as R5, those with NC values above 5.0 were classified as X4, while those with intermediate NC values (4.0–5.0) were classified as undetermined. Here, 54 (71.1%) of the isolates were CCR5 tropic, while 22 (28.9%) were CXCR4 tropic (Table 9). We did not conduct phenotypic tropism testing to validate phenotype derivations.

3.5.6 Concordance of the genotypic tropism testing tools

Cronbach's alpha coefficient was used to measure the reliability of the different genotypic tropism testing (GTT) tools (Cronbach, 1951). First, reliability was evaluated for the variations within each tropism prediction algorithm. For Geno2Pheno, agreeability was calculated for the different cut-off FPRs [5,10,15,20%]. The raw alpha for the four variations of this algorithm was 0.93 (Feldt 0.90-0.95, 95% CI), exceeding the 0.70 cutoffs for acceptable concordance. For WebPSSM, agreeability was calculated for the SINSI, SINSI-C, and R5X4 matrices. The raw alpha for these matrices was 0.52 (Feldt 0.29-0.68, 95% CI), which is below the 0.70 cut-off for acceptable reliability. Raw alpha increased most to 0.74 when the SINSI-C matrix was dropped. After G2P_5 and WebPSSM [SINSI-C] assigned viral tropism were dropped**,** reliability was recalculated for the G2P_10, G2P_15, and G2P_20, giving a raw alpha value of 0.98 (95% CI 0.97-0.98), and 0.75 (95% CI 0.60-0.84,) for WebPSSM [SINSI and R5X4 matrices] (Table 9).

In addition, the degree of agreement among the various algorithms was calculated. Reliability was calculated for Geno2Pheno [5,10,15 and 20% FPR], WebPSSM [Sinsi and R5X4 matrices], and the 11/25 rule viral tropism prediction algorithms were calculated. The raw alpha for all seven algorithms was 0.88 (95% CI 0.82-0.91,), exceeding the 0.70 cut-off for acceptable internal reliability (table 4). Raw alpha_1 in Table 9 column shows variation in reliability when an algorithm is dropped. G2P_5 and the 11/25 algorithms were dropped, and concordance was

calculated again for G2P_10, G2P_15 & G2P_20, and the WebPSSM [SINSI, and R5X4] matrices (Raw alpha_2 column in Table 9). G2P_15 assigned tropism was retained and used to stratify tropism outcomes within this data set (Mulinge et al., 2013).

Tropism prediction tool	Viral Tropism n (%)		Concordance 1	Concordance 2
	CCR5	CXCR4	Raw alpha	Raw alpha
Geno2Pheno 5 FPR	**68 (89)**	**8 (11)**	**0.85**	**--**
Geno2pheno 10 FPR	58 (76)	18 (24)	0.84	0.78
Geno2Pheno 15 FPR	56 (74)	20 (26)	0.84	0.76
Geno2Pheno 20 FPR	55 (72)	21 (28)	0.84	0.77
			0.93 [CI 0.90-0.95]	**0.98 [CI 0.97-0.98]**
WebPSSM Sinsi	67 (88)	9 (12)	0.87	0.86
WebPSSM R5X4	60 (79)	16 (21)	0.89	0.89
WebPSSM Sinsi C	**12 (16)**	**64 (84)**	**0.93**	**--**
			0.52 [CI 0.90-0.95]	**0. 75 [CI 0.90-0.95]**
11/25 Rule	**71 (93.4)**	**5 (6.6)**	**0.87**	**--**
Net Charge Rule	54 (71.1)	22 (28.9)	--	--
			0.88 [CI 0.82-0.91]	**[CI 0.90-0.95]**

Table 9: Distribution of viral tropism phenotypes and variation in tropism prediction reliability when an algorithm is dropped.

3.5.7 Tropism in relation to host and viral characteristics

COMET-assigned HIV-1 subtypes were used as the reference subtypes for each subject, while Geno2Pheno 15% FPR-based tropism, was used as the reference tropism. Six subtypes were reported. Subtype A1 62 (85.5%) was the major subtype in this cohort.

Coreceptor	HIV-1 Subtype n (%)						
tropism	A1	B	C	D	G	J	Total
CCR5	49 (88)	1 (1.8)	3 (5.4)	2 (3.6)	0 (0)	1 (1.8)	56 (73.7)
CXCR4	16 (80)	0 (0)	0 (0)	3(15)	1 (5)	0 (0)	20 (26.3)

Table 10: HIV-1 subtype and viral tropism distribution

Characteristic	Viral tropism n (%)	
	[a]CCR5	[b]CXCR4
Age, Median (IQR)	34 (29 – 40)	37 (30 – 39)
Age category		
20-30	15 (27)	5 (25)
30-40	25 (45)	10 (50)
40-50	11 (20)	5 (25)
>50	5 (8.9)	0 (0)
CD4 percent, Median (IQR)	23 (14 – 31)	19 (10 – 26)
[a]CD4 count, Median (IQR)	410 (249 – 630)	338 (141 – 444)
[b]CD4 count strata		
<200	11 (20)	6 (30)
200-350	14 (25)	5 (25)
350-500	10 (18)	4 (20)
>500	21 (38)	5 (25)
[c]SWOP clinic n (%)		
Thika Road	16 (29)	5 (25)
City	12 (21)	1 (5)
Donholm	4 (7.1)	0 (0)
Kawangware	9 (16)	10 (50)
Korogocho	9 (16)	2 (10)
Langata	2 (3.6)	1 (5)
Majengo	4 (7.1)	1 (5)
Sexual preference		
Heterosexual	52 (93)	19 (95)
Homosexual	4 (7.1)	1 (5.0)

Table 11: Demographic and clinical characteristics of 76 study subjects included in the analysis. Results are median (IQR) or frequency (%), as appropriate. Correlations were calculated using Pearson's Chi-Square test.
[a]**CCR5** = C-C motif chemokine receptor 5, [b]**CXCR4** = CXCR4 C-X-C motif chemokine receptor 4, [c]**SWOP**=sex worker outreach program

Stratification of viral strains based on tropism showed that 73.7% (n=56) were CCR5-tropic, while 26.3% (n=20) were CXCR4-tropic. There was no significant association between HIV-1 subtype and G2P_15-assigned coreceptor tropism (p = 0.18). All other host characteristics were stratified using Geno2Pheno 15% FPR-based tropism. Results are shown in Table 11. Subjects infected with CCR5-tropic viruses had a median age of 34 years (interquartile range [IQR], 29-40), whereas those infected with CXCR4-tropic viruses had a median age of 37 years (IQR, 30-39). Subjects infected with CCR5 viruses had 410 copies/mm3 (IQR, 229-630), whereas those

infected withCXCR4 viruses had 338 copies/mm3 (IQR, 141-444). Among subjects infected with CCR5-tropic viruses, 20% (n=11) had $CD4^+$ T cell counts <200 copies/mm^3, 25% (n=14) had $CD4^+$ T cell counts between 200 and 349 copies/mm^3, 18% (n=10) had $CD4^+$ T cell counts between 350 and 500 copies/mm^3, and 38% (n=21) had $CD4^+$ T cell counts >500 copies/mm^3. In subjects infected with CXCR4-tropic viruses, 30% (n=6), 25% (n=5), 20% (n=4) and 25% (n=5) had <200 copies/mm^3, 200-349 copies/mm^3, 350-500 copies/mm^3, and >500 copies/mm^3, respectively. Viruses that were CCR5-tropic had a median $CD4^+$ T cell percentage of 23 (IQR, 14-31) while CXCR4-tropic viruses had a median $CD4^+$ T cell percentage of 19 (IQR, 10-26).

When the subjects were stratified by location, 28% of subjects resided on Thika Road, 25% in Kawangware, and 14% in Korogocho. Kawangware had the largest proportion of X4 variants 10 (50%), while Donholm 0 (0%) had the least. CCR5-tropic viruses predominated at 76% (n=16), 92% (n=12), 100% (n=4), 82% (n=9), 67% (n=2), and 80% (n=4) for the Thika road, City, Donholm, Korogocho, Langata, and Majengo SWOP clinics, respectively (Table 11).

CD4+ T cell counts in subjects infected with CCR5-tropic and CXCR4-tropic viruses were compared to assess the extent to which the virus's replication without treatment has harmed the $CD4^+$ T cell population (**Figure *14***). Two outliers on the CXCR4 upper tails (p = .020) were excluded. Subjects harboring CXCR4-tropic viruses had statistically significant lower baseline $CD4^+$ T-cell counts 289±189 copies/mm^3 while those infected with CCR5-tropic viruses had 433±260 copies/mm^3 (p = .044). In graph (B) Subjects infected with CXCR4-tropic viruses had lower CD4 + T-cell percentages than those infected with CCR5-tropic viruses, Wilcoxon rank sum test (P = .042). The levels of CD4 + T cell count in subjects infected with HIV-1 A1, C, and subtype D was compared using Kruskal–Wallis test **(Figure *14* graph C)** but there were no statistically significant differences among the different HIV-1 subtypes, (P = .98).

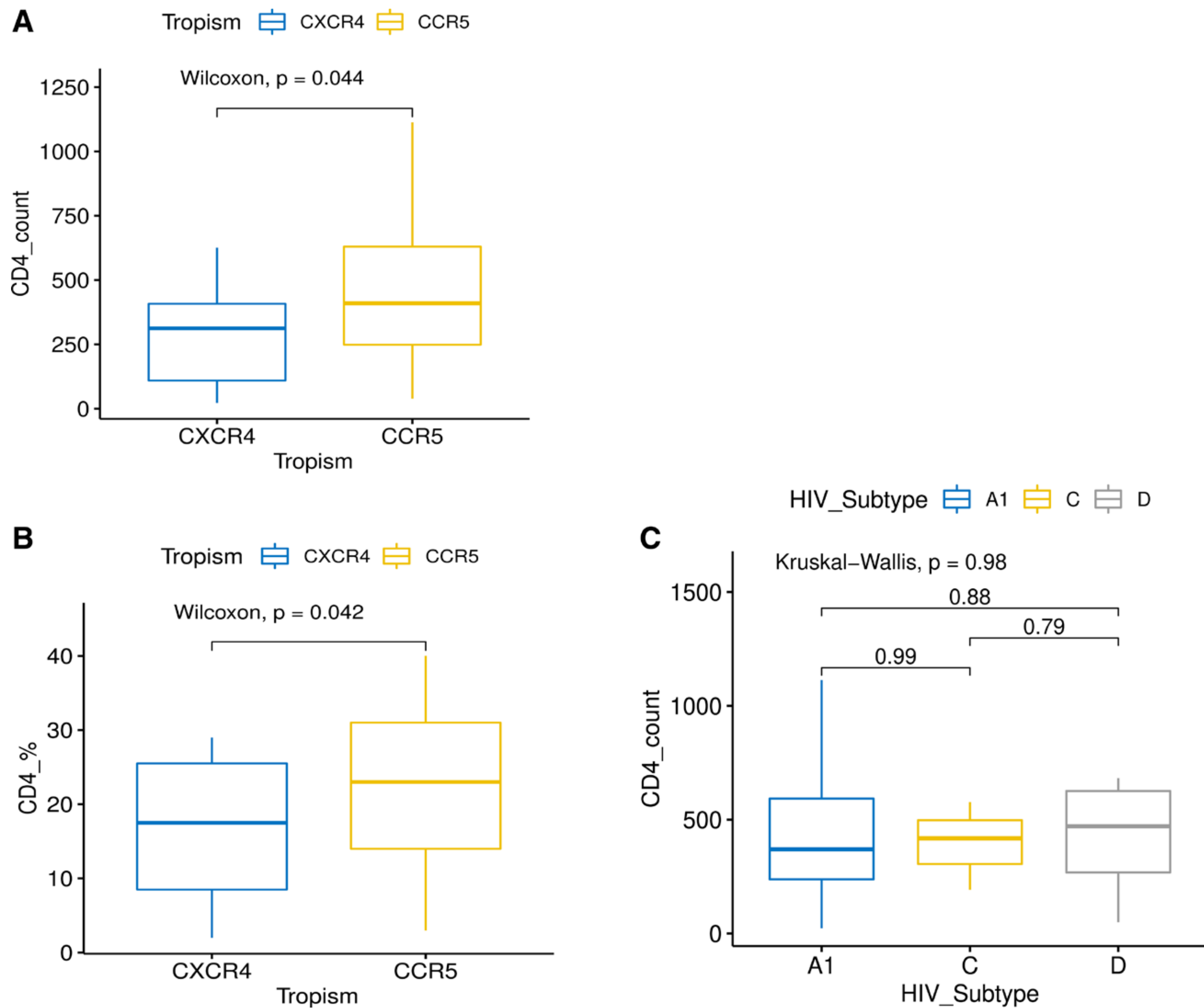
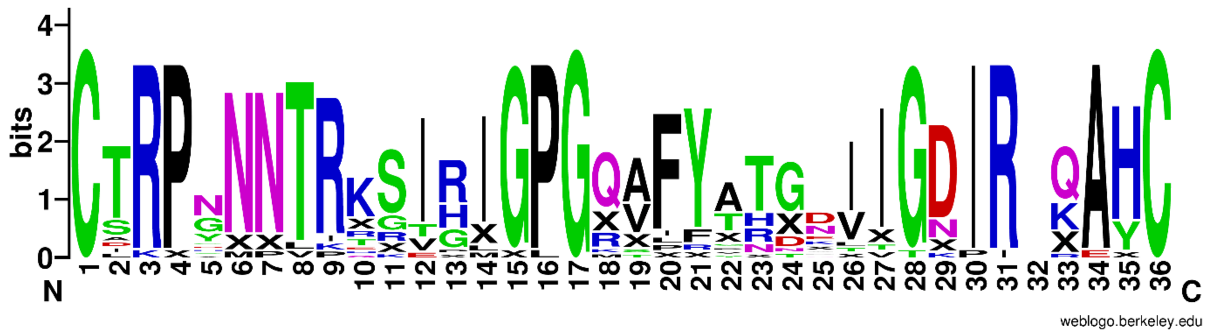


Figure 14: Comparisons of (a) median baseline CD4 + T cell counts and CD4 and (b) median baseline T cell percentages between CCR5 and CXCR4 tropisms. (c) Comparison of CD4 T count in subjects infected with different HIV-1 subtypes.
CCR5 = C-C motif chemokine receptor 5, CXCR4 = CXCR4 C-X-C motif chemokine receptor 4.

3.5.8 V3 loop sequence diversity

Genetic characteristics of the translated V3 loop sequences were profiled to assess their impact on tropism prediction. The average length of the V3 loop was 35 [34-36] amino acids. Most sequences had 35 amino acids 60 (79%), while 15 (20%) were 34 amino acids long. One sequence was 36 amino acids long. Five (33%) of the CXCR4 tropic amino acids were 34 AA long, while 14 (23%) were 35 AA long.

A



B

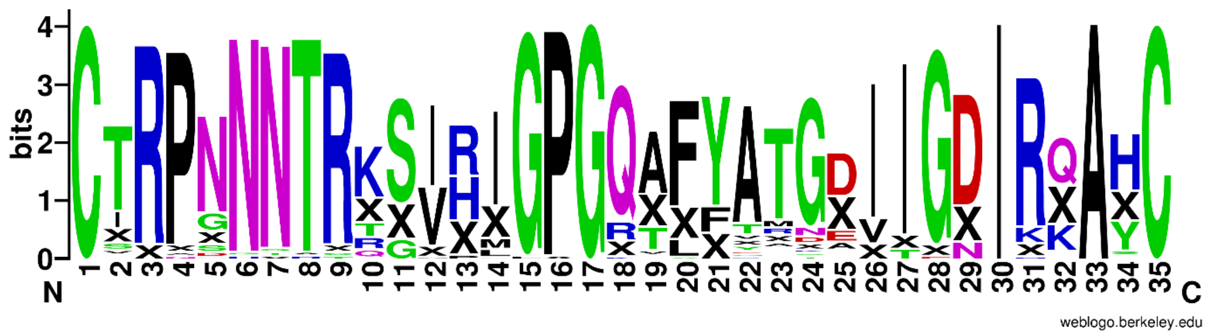


Figure 15: Sequence logos of CXC4 and CCR5 tropic HIV-1 V3 loop data sets.
(A) CXCR4 tropic variants. (B) CCR5 tropic variants

There was no correlation between the length of the V3 loop and tropism ($\chi2$=3.46, df=2, P = .178). The highly conserved GPGQ motif at the crown was noticeable in both the X4 and R5 viral strains **(Figure *15*).** The most notable result of this study was the relatively higher occurrence of Alanine at position 22 of the V3 loop in R5 strains than in X4 strains. Isolates with Alanine at position 22 of the V3 loop accounted for 80% (n=45) of the R5 strains. The occurrence of this specific residue at position 22 was also significantly associated with viral tropism ($\chi2$=15.586, df=8, P = .048). We did not observe any correlation between the occurrence of positively charged amino acid at position 11 ($\chi2$=6.478, df=2, P = .166), or position 25 ($\chi2$=19.967, df=10, P = .075) with viral tropism **Table 12**.

Characteristic	OR[1]	Univariable 95% CI	*P*-value	OR[2]	Multivariable 95% CI	*P*-value
Net charge	1.75	1.20, 2.70	**.006**	2.40	1.35, 5.00	**.007**
$CD4^+$ T cell count	0.88	0.76, 0.97	**.039**	1.00	0.93, 1.81	.093
Amino acid at position 22						
A	—	—		—	—	
Other (E,G,P,R,V)	2.41	0.45, 10.9	.27	1.21	0.17, 7.37	.84
T	12.7	3.32, 57.0	**<.001**	55.7	4.04, 84.1	**.003**
SWOP clinic						
Thika Road	—	—		—	—	
Kawangware	3.56	0.96, 14.7	.065	6.94	1.22, 48.81	**.034**
City	0.27	0.01, 1.95	.25	2.33	1.04, 14.6	.89
Donholm	0.00	0.0, ∞	>.99	0.00	0, ∞	>.99
Korogocho	0.71	0.09, 4.10	.72	0.22	0.01, 2.48	.26
Langata	1.60	0.07, 20.6	.27	4.89	0.16, 116	.30
Majengo	0.80	0.04, 7.21	.86	4.00	0.14, 71.0	.34
HIV-1 subtype						
A1	—	—				
C	0.00	0.00, ∞	>.99			
D	4.50	0.69, 36.4	.12			

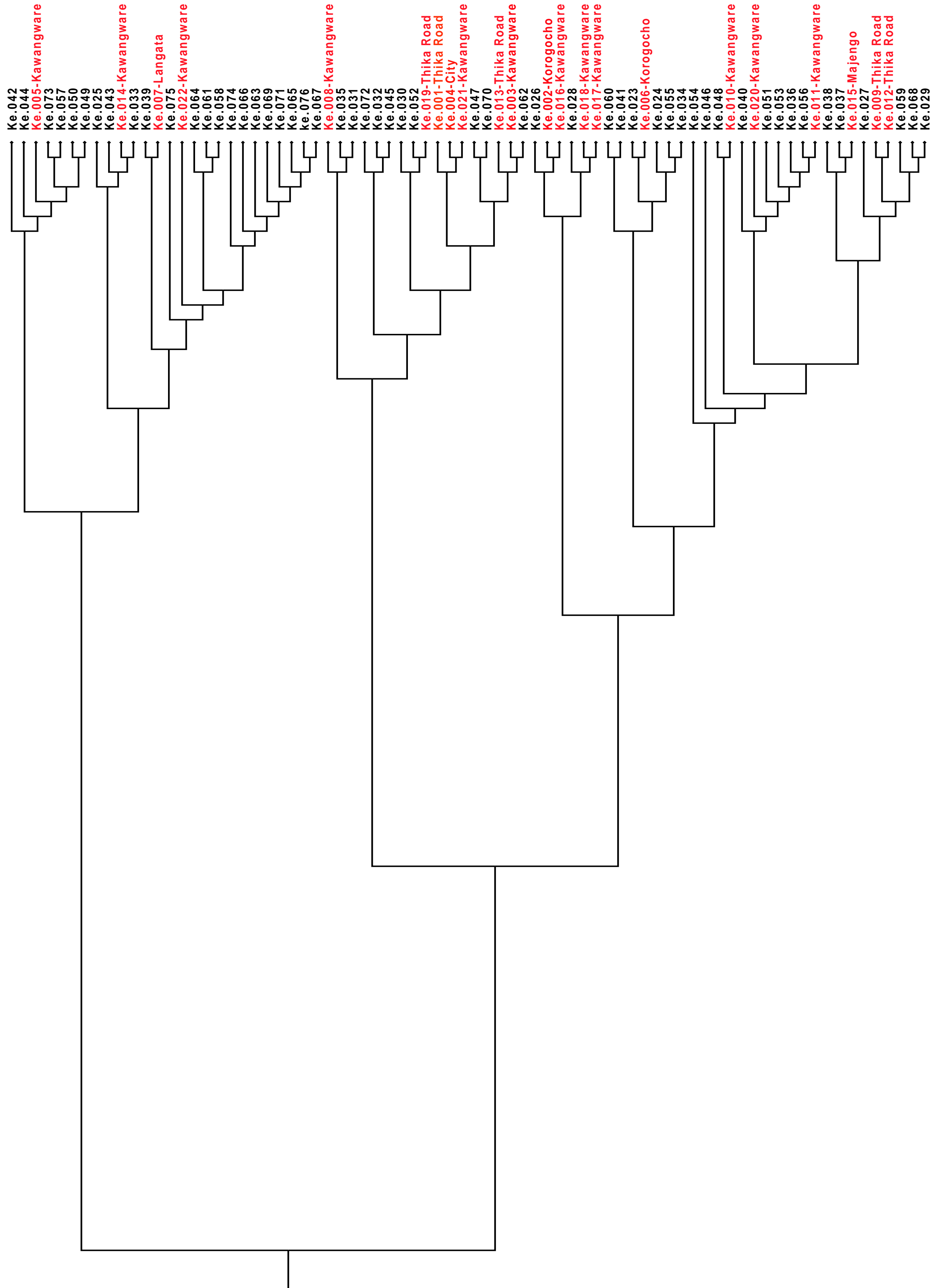
Table 12: Factors associated with coreceptor usage in 76 HIV-1 treatment naïve subjects
OR[1] = unadjusted Odds Ratio, **CI** = Confidence Interval, **OR[2]** = adjusted Odds Ratio, **p - values** from χ2 test for categorical variables or Wilcoxon rank sum test for continuous variables
SWOP = Sex work outreach program, **A** = Alanine, **E** = Glutamic acid, **G** = Glycine, **P** = Proline, **T** = Threonine, **R** = Arginine, **V** = Valine. Bold value indicates P < 0.05.

The presence of a positively charged residue at position 11 and or 25 of the V3, commonly termed the '11/25' rule was significantly associated with viral tropism11 (χ2=44.96, df=4, *P* <.001) 25 (χ2=54.848, df=10, *P* < .001). A significant association was observed when positive net charge and tropism were compared, with CXC4 tropic viruses exhibiting a higher net charge (+6) compared to their CCR5 counterparts (+4). The amino residue at position 22 was not associated with the net charge (χ2=61.741, df=56, *P* = .278), and did not introduce multicollinearity in the final multivariable logistic regression model. There was no association between HIV-1 and residues at position 11 subtype (χ2=5.894, df=8, *P* = .659), or 22 (χ2=20.605, df=16, *P* =0.194) of the V3 loop. In contrast, the amino acid at position 25 was significantly associated with HIV-1 subtype (χ2=45.973, df=20, *P* < .001).

3.5.9 Phylogenetic tree analyses

The ML phylogenetic tree of the 76 recently infected subjects in this study revealed three statistically supported nodes (bootstrap > 90%). In the phylogenetic tree, the sequences were shown to form two divergent clusters with an uneven distribution of CXCR4-tropic viruses (highlighted in red) in the lower arm (Fig 3). There was a tendency of subjects recruited from Kawangware SWOP clinic to be CXCR4- tropic though there is no clear evidence of an exclusive CXCR4-tropic cluster nor a SWOP clinic cluster.

Figure 16; Schema of the phylogenetic tree of HIV-1 viruses circulating in this cohort of treatment naïve commercial female sex workers. In red are CXCR4 tropic variants from different SWOP clinic locations, distributed throughout Nairobi.



Chapter 4

4.0 Discussion

HIV infection is disproportionately high in high-risk populations such as female sex workers (FSWs), men who have sex with men (MSM), and injection drug users (IDUs). Approximately 6.6% of the Kenyan population are infected by HIV-1, with a more than three-fold higher HIV-1 prevalence among so-called high-risk groups. These groups also account for at least one-third of new infections ((NASCOP), 2018). FSWs are amongst the most vulnerable of the key population groups and the risk for acquiring HIV is 30 times higher for sex workers (Bitty-Anderson et al., 2022). The reported national HIV incidence in FSWs was 29.3% in 2021, with a 4.5-fold increased risk compared non-sex worker females (Tago et al., 2021). This population is more likely to harbor diverse HIV-1 variants and recombinants, due to their social vulnerability and the risks associated with their occupation including multiple sexual partners, inconsistent condom use, use of injecting and non-injecting drugs, and co-infections with other STIs (Giguère et al., 2019; Gitahi-Kamau et al., 2015; Nduva et al., 2020; Tago et al., 2021). These women are at an increased risk of acquiring and transmitting HIV to the general population and thus play a central role in sustaining population-level infections. This highlights the importance of focusing on such a critical population, to understand the trajectory of the pandemic's evolution, and thus inform public health mitigation strategies.

4.1 Viral tropism prediction

There is currently no consensus on the tool of choice for tropism prediction. Other than Geno2Pheno 15% which is well validated and recommended by European virologists for clinical tropism testing (Vandekerckhove et al., 2011), several widely validated subtype-specific tools have been developed. More than 2/3 of the isolates in our data set were CCR5 tropic according to all the GTT tools except WebPSSM [sinsi C] matrix, which overestimated CXCR4 tropism at 84% (Table 8**).** The WebPSSM sinsi-C algorithm which was specifically developed for HIV-1C viruses showed improved sensitivity in picking up X4 tropism (1/3), in subtype C strains. No dual/mixed tropic variants were detected.

According to the '11/25' algorithm, which determines tropism based on the occurrence of a positively charged amino acid residue at position 11 and or 25 of the V3 loop (Fouchier et al., 1992; De Jong et al., 1992), only 6.6% of the isolates were CXCR4 tropic against a score of >20% for the highly concordant Geno2Pheno (Beerenwinkel et al., 2003b) and WebPSSM algorithms (Jensen et al., 2003) Table 9. This agrees with previous studies that have reported a relatively higher prevalence of X4 viruses in treatment experienced patients, compared to their treatment

naïve counterparts (Sánchez et al., 2010; Soulié et al., 2008). The training data for the 11/25 algorithm and many other GTT tools utilizing other aspects of the V3 loop such as position specific matrices were based on clade B virus characteristics such as a 35 amino acid V3 loop length. This may have limited the sensitivity of this tool since the length of the V3 loops in this data set ranged from 34-36 amino acids. The HIVCoR algorithm which was exclusively developed for CRF01_AE isolates (Hongjaisee et al., 2019), assigned CRR5 tropism to all the isolates within our data set. The CCR5 dominance within this cohort may be due to the limited sensitivity of GTT tools in predicting tropism for non-B subtypes, since many of these tools were developed using clade B training data sets. Epidemiological studies in the European, America and Australian populations where subtype B is endemic, have shown that 80–90% of HIV-1 positive individuals are infected by R5 tropic HIV-1 strains (Hoffman, 2007). Therefore, this bias in favor of B variants is likely to overestimate R5 tropism in areas where non-B HIV-1 subtypes predominate such as Kenya. These discrepancies in tropism prediction underscore the need for subtype-specific algorithms or algorithms with increased sensitivity to a wider range of HIV subtypes.

4.2 Concordance of tropism prediction algorithms

We found a 26% prevalence of CXCR4 tropic viruses among the five independent algorithms using the Kuder-Richardson 20-formula (Cronbach, 1951). This formula calculates the reliability of the internal consistency of the genotypic tropism testing (GTT) tools, or the similarity in tropism prediction scores for the GTT tools with an acceptable cutoff of 70%. We further assessed the reliability of the Geno2Pheno algorithm Table 7) and WebPSSM matrices (Table 8) in assigning tropism to viral populations dominated by COMET assigned HIV-1 subtype A, with subtype C and D as minority variants (Table 6**).** Overall, the 93% (95% CI 90-95%) concordance within FPR variations of the Geno2Pheno algorithm was relatively higher than the 52% (95% CI 29%-68%) within WebPSSM matrices (Table 9). The Geno2Pheno 5% FPR algorithm was the least concordant among the FPR variations of this tool, and dropping it increased the concordance to 98% (95% CI 0.97-0.98) for Geno2Pheno [coreceptor FPR:10- 20%]. Geno2Pheno [coreceptor FPR:10- 20%], and WebPSSM$_{[X4R5]}$ were highly concordant 88% (95% CI 79%–90%), in line with previous studies which demonstrated the best concordance (>85%) between these two algorithms (Seclén et al., 2011).

4.3 HIV-1 subtypes

HIV-1 subtypes and circulating recombinant forms (CRFs) continue to evolve globally, altering their transmissibility, cellular tropism, viral replication kinetics, and disease progression (Taylor and Hammer, 2008). HIV-1 subtype A1 accounted for more than three-quarters of the variants detected by all the HIV-1 subtyping tools within our data set. Subtype D came in second at 7%, and subtype C a close third at 4% (Table 6). These findings are in tandem with similar studies conducted in Kenya in which subtypes A1, and D were the most prevalent, while subtypes A2, C, and G accounted for the minority subtypes (Kitawi et al., 2017; Lihana et al., 2009; Lwembe et al., 2009; Mabeya et al., 2018). The predominance of subtype A1 has been attributed to the relatively higher replicative fitness of this clade, compared to clade D (Gounder et al., 2017; Nduva et al., 2020; Tovanabutra et al., 2010). Subtype D has been reported to be decreasing in prevalence, consistent with the notion that more pathogenic variants may be less fit for transmission (Frank et al., 2019; Mckinnon et al., 2012; Nduva et al., 2020).

4.4 Concordance of HIV-1 subtyping tools

Cronbach's alpha test was also employed to determine agreement among the HIV-1 subtyping tools. These automated subtyping tools are routinely used in clinical practice due to their speed, user-friendliness, and low cost (Pineda-Peña et al., 2013; Yebra et al., 2011). Concordance for COMET, REGA, and NCBI viral subtyping tools was lower than 0.7 (Table 9). This reduced concordance was attributed to the NCBI viral subtyping tool since dropping it increased the reliability to 0.84. The COMET and REGA tools were previously reported to be (93.8%) and (92.5%) concordant respectively, with Mphy (molecular phylogeny), the gold standard in subtyping for epidemiological surveillance and routine clinical practice by (Fabeni et al., 2017). In the same study, COMET also showed high sensitivity (92.3%) to A1 subtypes while G subtypes were better classified by REGA (98.6%) (Fabeni et al., 2017). Eighty six percent of the isolates within our data set were predicted to be A1 by COMET, while one was subtype G (Table 6). The REGA tool did not assign the HIV-1G subtype to any of the sequences (Table 6). Although REGA had the highest calculated reliability here, this tool was unable to assign tropism to 13% of the isolates within our data set. COMET based subtypes were therefore chosen due to their high sensitivity towards A1 variants-the most prevalent subtype in Kenya and their 100% success rate in assigning HIV-1 subtypes (Table 6**).**

4.5 Predictors of coreceptor usage

A multivariate logistic regression model was used to investigate the effect of multiple independent variables on the predicted tropism **(Table 12).** We sought to understand how each risk factor contributed to tropism variation (if any) within this high-risk group, and for the findings herein to mold better and more effective public health policies and research efforts targeted toward this group. Our model was 90% accurate when validated using the K-fold cross-validation approach. Consistent with previous studies (Claudia et al., 2014; Naganawa et al., 2008), net charge and tropism were significantly correlated within this data set ($p = 0.006$). The odds of an isolate being X4 were 240 times higher for every unit increase in net charge (0R = 2.40, 95% CI = 1.35-5.00, $p = 0.006$). Despite the correlation observed between CD4 T cell count and tropism in the univariate model, the effect of CD4 T cell count was rendered insignificant (95% CI 0.93-1.81, $p = 0.093$) in our multivariate model by the confounding effects of net charge, the amino acid at position 22 of the V3 loop, and convergence of CXCR4 tropic variants at the Kawangware SWOP clinic **(Table 12).**

4.5.1 HIV-1 subtype

The biggest caveat of GTT stems from subtype-specific differences that affect the prediction of tropism. Studies have demonstrated a correlation between genetic subtype and coreceptor tropism evolution throughout the disease course (Björndal et al., 1999; Cilliers et al., 2003; Kaleebu et al., 2007). HIV-1 subtype was not a significant predictor of viral tropism in this data set using the univariate logistic regression model **(Table 12).** Despite earlier research indicating that subtype D is linked to CXCR4 tropism in earlier stages of infection (Cilliers et al., 2003; Connell et al., 2008; Kaleebu et al., 2007), we did not find any conclusive evidence of such an association. We cautiously hypothesize that this is due to the dominance of subtype A1, which obscured any differences from the minority subtypes, thereby reducing the ability of the test to detect differences.

4.5.2 CD4 T cell count

A profound decline in CD4 T cell count, the target cell of the HIV virus, and a rise in plasma viral load are some of the distinctive features associated with HIV-1 infection (Leda et al., 2020; Moyle et al., 2005) . If HIV infection remains untreated, opportunistic infections, the development of AIDS, and AIDS-related deaths may occur (Langford et al., 2007; Naif, 2013). As a result, the CD4 T cell count has been reported as a major prognostic marker of the progression of disease in HIV infection and was previously used to determine the priority for initiating antiretroviral therapy

(Zhang et al., 2011). We found a statistically significant correlation between low CD4 T cell count and CXCR4 tropism in this data set (**Figure *14***). In addition, when CD4 T cell count was modeled as a confounder of viral tropism (cut off $p= <.25$), the odds of an isolate being CXCR4 tropic were 88 times higher when the CD4 count was low (OR= 0.88 95% CI [0.76-0.97] $p = 0.039$). This implies that CXCR4 tropic isolates are more pathogenic and may play a major role in the depletion of CD4 T cells during early HIV infection (EHI), and possibly result in faster disease progression to AIDS in the absence of ART, as observed in this cohort of treatment naïve study participants (Ghosn et al., 2017; Sechet et al., 2015). This is of great importance in shaping public health policies and strategies, since early detection of HIV infection, timely disease progression monitoring, and early initiation of antiretroviral therapy are crucial steps to containing the spread of HIV.

Indeed, a correlation between tropism and CD4 T cell count has been demonstrated in studies involving larger cohorts (Raymond et al., 2009). In a study conducted by (Brumme et al., 2005), the X4 genotype in baseline plasma was an independent predictor of poor immunological response and increased mortality. This observation is in tandem with faster progression of disease observed for subtype D infections, which mostly exhibit the X4 genotype (Raymond et al., 2011). We did not find any significant association between the circulating subtypes and disease progression represented by CD4 T cell count ($p = 0.98$). A study conducted by (Baeten et al., 2007), reported that subtype D was correlated with increased rates of CD4 T cell decline and a 12-fold higher risk of death, compared to subtype A, in a cohort of female sex workers from Mombasa, Kenya. Rapid disease progression is most pronounced for subtypes D and B, while subtypes A and C are usually considered less aggressive (Mckinnon et al., 2012). These findings are, however, still being actively debated.

4.5.3 SWOP clinic location

Despite the global scale-up on HAART that has seen a reduction in HIV/AIDS-related morbidities, mortalities, and prolonged average life expectancy of infected individuals, HIV incidence in marginalized groups remains disproportionately high. Female sex workers (FSWs) are 13.5 times more likely to contract the virus compared to their non- sex worker counterparts (Baral et al., 2012), highlighting the role of various HIV-1 transmission pathways, and risky behavior in this process. To address this vulnerable population, SWOP clinics were established in 2008 to scale-up accessible, acceptable, and friendly HIV prevention and treatment services. This program

operates seven clinics throughout Nairobi County, all of which are located near hot spots where sex workers are likely to find clients. We found a statistically significant convergence of CXCR4-tropic strains at the Kawangware SWOP clinic **(Table 12).** Since X4 viruses are associated with increased risk of disease progression and increased mortality in untreated individuals, increased CD4 T cell decline, and poor response to treatment (Brumme et al., 2005; De Jong et al., 1992; Koot et al., 1992; Ping et al., 2013), this site should be further studied to better understand HIV epidemiology at SWOP clinic location.

4.5.4 Features of the env V3 loop

HIV tropism has also been associated with various amino acid signatures within the V3 loop (Soulié et al., 2010). A notable finding within this study was the correlation between tropism and the amino acid at position 22, in the subtype A1 V3 loop (P =.003). The most dominant residue at position 22 of the V3 loop was Alanine for CCR5 tropic viruses and Threonine for CXCR4 strains **(Figure *15*).** Although this 22-AA genotype does not occur frequently in the widely studied subtype B, there is a notable increase in the infectivity of the virus, when it does (Douek et al., 2003; Wei et al., 2016). There is still no definitive explanation for the observed differences in tropism following the switch of uncharged amino acids (alanine to threonine) at this position. Marozsan et al., used the Garnier-Osguthorpe-Robson (GOR) method to further investigate the role of this switch. There were no notable differences in tropism, asserting the predominance of Alanine in R5 variants as opposed to X4 variants of clade B viruses (Cormier and Dragic, 2002). To assess the relative fitness of a given HIV clade, (Ariën et al., 2005) set up a laboratory growth competition experiment that measured viral populations infecting peripheral blood mononuclear cells. The order of relative fitness was subtypes A, B, D, CRF01_AE > subtype C > HIV-2 > group O. This warrants further investigation to ascertain whether the increased fitness HIV-1 subtype A is associated with this amino acid switch. Contrary to expectations, we did not find a significant correlation between CXCR4-tropism and the occurrence of positively charged amino acid residues at positions 11 and or 25 of the V3 loop ($p = 0.22$) (Fouchier et al., 1992; Soulié et al., 2016).

4.6 Degree of relatedness of the viruses

The persistence of the HIV epidemic has been linked to the existence of numerous transmission clusters, frequently associated with HIV transmission (Kouyos et al., 2010b; Petersen et al., 2018). This is especially important given that FSWs are one of the most vulnerable groups in HIV transmission. There was a statistically significant convergence of CXCR4 tropic variants at the

Kawangware SWOP clinic despite the absence of a predominant CXCR4 tropic cluster at each clinic location **(Figure 16).** This is expected since studies have shown patients are likely to visit clinics located far from their place of work due to the stigma associated with commercial sex work (Wanjiru et al., 2022).

4.7 Study limitations and recommendations

4.7.1 Study limitations

Some limitations of this study warrant mention. First, viral load data was unavailable. In low - resource settings such as Kenya, cost, and operational challenges limit viral load testing alongside other clinical and immunological biomarkers that may be key players in viral tropism (Joram et al., 2017). Since viral load testing is conducted on a need-to-know basis, lack of VL information may have been a possible limitation in the selection of 'viable' samples. According to the European guidelines on tropism testing, a minimum viral load of >500 RNA copes/ml, and triplicate PCR amplification and sequencing for each sample is recommended (Vandekerckhove et al., 2011).

We also experienced a relatively high rate of non-amplification (50%), which affected the sample size. A sample size of 76 subjects may be small and underpowered on the surface, but this was not the case. In logistic regression, the occurrence of an event defines the dependent (outcome) variable, and the rule of thumb takes into consideration the number of subjects to whom the event occurs rather than the total number of subjects. In this study, we predicted the occurrence of CXCR4-tropic viruses by assuming that approximately 10% of the subjects had CXCR4 and applied a model with 5 independent (predictor) variables. As a result, the minimum sample size would need to be 10-15 events per independent variable, which adds up to 50-75. Despite the justification, we cannot completely rule out the possibility of low study power in circumstances where an uneven distribution of cases in a specific variable led to contradictory findings.

Another limitation of this study was the lack of phenotypic tropism testing (PTT) on the same plasma samples. Since PTT is considered the gold standard for assigning coreceptor tropism, the accuracy and efficiency of the algorithms used herein cannot be validated. Tropism prediction for pure non-B, and co-circulating multiple subtypes, and inter-subtype recombinant forms remains a challenge, necessitating the need to document coreceptor tropism of a diverse panel of non-subtype B HIV-1 variants and emerging recombinant forms, to develop novel GGT tools and or optimize

training data sets of the available GTT tools for improved sensitivity and specificity to non-B subtypes.

4.7.2 Study recommendations

The source of viral RNA in the study was Plasma samples from HIV-1 positive treatment naïve female sex workers. We experienced low yield (50%), with the minority variants (CXCR4) constituting only a quarter of the isolates (**Table 9**). Studies have shown that using proviral DNA may improve yield and detection of X4 variants since CXCR4 tropic isolates are commonly found in peripheral blood mononuclear cells (Vandekerckhove et al., 2011; Verhofstede et al., 2009).

Using next-generation, deep sequencing methods has also been reported to improve the sensitivity of GTT tools to X4 an dual/mixed patient samples (Leda et al., 2020; Panos and Nelson, 2007). Conventional sanger sequencing methods have an intrinsic limited sensitivity to minority viral species such as X4 and X4R5. These viruses most often occur with R5 variants forming a mixed quasispecies, and are undetectable when they represent <10-25% of the population (Verhofstede et al., 2011). With deep-sequencing, thousands of PCR products (even from different HIV-1 variants), can be sequenced (read) individually and in parallel, enabling the detection and quantification of minority subpopulations (Liang et al., 2011). The technique however, is not error free and must be analyzed carefully so as not to overestimate diversity within the viral population (Leda et al., 2020).

Improving tropism prediction may not solely rely on deeper sequencing but also on the analysis of non-V3 regions that have been cited as possible additional tropism determinants. Even though tropism determinants of subtype B viruses have been widely studied, there is scarcity of information on more prevalent subtypes such subtype C, and subtype A and its CRFs, which account for than 50% and 25% global infections, respectively (Digban et al., 2020; Gounder et al., 2017). These regions, some of which are yet to be fully characterized, differ between HIV-1 subtypes and influence the propensity of certain virus subtypes to use either the CRR5 or CXCR4 coreceptor. The V1/V2 and the C4 regions have been shown to influence coreceptor use (Hartley et al., 2005; Pastore et al., 2006b). Profiling of clade A gp120 sequences by (Riemenschneider et al., 2016) led to the detection of a statistically higher association between the V2 loop with tropism, compared to the V3 loop. The V1-V3 loops were also cited as the major tropism determinants by (Esbjörnsson et al., 2010). Additional studies on clade A and C viruses have suggested the varying

role of other coreceptors, namely CXCR6 and GPR15, in subtype specific HIV-1 entry (Karlsson et al., 2012; Pohlmann et al., 1999). Isolates from Central African Republic where subtype A is predominant have been shown to extensively use these minor coreceptors (Bégaud et al., 2003).

4.8 Conclusion

Although the REGA tool had better calculated reliability in assigning HIV-1 subtype to the isolates in this data set, this tool failed to assign a HIV subtype to more than ten isolates. The COMET subtyping tool successfully assigned a HIV-1 subtype to all the 76 partial envelope sequences analyzed, with high reliability. In line with these findings and previous studies that have reported the high sensitivity of this tool to HIV-1 subtype A variants, we recommend this tool for subtyping Kenyan HIV isolates which are dominated by subtype A1.

We found a relatively higher prevalence of X4 viruses in HIV-1 subtype A1 (26%) than previously thought to be present, a major implication in the adoption of the selective CCR5 antagonist Maraviroc. Since tropism testing is conducted to rule out the outgrowth of CXCR4 viruses as a pathway to treatment failure, the higher occurrence of CXCR4 variants may adversely affect the adoption of the CCR5 antagonist Maraviroc. Further studies are also required to assess the sensitivity of the dominant Kenyan HIV-1 variants to Maraviroc due to reported subtype-specific differences in sensitivity to this CCR5 antagonist (Siddik et al., 2018).

Variations of the Geno2Pheno tropism prediction tools had better agreeability of 93% compared to variations of the WebPSSM matrix (52%). Geno2Pheno 10%, 15%, and 20% FPR are usually recommended for tropism testing. These algorithms also had the highest calculated reliability in assigning tropism to the isolates in this data. We recommend Geno2Pheno 15% FPR for clinical management of viral tropism in line with the European guidelines and previous studies, because at this FPR there is a balance between favorable sensitivity to the minority variants (X4/R5X4) and concordance with other tools for detection of majority variants.

The most notable finding was the significant association between CCR5 tropism A1 viruses and the presence of the amino acid Alanine at position 22 of the V3 loop. Additional studies with a large data set are warranted to confirm our findings.

4.9 Appendices

4.9.1 Waiver of consent application

KNH-UoN/ERC/FORM/IC05

UNIVERSITY OF NAIROBI (UoN)
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
(254-020) 2726300 Ext 44355

KNH-UoN ERC
Email: uonknh_erc@uonbi.ac.ke
Website: http://www.erc.uonbi.ac.ke
Facebook: ttps://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC

KENYATTA NATIONAL HOSPITAL (KNH)
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

(To be submitted with Application for ERC Review of Research)
Exempt studies to be defined

KNH-UoN ERC
REQUEST FOR WAIVER OF INFORMED CONSENT
(Not Required for Exempt Studies)

Project Title: **HIV-1 genetic diversity and transmission dynamics in sex workers cohort in Nairobi – Kenya.**

Principal Investigator and Institutional affiliation: **Dr. Martin MULINGE, University of Nairobi**

Date: Sept 29, 2019

Under special circumstances, investigators may request one of three types of waivers to obtaining written informed consent from research participants.

1. **Alteration of informed consent.**
   With this waiver, the investigator may provide to the participants a consent which does not include or which alters one or all of the required elements. Examples of when this waiver might be applicable would be, when a researcher is conducting secondary data analysis and the participants cannot be located or when requiring informed consent might somehow actually have negative consequences for research participants.
2. **Waiver of parental permission.**
   This waiver would be used in cases where something may be legal for a child to do (i.e. contraception) without parental permission and obtaining parental permission would violate that privacy. An example of this type of waiver would be a survey on children (which would require parental permission) but the survey is about their experience on contraception usage.
3. **Waiver of written documentation** that informed consent was obtained. With this waiver, the investigator would be required to read or provide the informed consent form to a participant, but would not need to obtain the participant's signature on the consent form.

Examples of when this waiver might be applicable would be some internet or phone surveys or when signing the form might have some negative consequence for the participant. It must be emphasized that these waivers will be given only when there are compelling reasons for doing so.

The Ethics and Research Committee determines which type of consent applies to your research, but please indicate the type that you are requesting.

☒ **Waiver or alteration of the informed consent process.** *(Complete Section I)*

☐ **Request for waiver of parental permission.** *(Complete Section II)*

☐ **Waiver of written documentation of consent.** *(Complete Section III)*

I. Request for waiver or alteration of the consent process (Not required for Exempt studies)

I believe that this protocol is eligible for waiver or alteration of required elements of the informed consent process because the protocol meets all of the following criteria: (Provide protocol-specific supporting information for each criterion that justifies the findings for the following :)

1. The research presents no more than "minimal risk" of harm to participants.

From its definition., **"Minimal risk" is "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in the daily life or during the performance of routine physical or psychological examinations or tests" (Protection of Human Subjects 2009).**

This study represents no more than minimal risk to subjects and will not adversely affect their rights and welfare. Due to the impracticality of contacting those individuals we are requesting waiver of consent. By the time the 1mL of archived plasma is delivered to the bench for RNA extraction, all study participant identifiers will have be removed and the sample will be identifiable using a code. Furthermore, this study is not profiling the genome of the subjects but the viral genome.

2. The waiver or alteration will not adversely affect the rights and welfare of the participants.
The risks and welfare of the subjects will not be adversely affected by the waiver. The study subjects cannot be contacted by the researchers since they receive the samples when they are already coded. All data collected for this study will be restricted solely to the codes and will not be linked to the individuals who donated it.

3. The research could not practicably be carried out without the waiver or alteration.
Securing a second consent from study participants for the second time since it was included at the time of collection of sample may not be not feasible for all the 150 study participants, due to constraints of cost and time.

4. Whenever appropriate, the participants will be provided with additional pertinent information after participation.

Debriefing (process is an opportunity to provide subjects with information not disclosed during the initial consent process) will not be done.

For this study, we belief debriefing if it were to be done, it would cause more harm than good because of the nature of sexual work.

5. Elements of informed consent for which a waiver or alteration is requested and the rationale for each:

No new subjects will be recruited for this cross-sectional study. However, at the time of enrolment, the study subjects consented with a broad approval to their left-over blood samples being archived for futuristic study on HIV. All elements of informed consent are being asked to be waived for this study.

6. The research does not involve non-viable neonates:

Neonates are NOT involved in this Research

7. The research is not subject to FDA and/or national research regulation:

This is not applicable

II. Request for waiver of parental permission (Not required for Exempt studies)

I believe that this protocol is eligible for waiver of parental permission because the protocol meets all of the following criteria: (Provide protocol-specific supporting information for each criterion that justifies the findings for one of the following two options :)

Option 1

1. The research presents no more than "minimal risk" of harm to participants.

___

2. The waiver or alteration will not adversely affect the rights and welfare of the participants.

___

3. The research could not practicably be carried out without the waiver or alteration.

___

4. Whenever appropriate, the participants will be provided with additional pertinent information after participation.

___

5. Elements of informed consent for which a waiver or alteration is requested and the rationale for each:

____

____

6. The research does not involve non-viable neonates:

____

____

7. The research is not subject to FDA and/or national research regulation:

____

____

Option 2:

1. The research protocol is designed for conditions or for a participant population for which parental or guardian permission is not a reasonable requirement to protect the participants (for example, neglected or abused children)

____

____

2. An appropriate mechanism for protecting the children who will participate as participant in the research will be substituted

____

____

3. The research is not subject to FDA and/or national research regulation:

____

____

4. The waiver is consistent with international and national law:

____

____

III. Request for waiver of written documentation of consent (Not required for Exempt studies and not required when the consent process is waived.)

I believe that this protocol is eligible for a waiver of written documentation of informed consent because the protocol meets one of the following criteria: (Provide protocol-specific supporting information for each criterion that justifies the findings for one of the following two options :) **(NOTE: Even when documentation of informed consent is waived, the investigator is required to give participants full consent information, and to obtain their voluntary consent orally.)**

Option 1

(Example: Conducting interviews with street children engaged in drug abuse. The only record of the name or other identifying information of the participants would be the signed consent form and

knowledge of an individual's participation or information provided could lead to potential legal, social, or physical harm.)

Explain:

1. The only record linking the participant and the research would be the consent document.

2. The principle risk would be potential harm resulting from breach of confidentiality.

3. Each participant will be asked whether the subject wants documentation linking the participant with the research and the participant's wishes will govern.

4. The research is not subject to FDA and / national research regulation.

Option 2

(Example: Using an anonymous survey consent or conducting telephone interviews with politicians about how constitutional provision for funding of political parties will affecting the campaign process of smaller parties

1. The research presents no more than minimal risk of harm to participants.

2. The research involves no procedures for which written consent is normally required outside of the research context.

Approval (KNH-UoN ERC Chairperson: Check all that apply to indicate that the waiver or alteration is approved and to indicate agreement with the investigators protocol specific findings justifying the waiver.)

☐ Waiver or Alteration of the Consent Process

☐ Waiver of parental permission

☐ Waiver of Written Documentation of Consent

NOTE: To approve a waiver of written documentation of informed consent the investigator must provide a written document describing the information to be disclosed. This document has to include all required and appropriate additional elements of consent disclosure, unless the consent process has been altered.

Chose one of the following when approving a waiver of written documentation:

- ☐ The investigator must provide a written description of the information provided orally to the participant.
- ☐ The investigator does not have to provide a written description of the information provided orally to the participant.

APPROVED BY CHAIR KNH-UoN ERC:

Name: ______________________ Signature ______________________

Date and Stamp: ______________________

4.9.2 Ethical approval for nested study

UNIVERSITY OF NAIROBI
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KNH-UON ERC
Email: uonknh_erc@uonbi.ac.ke
Website: http://www.erc.uonbi.ac.ke
Facebook: https://www.facebook.com/uonknh.erc
Twitter: @UONKNH_ERC https://twitter.com/UONKNH_ERC

KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref: KNH-ERC/A/388

16th October, 2019

Dr. Martin Mulinge
Principal Investigator
KAVI- Institute of Clinical Research
College of Health Sciences
University of Nairobi

Dear Dr. Mulinge,

RESEARCH PROPOSAL: HIV-1 GENETIC DIVERSITY AND TRANSMISSION DYNAMICS IN SEX WORKERS COHORT IN NAIROBI, KENYA (P556/07/2019)

This is to inform you that the KNH- UoN Ethics & Research Committee (KNH- UoN ERC) has reviewed and **approved** your above research proposal. The approval period is 16th October 2019 – 15th October 2020.

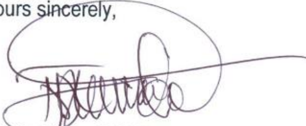
This approval is subject to compliance with the following requirements:

a. Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
b. All changes (amendments, deviations, violations etc.) are submitted for review and approval by KNH-UoN ERC before implementation.
c. Death and life threatening problems and serious adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH-UoN ERC within 72 hours of notification.
d. Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
e. Clearance for export of biological specimens must be obtained from KNH- UoN ERC for each batch of shipment.
f. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
g. Submission of an *executive summary* report within 90 days upon completion of the study. This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/ or plagiarism.

Protect to discover

For more details consult the KNH- UoN ERC website **http://www.erc.uonbi.ac.ke**

Yours sincerely,



PROF. M. L. CHINDIA
SECRETARY, KNH-UoN ERC

c.c. The Principal, College of Health Sciences, UoN
The Director, CS, KNH
The Chairperson, KNH- UoN ERC
The Assistant Director, Health Information, KNH
Mentors: Prof. Omu Anzala (KAVI- ICR, UoN),
Prof. Walter Jaoko (KAVI- ICR, UoN),
Dr. Joshua Herbeck (Dept. of Global Health, University of Washington)

Protect to discover

4.9.3 Consent to futuristic use of collected sample

PARENT STUDY PROTOCOL

Statement of Consent:

~~ you agree to participate in the study, please sign below.

I, ____________________, have read or have had read to me, the consent form for the above study and have

scussed the study with ____________________.

I understand that the following (check the box only if you fully understand and agree with each statement):

- [ ] the goals of this research program are to study resistance and susceptibility to sexually transmitted infections
- [ ] enrolment is completely voluntary and I can withdraw from the study at any time
- [ ] blood, cervical and vaginal specimens will be required for this study and may be used for genetic studies
- [ ] any blood specimens previously collected may be used for this study
- [ ] a portion of my blood, cervical and vaginal specimens will be stored for future studies of the genes involved in resistance and susceptibility to HIV and other infections.

am willing to participate in the study.

Name of Study Participant____

gnature/Thumb print: ______________________________ Date: ____________

For clinic staff:

, ______________________________, have explained the nature and purpose of the above study to

____

Name of Clinic Staff: ____

gnature: ______________________________ Date: ____________

ssigned Study Number / Clinic Number ___ ___ ___

NB: All study participants will be issued with a copy of this information and consent forms

ate___ /___ /___ (dd/mm/yy)   Initials __________   Study # ___ ___ ___

References

(NASCOP), N.A. and S.C.P. (2018). Kenya HIV Estimates Report 2018.

Abbink, T.E.M., and Berkhout, B. (2003). A novel long distance base-pairing interaction in human immunodeficiency virus type 1 rna occludes the gag start codon. J. Biol. Chem. *278*, 11601–11611.

Agrawal, L., Lu, X., Qingwen, J., VanHorn-Ali, Z., Nicolescu, I. V., McDermott, D.H., Murphy, P.M., and Alkhatib, G. (2004). Role for CCR5 32 Protein in Resistance to R5, R5X4, and X4 Human Immunodeficiency Virus Type 1 in Primary CD4+ Cells. J. Virol. *78*, 2277–2287.

Agrawal, L., Jin, Q., Altenburg, J., Meyer, L., Tubiana, R., Theodorou, I., and Alkhatib, G. (2007). CCR5 32 Protein Expression and Stability Are Critical for Resistance to Human Immunodeficiency Virus Type 1 In Vivo. J. Virol. *81*, 8041–8049.

Ahmad, N., and Venkatesan, S. (1988). Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science (80-.). *241*, 1481–1485.

Albanese, A., Arosio, D., Terreni, M., and Cereseto, A. (2008). HIV-1 Pre-Integration Complexes Selectively Target Decondensed Chromatin in the Nuclear Periphery. PLoS One *3*, 2413.

Alimonti, J., Ball, T., and Fowke, K. (2003). Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. J. Gen. Virol. *84*, 1649–1661.

Allam, A., Majji, S., Peachman, K., Jagodzinski, L., Kim, J., Ratto-Kim, S., Wijayalath, W., Merbah, M., Kim, J.H., Michael, N.L., et al. (2015). TFH cells accumulate in mucosal tissues of humanized-DRAG mice and are highly permissive to HIV-1. Sci. Rep. *5*, 10443–10459.

Aloja, P.D.', Olivetta, E., Bona, R., Nappi, F., Pedacchia, D., Pugliese, K., Ferrari, G., Verani, P., and Federico, M. (1998). gag, vif, and nef Genes Contribute to the Homologous Viral Interference Induced by a Nonproducer Human Immunodeficiency Virus Type 1 (HIV-1) Variant: Identification of Novel HIV-1-Inhibiting Viral Protein Mutants. J. Virol. *72*, 4308–4319.

Altfeld, M., Addo, M.M., Eldridge, R.L., Yu, X.G., and Et al. (2001). Vpr Is Preferentially Targeted by CTL During HIV-1 Infection. J. Immunol. *167*, 2743–2752.

Alvarez, Y., Tuen, M., Shen, G., Nawaz, F., Arthos, J., Wolff, M.J., Poles, M.A., and Hioe, C.E. (2013). Preferential HIV Infection of CCR6 Th17 Cells Is Associated with Higher Levels of Virus Receptor Expression and Lack of CCR5 Ligands. J. Virol. *87*, 10843–10854.

Ames, J., Ahn, O.K., Ruce, B., and Alker, D.W. (1998). Acute Human Immunodeficiency Virus Type 1 Infection. N. Engl. J. Med. *339*, 33–39.

An, M., Han, X., Zhao, B., English, S., Frost, S.D.W., Zhang, H., and Shang, H. (2020). Cross-Continental Dispersal of Major HIV-1 CRF01_AE Clusters in China. Front. Microbiol. *11*, 61.

Anderson, J.P., Rodrigo, A.G., Learn, G.H., Madan, A., Delahunty, C., Coon, M., Girard, M., Osmanov, S., Hood, L., and Mullins, J.I. (2000). Testing the Hypothesis of a Recombinant Origin of Human Immunodeficiency Virus Type 1 Subtype E. J. Virol. *74*, 10752.

Ariën, K.K., Abraha, A., Quiñones-Mateu, M.E., Kestens, L., Vanham, G., and Arts, E.J. (2005). The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J. Virol. *79*, 8979–8990.

Arrildt, K.T., Joseph, S.B., and Swanstrom, R. (2012). The HIV-1 Env protein: A coat of many colors. Curr. HIV/AIDS Rep. *9*, 52–63.

Arthos, J., Cicala, C., Martinelli, E., Macleod, K., Van Ryk, D., Wei, D., Xiao, Z., Veenstra, T.D., Conrad, T.P., Lempicki, R.A., et al. (2008). HIV-1 envelope protein binds to and signals through integrin α4β7, the gut mucosal homing receptor for peripheral T cells. Nat. Immunol. *9*, 301–309.

Asjö, B., Morfeldt-Månson, L., Albert, J., and Et. al. (1986). plicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet *2*, 660–662.

Ayyavoo, V., Mahboubi, A., and Et al. (1997). HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor κB. Nat. Med. *3*, 1117–1123.

Bachelerie, F., Ben-Baruch, A., Burkhardt, A.M., Combadiere, C., Farber, J.M., Graham, G.J., Horuk, R., Sparre-Ulrich, A.H., Locati, M., Luster, A.D., et al. (2014). International union of pharmacology. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors. Pharmacol. Rev. *66*, 1–79.

Van Baelen, K., Vandenbroucke, I., Rondelez, E., Van Eygen, V., Vermeiren, H., Stuyver, L.J., Baelen, K. Van, Vandenbroucke, I., Rondelez, E., Eygen, V. Van, et al. (2007). HIV-1 coreceptor usage determination in clinical isolates using clonal and population-based genotypic and phenotypic assays. J. Virol. Methods *146*, 61–73.

Baeten, J.M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R.S., Certain, L., Mandaliya, K., Jaoko, W., and Overbaugh, J. (2007). HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J Infect Dis *195*, 1177–1180.

Bagashev, A., and Sawaya, B.E. (2013). Roles and functions of HIV-1 Tat protein in the CNS: an overview. Virol. J. *10*, 1–20.

Balasubramaniam, M., and Freed, E.O. (2011). New insights into HIV assembly and trafficking. Physiology *26*, 236–251.

Balotta, C., Bagnarelli, P., Violin, M., Ridolfo, A.L., Zhou, D., Berlusconi, A., Corvasce, S., Corbellino, M., Clementi, M., Clerici, M., et al. (1997). Deletion Ccr5 - Hiv Protection. Aids *78*, 1–5.

Baral, S., Beyrer, C., Muessig, K., Poteat, T., Wirtz, A.L., Decker, M.R., Sherman, S.G., and Kerrigan, D. (2012). Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect. Dis. *12*, 538–549.

Baron, S., Poast, J., and Cloyd, M.W. (1999). Why Is HIV Rarely Transmitted by Oral Secretions?: Saliva Can Disrupt Orally Shed, Infected Leukocytes. Arch. Intern. Med. *159*, 303–310.

Baumann, R.E., Rogers, A.A., Hamdan, H.B., Burger, H., Weiser, B., Gao, W., Anastos, K., Young, M., Meyer, W.A., Pesano, R.L., et al. (2015). Determination of HIV-1 coreceptor tropism using proviral DNA in women before and after viral suppression. AIDS Res. Ther. *12*, 1–6.

Beerenwinkel, N., Dä Umer, M., Oette, M., Korn, K., Hoffmann, D., Kaiser, R., Lengauer, T., Selbig, J., and Walter, H. (2003a). Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res. *31*, 3850–3855.

Beerenwinkel, N., Dä Umer, M., Oette, M., Korn, K., Hoffmann, D., Kaiser, R., Lengauer, T.,

Selbig, J., and Walter, H. (2003b). Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res. *31*, 3850–8355.

Bégaud, E., Barré-Sinoussi, F., Pancino, G., Feindirongai, G., Versmisse, P., Ipero, J., Léal, J., Germani, Y., Morvan, J., Fleury, H., et al. (2003). Broad spectrum of coreceptor usage and rapid disease progression in HIV-1-infected individuals from Central African Republic. AIDS Res. Humn Retroviruses *19*, 551–560.

Benko, Z., Liang, D., Agbottah, E., Hou, J., Chiu, K., Yu, M., Innis, S., Reed, P., Kabat, W., Elder, R.T., et al. (2004). Anti-Vpr Activity of a Yeast Chaperone Protein. J. Virol. *78*, 11016–11029.

Berger, E.A., Doms, R.W., Fenyö, E.M., Korber, B.T., Littman, D.R., Moore, J.P., Sattentau, Q.J., and Schuitemaker, H. (1998). A new classification for HIV-1. Nature *391*, 240.

Berger, E.A., Murphy, P.M., and Farber, J.M. (1999). Chemokine Receptors as HIV-1 Coreceptors: Roles in viral entry and disease. Annu. Rev. Immunol *17*, 657–700.

Berkowitz, R.D., Beckerman, K.P., McCune, J.M., and Schall, T.J. (1998). CXCR4 and CCR5 expression delineates targets for HIV-1 disruption of T cell differentiation. J. Immunol. *161*, 3702–3710.

Bernstein, H.B., Tucker, S.P., Kar, S.R., McPherson, S.A., McPherson, D.T., Dubay, J.W., Lebowitz, J., Compans, R.W., and Hunter, E. (1995). Oligomerization of the hydrophobic heptad repeat of gp41. J. Virol. *69*, 2745–2750.

Bhatti, A.B., Usman, M., and Kandi, V. (2016). Current Scenario of HIV/AIDS, Treatment Options, and Major Challenges with Compliance to Antiretroviral Therapy. Cureus *8*, e515.

Bieniasz, P.D., Grdina, T.A., Bogerd, H.P., and Cullen, B.R. (1999). Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. Proc. Natl. Acad. Sci. U. S. A. *96*, 7791–7796.

Binley, J.M., Sanders, R.W., Clas, B., Schuelke, N., Master, A., Guo, Y., Kajumo, F., Anselma, D.J., Maddon, P.J., Olson, W.C., et al. (2000). A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated

Structure. J. Virol. *74*, 627–643.

Bitty-Anderson, A.M., Gbeasor-Komlanvi, F.A., Tchankoni, M.K., Sadio, A., Salou, M., Coffie, P.A., Dagnra, C.A., and Ekouevi, D.K. (2022). HIV prevalence and risk behaviors among female sex workers in Togo in 2017: a cross-sectional national study. Arch. Public Heal. 2022 801 *80*, 1–8.

Björndal, A., Sönnerborg, A., Tscherning, C., Albert, J., and Fenyö, E.M. (1999). Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of ethiopian AIDS patients. AIDS Res. Hum. Retroviruses *15*, 647–653.

Blaak, H., Van't Wout, A.B., Brouwer, M., Hooibrink, B., Hovenkamp, E., and Schuitemaker, H. (2000). In vivo HIV-1 infection of CD45RA+CD4+ T cells is established primarily by syncytium-inducing variants and correlates with the rate of CD4+ T cell decline. Proc. Natl. Acad. Sci. U. S. A. *97*, 1269–1274.

Blanpain, C., Migeotte, I., Lee, B., Vakili, J., Doranz, B.J., Govaerts, C., Vassart, G., Doms, R.W., and Parmentier, M. (1999). CCR5 binds multiple CC-chemokines: MCP-3 acts as a natural antagonist. Blood *94*, 1899–1905.

Bosco, D.A., Eisenmesser, E.Z., Pochapsky, S., Sundquist, W.I., and Kern, D. (2002). Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. Proc. Natl. Acad. Sci. U. S. A. *99*, 5247–5252.

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

Bour, S., Geleziunas, R., and Wainberg, M.A. (1995). The Human Immunodeficiency Virus Type 1 (HIV-1) CD4 Receptor and Its Central Role in Promotion of HIV-1 Infection. Microbiol. Rev. *59*, 63–93.

Brenchley, J.M., and Douek, D.C. (2008). The mucosal barrier and immune activation in HIV pathogenesis. Curr. Opin. HIV AIDS *3*, 356–361.

Brenchley, J.M., Price, D.A., Schacker, T.W., Asher, T.E., Silvestri, G., Rao, S., Kazzaz, Z., Bornstein, E., Lambotte, O., Altmann, D., et al. (2006). Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat. Med. *12*, 1365–1371.

Brenner, C., and Kroemer, G. (2003). The mitochondriotoxic domain of Vpr determines HIV-1 virulence. J. Clin. Invest. *111*, 1455–1458.

Brenner, B.G., Roger, M., Routy, J.P., Moisi, D., Ntemgwa, M., Matte, C., Baril, J.G., Thomas, R., Rouleau, D., Bruneau, J., et al. (2007a). High rates of forward transmission events after acute/early HIV-1 infection. J. Infect. Dis. *195*, 951–959.

Brenner, B.G., Roger, M., Routy, J.-P., Moisi, D., Ntemgwa, M., Matte, C., Baril, J.-G., Jean Thomas, R., Rouleau, D., Bruneau, J., et al. (2007b). High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection. J. Infect. Dis. *195*.

Briggs, J.A.G., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G., and Kräusslich, H.G. (2009). Structure and assembly of immature HIV. Proc. Natl. Acad. Sci. U. S. A. *106*, 11090–11095.

Brinchmann, J.E., Gaudernack, G., and Vartdal, F. (1990). CD8+ T cells inhibit HIV replication in naturally infected CD4+ T cells. Evidence for a soluble inhibitor. J. Immunol. *144*.

Brown, T.R. (2015). I Am the Berlin Patient: A Personal Reflection. AIDS Res. Hum. Retroviruses *31*.

Brown, P. 0, Bowermant, B., Varmust, H.E., and Michael Bishop, J. (1989). Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein (transposition/recombination/retrotransposon/AIDS). Proc. Nati. Acad. Sci. USA *86*, 2525–2529.

Brumme, Z.L., Goodrich, J., Mayer, H.B., Brumme, C.J., Henrick, B.M., Wynhoven, B., Asselin, J.J., Cheung, P.K., Hogg, R.S., Montaner, J.S.G., et al. (2005). Molecular and Clinical Epidemiology of CXCR4-Using HIV-1 in a Large Population of Antiretroviral-Naive Individuals. J. Infect. Dis. 466–474.

Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc. Natl. Acad. Sci. U. S. A. *87*, 523–527.

Bukrinsky, M., and Adzhubei, A. (1999). Viral protein R of HIV-1. Rev. Med. Virol. *9*, 39–49.

Bukrinsky, M.I., Haggerty, S., Dempsey, M.P., Sharova, N., Adzhubei, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993). A nuclear localization signal within

HIV-1 matrix protein that governs infection of non-dividing cells. Nature *365*, 666–669.

Butler, S.L., Hansen, M.S.T., and Bushman, F.D. (2001). A quantitative assay for HIV DNA integration in vivo. Nat. Med. *7*, 631–634.

Cabral, G.B., Ferreira, J.L.D.P., Coelho, L.P.O., Fonsi, M., Estevam, D.L., Cavalcanti, J.S., and Brígido, L.F.D.M.E. (2012). Concordance of HIV Type 1 Tropism Phenotype to Predictions Using Web-Based Analysis of V3 Sequences: Composite Algorithms May Be Needed to Properly Assess Viral Tropism. AIDS Res. Hum. Retroviruses *28*, 734–738.

Callebaut, C., Krust, B., Jacotot, E., and Hovanessian, A.G. (1993). T cell activation antigen, CD26, as a cofactor for entry of HIV in CD4 + cells. Science (80-.). *262*, 2045–2050.

Callebaut, C., Jacotot, E., Blanco, J., Krust, B., and Hovanessian, A.G. (1997). The level of CD26 determines the rate of HIV entry in a CD4+ T-cell line. Adv. Exp. Med. Biol. *421*, 179–184.

Campbell, S., and Rein, A. (1999). In Vitro Assembly Properties of Human Immunodeficiency Virus Type 1 Gag Protein Lacking the p6 Domain. J. Virol. *73*, 2270–2279.

Campbell, E.M., Nunez, R., and Hope, T.J. (2004). Disruption of the Actin Cytoskeleton Can Complement the Ability of Nef To Enhance Human Immunodeficiency Virus Type 1 Infectivity. J. Virol. *78*, 5745–5755.

Campbell, S.M., Crowe, S.M., and Mak, J. (2001). Lipid rafts and HIV-1: from viral entry to assembly of progeny virions. J. Clin. Virol. *22*, 217–227.

Cancio, R., Spadari, S., and Maga, G. (2004). Vif is an auxiliary factor of the HIV-1 reverse transcriptase and facilitates abasic site bypass. Biochem. J *383*, 475–482.

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

Carr, J., Foley, B.T., Mokili, J.L.., Simmonds, P., and Bopopi, J.M. (2002). Identification of Novel Clade Of HIV-1 in DRC. AIDS Res. Hum. Retroviruses *18*, 817–823.

Casazza, J.P., Brenchley, J.M., Hill, B.J., and Ambrozak, A.R. (2009). Autocrine Production of

b-Chemokines Protects CMV-Specific CD4 + T Cells from HIV Infection. PLoS Pathog. *5*, e1000646.

Cashin, K., Gray, L.R., Jakobsen, M.R., Sterjovski, J., Churchill, M.J., Gorry, P.R., Gray, R., and Roche, M. (2013). CoRSeqV3-C: A novel HIV-1 subtype C specific V3 sequence based coreceptor usage prediction algorithm. Retrovirology *10*, 1–10.

Cashin, K., Gray, L.R., Harvey, K.L., Perez-Bercoff, D., Lee, G.Q., Sterjovski, J., Roche, M., Demarest, J.F., Drummond, F., Harrigan, P.R., et al. (2015). Reliable Genotypic Tropism Tests for the Major HIV-1 Subtypes. Sci. Rep. *5*.

CDC (1981). Morbidity and Mortality Weekly Report. CDC *50*, 430–471.

Chen, J., Pathak, V.K., Peng, W., and Hu, W.-S. (2008). Capsid Proteins from Human Immunodeficiency Virus Type 1 and Simian Immunodeficiency Virus SIV mac Can Coassemble into Mature Cores of Infectious Viruses . J. Virol. *82*, 8253–8261.

Chen, R., Le Rouzic, E., Kearney, J.A., Mansky, L.M., and Benichou, S. (2004). Vpr-mediated incorporation of UNG2 into HIV-1 particles is required to modulate the virus mutation rate and for replication in macrophages. J. Biol. Chem. *279*, 28419–28425.

Cheng-Mayer, C., Iannello, P., Shaw, K., Luciw, P.A., and Levy, J.A. (1989). Differential Effects of nef on HIV Replication: Implications for Viral Pathogenesis in the Host. Science (80-.). *226*, 1629–1632.

Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1992). Macrophage-Tropic Human Immunodeficiency Virus Isolates from Different Patients Exhibit Unusual V3 Envelope Sequence Homogeneity in Comparison with T-Cell-Tropic Isolates: Definition of Critical Amino Acids Involved in Cell Tropism. J. Virol. *66*, 6547–6554.

Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1996). Mapping of Independent V3 Envelope Determinants of Human Immunodeficiency Virus Type 1 Macrophage Tropism and Syncytium Formation in Lymphocytes. J. Virol. *70*, 9055–9059.

Chiappetta, S., Pogliaghi, M., Ripa, M., Lazzarin, A., Tambussi, G., and Nozza, S. (2014). HIV-1 Tropism Test Evaluation: Assessment and Clinical Implications. ISRN Virol.

Chowers, M.Y., Spina, C.A., Kwoh, T.J., Fitch, N.J.S., Richman, D.D., Guatellii, J.C., and Diego, S. (1994). Optimal Infectivity In Vitro of Human Immunodeficiency Virus Type 1 Requires an Intact nef Gene. J. Virol. *68*, 2906–2914.

Cilliers, T., Nhlapo, J., Coetzer, M., Orlovic, D., Ketas, T., Olson, W.C., Moore, J.P., Trkola, A., and Morris, L. (2003). The CCR5 and CXCR4 Coreceptors Are Both Used by Human Immunodeficiency Virus Type 1 Primary Isolates from Subtype C. J. Virol. *77*, 4449–4456.

Cimarelli, A., and Darlix, J.L. (2014). HIV-1 reverse transcription. Methods Mol. Biol. *1087*, 55–70.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat. Med. *11*, 1287–1289.

Clark, E., Nava, B., and Caputi, M. (2017). Tat is a multifunctional viral protein that modulates cellular gene expression and functions. Oncotarget *8*, 27569–27581.

Claudia, M., Elisa, D.C., Isabella, B., Carla, R.M., Ivano, M., Ombretta, T., Cecilia, G., and Guido, A. (2014). V3 Net Charge : Additional tool in HIV-1 Tropism Prediciton. AIDS Res. Hum. Retroviruses *30*.

Clavel, F., Guyader, M., Guétard, D., Sallé, M., Montagnier, L., and Alizon, M. (1986). Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nat. 1986 3246098 *324*, 691–695.

Clevestig, P., Pramanik, L., Leitner, T., and Ehrnst, A. (2006). CCR5 use by human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N -linked glycosylation site. J. Gen. Virol. *87*, 607–612.

Coakley, E., Reeves, J.D., Huang, W., Mangas-ruiz, M., Maurer, I., Harskamp, A.M., Gupta, S., Lie, Y., Petropoulos, C.J., and Schuitemaker, H. (2009). Comparison of Human Immunodeficiency Virus Type 1 Tropism Profiles in Clinical Samples by the Trofile and MT-2 Assays. Antimicrob. Agents Chemother. *53*, 4686–4693.

Cocchi, F., DeVico, A.L., Garzino-Demo, A., Arya, S.K., Gallo, R.C., and Lusso, P. (1995). Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors

produced by CD8+ T cells. Science *270*, 1811.

Cochrane, A.W., Perkins, A., and Rosen, C.A. (1990). Identification of Sequences Important in the Nucleolar Localization of Human Immunodeficiency Virus Rev: Relevance of Nucleolar Localization to Function. J. Virol. *64*, 881–885.

Coetzer, M., Nedellec, R., Salkowitz, J., Mclaughlin, ‡ Sherry, Liu, Y., Heath, L., Mullins, J.I., and Mosier, D.E. (2008). Evolution of CCR5 Use before and during Coreceptor Switching †. J. Virol. *82*, 11758–11766.

Coetzer, M., Nedellec, R., Cilliers, T., Meyers, T., Morris, L., and Mosier, D.E. (2011). Extreme Genetic Divergence Is Required for Coreceptor Switching in HIV-1 Subtype C. JAIDS J. Acquir. Immune Defic. Syndr. *56*, 9–15.

Coffin, J., Hughes, S., and Varmus, H. (1997). Retroviral Virions and Genomes. In Retroviruses, (Cold Spring Harbor Laboratory Press), p.

Cohen, M.S., Shaw, G.M., Mcmichael, A.J., Ch, B., and Haynes, B.F. (2011a). Acute HIV-1 Infection. J. Med. *364*, 1943–1964.

Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H.S., et al. (2011b). Prevention of HIV-1 Infection with Early Antiretroviral Therapy. N. Engl. J. Med. *365*, 493.

Collier, A.C., Coombs, R.W., Katzenstein, D., Holodniy, M., Gibson, J., Mordenti, J., Izu, A.E., Duliege, A.M., Ammann, A.J., Merigan, T., et al. (1995). Safety, pharmacokinetics, and antiviral response of CD4-immunoglobulin G by intravenous bolus in AIDS and AIDS-related complex. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. *10*, 150–156.

Connell, B.J., Michler, K., Capovilla, A., Venter, W.D.F., Stevens, W.S., and Papathanasopoulos, M.A. (2008). Emergence of X4 usage among HIV-1 subtype C: Evidence for an evolving epidemic in South Africa. AIDS *22*, 896–899.

Connor, R.I., Sheridan, K.E., Ceradini, D., Choe, S., and Landau, N.R. (1997). Change in Coreceptor Use Correlates with Disease Progression in HIV-1-Infected Individuals. J. Exp. Med. *185*, 621–628.

Conti, L., Rainaldi, G., Matarrese, P., Varano, B., Rivabene, R., Columba, S., Sato, A., Belardelli, F., Malorni, W., and Gessani, S. (1998). The HIV-1 vpr Protein Acts as a Negative Regulator of Apoptosis in a Human Lymphoblastoid T Cell Line: Possible Implications for the Pathogenesis of AIDS. J. Exp. Med. *187*, 403–413.

Conticello, S.G., Harris, R.S., and Neuberger, M.S. (2003). The Vif Protein of HIV Triggers Degradation of the Human Antiretroviral DNA Deaminase APOBEC3G. Curr. Biol. *13*, 2009–2013.

Cormier, E.G., and Dragic, T. (2002). The Crown and Stem of the V3 Loop Play Distinct Roles in Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Interactions with the CCR5 Coreceptor. J. Virol. *76*, 8953–8957.

Craigie, R., and Bushman, F.D. (2012). HIV DNA Integration. Cold Spring Harb. Perspect. Med. *2*, 1–18.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika *16*, 297–334.

Cullen, B.R. (2001). Journey to the Center of the Cell. *105*, 697–700.

D'Souza, V., and Summers, M.F. (2005). How retroviruses select their genomes. Nat. Rev. Microbiol. *3*, 643–655.

Dalgeish, A.G., Beverley, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Lett. to Nat. *312*, 763–766.

Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J., and Kjems, J. (2004). RNA Interactions in the 5′ Region of the HIV-1 Genome. J. Mol. Biol. *336*, 369–379.

Datta, S.A.K., Temeselew, L.G., Crist, R.M., Soheilian, F., Kamata, A., Mirro, J., Harvin, D., Nagashima, K., Cachau, R.E., and Rein, A. (2011). On the Role of the SP1 Domain in HIV-1 Particle Assembly: a Molecular Switch? J. Virol. *85*, 4111–4121.

Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Michael, W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E., et al. (1996). Genetic Restriction of HIV-1

Infection and Progression to AIDS by a Deletion Allele of the CKR5 Structural Gene. Science (80-.). *273*, 1856–1862.

Désiré, N., Cerutti, L., Le Hingrat, Q., Perrier, M., Emler, S., Calvez, V., Descamps, D., Marcelin, A.G., Hué, S., and Visseaux, B. (2018). Characterization update of HIV-1 M subtypes diversity and proposal for subtypes A and D sub-subtypes reclassification. Retrovirology *15*, 1–7.

Dietrich, L., Ehrlich, L.S., Lagrassa, T.J., Ebbets-reed, D., and Carter, C. (2001). Structural Consequences of Cyclophilin A Binding on Maturational Refolding in Human Immunodeficiency Virus Type 1 Capsid Protein. J. Virol. *75*, 4721–4733.

Digban, T.O., Iweriebor, B.C., Nwodo, U.U., Okoh, A.I., and Obi, L.C. (2020). Chemokine Coreceptor Usage Among HIV-1 Drug-Naive Patients Residing in the Rural Eastern Cape, South Africa. AIDS Res. Hum. Retroviruses *36*, 688–696.

DImeglio, C., Raymond, S., Jeanne, N., Reynes, C., Carcenac, R., Lefebvre, C., Cazabat, M., Nicot, F., Delobel, P., and Izopet, J. (2020). THETA: a new genotypic approach for predicting HIV-1 CRF02-AG coreceptor usage. Bioinformatics *36*, 416–421.

Ding, J., Tasker, C., Lespinasse, P., Dai, J., Fitzgerald-Bocarsly, P., Lu, W., Heller, D., and Chang, T.L.-Y. (2015). Integrin α4β7 Expression Increases HIV Susceptibility in Activated Cervical CD4+ T Cells by an HIV Attachment-Independent Mechanism. JAIDS J. Acquir. Immune Defic. Syndr. *69*, 509–518.

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., Napier, C., et al. (2005). Maraviroc (UK-427,857), a Potent, Orally Bioavailable, and Selective Small-Molecule Inhibitor of Chemokine Receptor CCR5 with Broad-Spectrum Anti-Human Immunodeficiency Virus Type 1 Activity. Antimicrob. Agents Chemother. *49*, 4721–4732.

Douek, D.C., Picker, L.J., and Koup, R.A. (2003). T cell dynamics in HIV-1 infection. Annu. Rev. Immunol. *21*, 265–304.

Dragic, T. (2001). An overview of the determinants of CCR5 and CXCR4 co-receptor function. J. Gen. Virol. *82*, 1807–1814.

Dragic, T., Trkola, A., D Thompson, D.A., Cormier, E.G., Kajumo, F.A., Maxwell, E., Lin, S.W., Ying, W., Smith, S.O., Sakmar, T.P., et al. (2000). A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. PNAS *97*, 5639–5644.

Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, et al. (1996). HIV-1 entry into CD4+ cells is mediated bythe chemokine receptor CC-CKR-5. Nature *381*, 667–673.

Dubois, N., Khoo, K.K., Ghossein, S., Seissler, T., Wolff, P., Mckinstry, W.J., Mak, J., Paillart, J.-C., Marquet, R., and Bernacchi, S. (2018). The C-terminal p6 domain of the HIV-1 Pr55 Gag precursor is required for specific binding to the genomic RNA. RNA Biol. *15*, 923–936.

Dunn, D.T., Newell, M.L., Ades, A.E., and Peckham, C.S. (1992). Risk of human immunodeficiency virus type 1 transmission through breastfeeding. Lancet *340*, 585–588.

Dupont, S., Sharova, N., DeHoratius, C., and Virbasius, C.-M.A. (1999). A novel nuclear export activity in HIV-1 matrix protein required for viral replication. Nature *402*, 681–605.

Eberle, J., and Gürtler, L. (2012). HIV types, groups, subtypes and recombinant forms: Errors in replication, selection pressure and quasispecies. Intervirology *55*, 79–83.

Esbjörnsson, J., Månsson, F., Martínez-Arias, W., Vincic, E., Biague, A.J., da Silva, Z.J., Fenyö, E.M., Norrgren, H., and Medstrand, P. (2010). Frequent CXCR4 tropism of HIV-1 subtype A and CRF02_AG during late-stage disease - indication of an evolving epidemic in West Africa. Retrovirology *7*, 1–13.

Estes, J.D., Haase, A.T., and Schacker, T.W. (2008). The Role of Collagen Deposition in Depleting CD4+ T Cells and Limiting Reconstitution in HIV-1 and SIV Infections through Damage to the Secondary Lymphoid Organ Niche. Immunology *20*, 181–186.

Fabeni, L., Berno, G., Fokam, J., Bertoli, A., Alteri, C., Gori, C., Forbici, F., Takou, D., Vergori, A., Zaccarelli, M., et al. (2017). Comparative Evaluation of Subtyping Tools for Surveillance of Newly Emerging HIV-1 Strains. J. Clin. Invest. *55*, 2827–2837.

Fauci, A.S., Pantaleo, G., Stanley, S., and Weissman, D. (1996). Immunopathogenic mechanisms of HIV infection. In Annals of Internal Medicine, (American College of Physicians), pp. 654–663.

Feinberg, M.B., Baltimore, D., and Frankel, A.D. (1991). The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. Proc. Natl. Acad. Sci. U. S. A. *88*, 4045–4049.

Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T., and Pavlakis, G.N. (1989). rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. U. S. A. *86*, 1495–1499.

Feng, Y., Broder, C., Kennedy, P., and Berger, E. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science (80-.). *272*, 872–877.

Fenyö, E.M., Esbjö Rnsson, J., Medstrand, P., and Jansson, & M. (2011). Human immunodeficiency virus type 1 biological variation and coreceptor use: from concept to clinical significance. J. Intern. Med. *270*, 520–531.

Fischer, U., Huber, J., Boelens, W.C., Mattajt, L.W., and Lührmann, R. (1995). The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475–483.

Ford, E.S., Puronen, C.E., and Sereti, I. (2009). Immunopathogenesis of asymptomatic chronic HIV Infection: The calm before the storm. Curr. Opin. HIV AIDS *4*, 206–214.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051–1060.

Fortin, J.-F.O., Cantin, J., and Tremblay, M.J. (1998). T Cells Expressing Activated LFA-1 Are More Susceptible to Infection with Human Immunodeficiency Virus Type 1 Particles Bearing Host-Encoded ICAM-1. J. Virol. *72*, 2105–2112.

Fouchier, R.A.M., and Malim, M.H. (1999). Nuclear import of human immunodeficiency virus type-1 preintegration complexes. Adv Virus Res *52*, 275–299.

Fouchier, R.A., Groenink, M., Kootstra, N.A., Tersmette, M., Huisman, H.G., Miedema, F., and Schuitemaker, H. (1992). Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. *66*, 3183–3187.

Fouchier, R.A.M., Brouwer, M., Broersen, S.M., and Schuitemaker, H. (1995). Simple determination of human immunodeficiency virus type 1 syncytium- inducing V3 genotype by PCR. J. Clin. Microbiol. *33*, 906–911.

Frank, T.D., Carter, A., Jahagirdar, D., Biehl, M.H., Douwes-Schultz, D., Larson, S.L., Arora, M., Dwyer-Lindgren, L., Steuben, K.M., Abbastabar, H., et al. (2019). Global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2017, and forecasts to 2030, for 195 countries and territories: A systematic analysis for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. Lancet HIV *6*, e831–e859.

Freed, E.O. (2015). HIV-1 assembly, release and maturation. Nat. Rev. Microbiol. *13*, 484–496.

Freed, E.O., and Mouland, A.J. (2006). The cell biology of HIV-1 and other retroviruses. Retrovirology *3*, 1–10.

Freed, E. 0, Orenstein, J.M., Buckler-White, A.J., and Martin', M.A. (1994). Single Amino Acid Changes in the Human Immunodeficiency Virus Type 1 Matrix Protein Block Virus Particle Production. J. Virol. *68*, 5311–5320.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Matija Peterlin, B. (2004). Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Mol. Cell. Biol. *24*, 787–795.

Fujiwara, T., and Mizuuchi, K. (1988). Retroviral DNA integration: Structure of an integration intermediate. Cell *54*, 497–504.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. (1997). CRM1is responsible for intracellular transport mediatedby thenuclear export signal Makoto. Lett. to Nat. *390*, 308–311.

Gabuzda, D. (2000). Role of Vif in Stability of the Human Immunodeficiency Virus Type 1 Core. J. Virol. *74*, 11055–11066.

Garcia, V., and Miller, D.A. (1991). Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. Nature *350*, 508–511.

Garrido, C., Roulet, V., Chueca, N., Poveda, E., Aguilera, A., Skrabal, K., Zahonero, N., Carlos, S., García, F., Faudon, J.L., et al. (2008). Evaluation of eight different bioinformatics tools to predict viral tropism in different human immunodeficiency virus type 1 subtypes. J. Clin. Microbiol. *46*, 887–891.

Garrus, J.E., Von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Côté, M., Rich, R.L., et al. (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell *107*, 55–65.

Geijtenbeek, T.B.H., Kwon, D.S., Torensma, R., Van Vliet, S.J., Van Duijnhoven, G.C.F., Middel, J., Cornelissen, I.L.M.H.A., Nottet, H.S.L.M., KewalRamani, V.N., Littman, D.R., et al. (2000). DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell *100*, 587–597.

Ghosn, J., Bayan, T., Meixenberger, K., Tran, L., Frange, P., Monforte, A., Zangerle, R., de Mendoza, C., Krastinova, E., Porter, K., et al. (2017). CD4 T cell decline following HIV seroconversion in individuals with and without CXCR4-tropic virus on behalf of the CASCADE Collaboration in. J. Antimicrob. Chemother. *72*, 2862–2868.

Giguère, K., Béhanzin, L., Guédou, F.A., Leblond, F.A., Goma-Matsétsé, E., Zannou, D.M., Affolabi, D., Kêkê, R.K., Gangbo, F., Bachabi, M., et al. (2019). Biological validation of self-reported unprotected sex and comparison of underreporting over two different recall periods among female sex workers in Benin. Open Forum Infect. Dis. *6*.

Gilliam, B.L., Riedel, D.J., and Redfield, R.R. (2010). Clinical use of CCR5 inhibitors in HIV and beyond. J. Transl. Med. *9*, 1–14.

Gitahi-Kamau, N.T., Kiarie, J.N., Mutai, K.K., Gatumia, B.W., Gatongi, P.M., and Lakati, A. (2015). Socio-economic determinants of disease progression among HIV infected adults in Kenya. BMC Public Health *15*, 1–6.

Giuliani, M., Santoro, M.M., Lo Presti, A., Cella, E., Scognamiglio, P., Lai, A., Latini, A., Fabeni, L., Gori, C., Pinnetti, C., et al. (2013). Circulation of HIV-1 CRF02-AG among MSM population in Central Italy: A molecular epidemiology-based study. Biomed Res. Int. *2013*.

Gorry, P. r, and Ancuta, P. (2011). Coreceptors and HIV-1 Pathogenesis. Curr HIV/AIDS 45–

532.

Gounder, K., Oyaro, M., Padayachi, N., Zulu, T.M., de Oliveira, T., Wylie, J., and Ndung'u, T. (2017). Complex Subtype Diversity of HIV-1 Among Drug Users in Major Kenyan Cities. AIDS Res. Hum. Retroviruses *33*, 500–510.

Groen, G.V.A.N.D.E.R., Sharp, P.M., and Saman, E. (2000). SHORT COMMUNICATION Sequence Analysis of a Highly Divergent HIV-1-Related Lentivirus Isolated from a Wild Captured Chimpanzee. *350*, 346–350.

Groot, F., van Capel, T.M.M., Schuitemaker, J.H.N., Berkhout, B., and de Jong, E.C. (2006). Differential susceptibility of naïve, central memory and effector memory T cells to dendritic cell-mediated HIV-1 transmission. Retrovirology *3*, 1–10.

Grossman, Z., Meier-Schellersheim, M., Paul, W.E., and Picker, L.J. (2006). Pathogenesis of HIV infection: What the virus spares is as important as what it destroys. Nat. Med. *12*, 289–295.

Guedon, J.T., Luo, K., Zhang, H., and Markham, R.B. (2015). Monoclonal and Single Domain Antibodies Targeting β-Integrin Subunits Block Sexual Transmission of HIV-1 in In Vitro and In Vivo Model Systems. JAIDS J. Acquir. Immune Defic. Syndr. *69*, 278–285.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803–809.

Hartley, O., Klasse, P.J., Sattentau, Q.J., and Moore, J.P. (2005). V3: HIV's switch-hitter. AIDS Res. Hum. Retroviruses *21*, 171–189.

Hebb, 1, Singer, D.O., and Gray, W. (1999). Origin of HIV-1 in the chimpanzee Pan. Annu. Rev. Neurosci *18*, 555–586.

Hemelaar, J., Gouws, E., Ghys, P.D., and Osmanov, S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS *25*, 679–689.

Hemelaar, J., Elangovan, R., Yun, J., Dickson-Tetteh, L., Fleminger, I., Kirtley, S., Williams, B., Gouws-Williams, E., Ghys, P.D., Abimiku, A.G., et al. (2019). Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis.

Lancet Infect. Dis. *19*, 143–155.

Henriet, S., Richera, D., Bernacchia, S., Decroly, E., Vigne, R., and Et. al. (2005). Cooperative and Specific Binding of Vif to the 5′ Region of HIV-1 Genomic RNA. J. Mol. Biol. *354*, 55–72.

Herrmann, C.H., and Rice, A.P. (1995). Lentivirus Tat Proteins Specifically Associate with a Cellular Protein Kinase, TAK, That Hyperphosphorylates the Carboxyl-Terminal Domain of the Large Subunit of RNA Polymerase II: Candidate for a Tat Cofactor. J. Virol. *69*, 1612–1620.

Hioe, C.E., Tuen, M., Vasiliver-Shamis, G., Alvarez, Y., Prins, K.C., Banerjee, S., Ná Das, A., Cho, M.W., Dustin, M.L., and Kachlany, S.C. (2011). HIV Envelope gp120 Activates LFA-1 on CD4 T-Lymphocytes and Increases Cell Susceptibility to LFA-1-Targeting Leukotoxin (LtxA). PLoS One *6*, 23202–23211.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., and Johnson, P.R. (1989). An African primate lentivirus (SIVsmclosely related to HIV-2. Nat. 1989 3396223 *339*, 389–392.

Hladik, F., and McElrath, M.J. (2008). Setting the stage: host invasion by HIV. Nat. Rev. Immunol. *8*, 447–457.

Hoffman, C. (2007). The epidemiology of HIV coreceptor tropism. Eur. J. Med. Res. *12*, 385–390.

Höglund, S., Ohagen, A., Lawrence, K., and Gabuzda, D. (1994). Role of vif during packing of the core of HIV-1. Virology *201*, 351–355.

Hongjaisee, S., Nantasenamat, C., Carraway, T., and Shoombuatong, W. (2019). HIVCoR: A sequence-based tool for predicting HIV-1 CRF01_AE coreceptor usage. Comput. Biol. Chem. *80*, 419–432.

Hope, T.J., McDonald, D., Huang, X.J., Low, J., and Parslow, T.G. (1990). Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. J. Virol. *64*, 5360–5366.

Hope, T.J., Klein,’, N.P., Elder, M.E., and Parslow’, T.G. (1992). trans-Dominant Inhibition of Human Immunodeficiency Virus Type 1 Rev Occurs through Formation of Inactive Protein Complexes. J. Virol. *66*, 1849–1855.

Hsu, K., Seharaseyon, J., Dong, P., Bour, S., and Marbá N, E. (2004). Mutual Functional Destruction of HIV-1 Vpu and Host TASK-1 Channel. Mol. Cell *14*, 259–267.

Hu, W.-S., and Hughes, S.H. (2012). HIV-1 Reverse Transcription. Perspect. Med. *2*, a006882.

Hu, W.S., and Temin, H.M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. U. S. A. *87*, 1556–1560.

Huang, C.-C., Tang, M., Zhang, M.-Y., Majeed, S., Montabana, E., Stanfield, R.L., Dimitrov, D.S., Korber, B., Sodroski, J., Wilson, I.A., et al. (2005). Structure of a V3-Containing HIV-1 gp120 Core. Science (80-.). *310*, 1025–1028.

Hurley, J.H. (2008). ESCRT complexes and the biogenesis of multivesicular bodies. Curr. Opin. Cell Biol. *20*, 4–11.

Hütter, G., Nowak, D., Mossner, M., Ganepola, S., Müßig, A., Allers, K., Schneider, T., Hofmann, J., Kücherer, C., Blau, O., et al. (2009). Long-Term Control of HIV by CCR5 Delta32/ Delta32 Stem-Cell Transplantation. N Engl J Med *360*, 692–700.

Hwang, S.S., Cullen, B.R., Boyle, T.J., and Lyerly, K.H. (1991). Identification of the Envelope V3 Loop as the Primary Determinant of Cell Tropism. Science (80-.). *253*, 71–74.

Iordanskiy, S., Zhao, Y., Dubrovsky, L., Iordanskaya, T., Chen, M., Liang, D., and Bukrinsky, M. (2004a). Heat Shock Protein 70 Protects Cells from Cell Cycle Arrest and Apoptosis Induced by Human Immunodeficiency Virus Type 1 Viral Protein R. J. Virol. *78*, 9697–9704.

Iordanskiy, S., Zhao, Y., Dimarzio, P., Agostini, I., Dubrovsky, L., and Bukrinsky, M. (2004b). Heat-shock protein 70 exerts opposing effects on Vpr-dependent and Vpr-independent HIV-1 replication in macrophages. Blood *104*, 1867–1872.

Isaacman-Beck, J., Hermann, E.A., Yi, Y., Ratcliffe, S.J., Mulenga, J., Allen, S., Hunter, E., Derdeyn, C.A., and Collman, R.G. (2009). Heterosexual Transmission of Human Immunodeficiency Virus Type 1 Subtype C: Macrophage Tropism, Alternative Coreceptor Use, and the Molecular Anatomy of CCR5 Utilization. J. Virol. *83*, 8208–8220.

Izaurralde, E., Kutay, U., Von Kobbe, C., Mattaj, I.W., and Gö Rlich, D. (1997). The asymmetric

distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. EMBO J. *16*, 6535–6547.

Jacotot, E., Ravagnan, L., Loeffler, M., Ferri, K.F., Vieira, H.L.A., Zamzami, N., Costantini, P., Druillennec, S., Hoebeke, J., Briand, J.P., et al. (2000). The HIV-1 Viral Protein R Induces Apoptosis via a Direct Effect on the Mitochondrial Permeability Transition Pore. J. Exp. Med *191*, 33–45.

Jacotot, E., Ferri, K.F., Hamel, C. El, Brenner, C., Druillennec, S., Hoebeke, J., Rustin, P., Métivier, D., Lenoir, C., Geuskens, M., et al. (2001). Control of Mitochondrial Membrane Permeabilization by Adenine Nucleotide Translocator Interacting with HIV-1 Viral Protein R and Bcl-2. J. Exp. Med *193*, 509–519.

Jakobsen, M.R., Ellett, A., Churchill, M.J., and Gorry, P.R. (2010). Viral tropism, fitness and pathogenicity of HIV-1 subtype C. Future Virol. *5*, 219–231.

Janssens, W., Fransen, K., Peeters, M., Heyndrickx, L., Motte, J., Bedjabaga, L., Delaporte, E., Piot, P., and Van Der Groen, G. (1994). Phylogenetic Analysis of a New Chimpanzee Lentivirus SIVcpz-gab2 from a Wild-Captured Chimpanzee from Gabon. AIDS Res. Hum. Retroviruses *10*, 1191–1192.

Japour, A.J., Fiscus, S.A., Arduino, J., Mayers, D.L., Reichelderfer, P.S., and Kuritzkest, D.R. (1994). Standardized Microtiter Assay for Determination of Syncytium- Inducing Phenotypes of Clinical Human Immunodeficiency Virus Type 1 Isolates. J. Clin. Microbiol. *32*, 2291–2294.

Javaherian, K., Langloist, A.J., Mcdanalt, C., Ross, K.L., Eckler, L.I., Jellis, C.L., Profy, A.T., Rusche, J.R., Bolognesit, D.P., Putney, S.D., et al. (1989). Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. U. S. A. *86*, 6768–6772.

Jeang, K. (1996). HIV-1 Tat : Structure and Function. Hum. Retroviruses AIDS 1996 Compend. 3–18.

Jeanne, N., Saliou, A., Carcenac, R., Lefebvre, C., Dubois, M., Cazabat, M., Nicot, F., Loiseau, C., Raymond, S., Izopet, J., et al. (2015). Position-specific automated processing of V3 env ultra-deep pyrosequencing data for predicting HIV-1 tropism. Sci. Rep. *5*, 16944.

Jensen, M.A., Li, F.-S., van 't Wout, A.B., Nickle, D.C., Shriner, D., He, H.-X., McLaughlin, S., Shankarappa, R., Margolick, J.B., and Mullins, J.I. (2003). Improved Coreceptor Usage Prediction and Genotypic Monitoring of R5-to-X4 Transition by Motif Analysis of Human Immunodeficiency Virus Type 1 env V3 Loop Sequences. J. Virol. *77*, 13376–13388.

Jensen, M.A., Coetzer, M., Angélique, ‡, Van 't Wout, B., Morris, § Lynn, and Mullins, J.I. (2006). A Reliable Phenotype Predictor for Human Immunodeficiency Virus Type 1 Subtype C Based on Envelope V3 Sequences †. J. Virol. *80*, 4698–4704.

Jensen A. Mark, Wout Angelique (2003). Predicting HIV-1 Coreceptor Usage with Sequence Analysis. AIDS Rev. *5*, 104–112.

Ji, C., Brandt, M., Dioszegi, M., Jekle, A., Schwoerer, S., Challand, S., Zhang, J., Chen, Y., Zautke, L., Achhammer, G., et al. (2007). Novel CCR5 monoclonal antibodies with potent and broad-spectrum anti-HIV activities. Antiviral Res. *74*, 125–137.

Jin, Q., Marsh, J., Cornetta, K., and Alkhatib, G. (2008). Resistance to human immunodeficiency virus type 1 (HIV-1) generated by lentivirus vector-mediated delivery of the CCR5Δ32 gene despite detectable expression of the HIV-1 co-receptors. J. Gen. Virol. *89*, 2611–2621.

De Jong, J.J., De Ronde, A., Keulen, W., Tersmette, M., and Goudsmit, J. (1992). Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J. Virol. *66*, 6777–6780.

Joram, S.L., Paul, G., Moses, K., Stanley, B., Isaac, M., Allan, G., Tom, M., Lilian, K., and Mildred, M. (2017). Misdiagnosis of HIV treatment failure based on clinical and immunological criteria in Eastern and Central Kenya. BMC Infect. Dis. *17*, 1–7.

Judicate, G.P., Barabona, G., Kamori, D., Mahiti, M., Tan, T.S., Ozono, S., Mgunya, A.S., Kuwata, T., Matsushita, S., Sunguya, B., et al. (2021). Phenotypic and Genotypic Co-receptor Tropism Testing in HIV-1 Epidemic Region of Tanzania Where Multiple Non-B Subtypes Co-circulate. Front. Microbiol. *12*, 1788.

Kaleebu, P., Nankya, I.L., Yirrell, D.L., Shafer, L.A., Kyosiimire-Lugemwa, J., Lule, D.B., Morgan, D., Beddows, S., Weber, J., and Whitworth, J.A.G. (2007). Relation Between

Chemokine Receptor Use, Disease Stage, and HIV-1 Subtypes A and D. JAIDS J. Acquir. Immune Defic. Syndr. *45*, 28–33.

Kameoka, J., Tanaka, T., Nojima, Y., Schlossman, S.F., and Morimoto, C. (1993). Direct association of adenosine deaminase with a T cell activation antigen, CD26. Science (80-.). *261*, 466–469.

Kanki, P.J., Travers, K.U., Marlink, R.G., Essex, M.E., MBoup, S., Gueye-NDiaye, A., Siby, T., Thior, I., Sankalé, J.L., Hsieh, C.C., et al. (1994). Slower heterosexual spread of HIV-2 than HIV-1. Lancet *343*, 943–946.

Karlsson, U., Antonsson, L., Ljungberg, B., Medstrand, P., Esbjörnsson, J., Jansson, M., and Gisslen, M. (2012). Dual R3R5 tropism characterizes cerebrospinal fluid HIV-1 isolates from individuals with high cerebrospinal fluid viral load. Aids *26*, 1739–1744.

Karn, J., Stoltzfus, C.M., Bushman, F.D., Nabel, G.J., and Swanstrom, R. (2012). Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. Perspect. Med. *4*.

Kato, K., Sato, H., and Takebe, Y. (1999). Role of Naturally Occurring Basic Amino Acid Substitutions in the Human Immunodeficiency Virus Type 1 Subtype E Envelope V3 Loop on Viral Coreceptor Usage and Cell Tropism. J. Virol. *73*, 5520–5526.

Kawamura, T., Gulden, F.O., Sugaya, M., McNamara, D.T., Borris, D.L., Lederman, M.M., Orenstein, J.M., Zimmerman, P.A., and Blauvelt, A. (2003). R5 HIV productively infects langerhans cells, and infection levels are regulated by compound CCR5 polymorphisms. Proc. Natl. Acad. Sci. U. S. A. *100*, 8401–8406.

Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., et al. (2015). Structure of the HIV-1 RNA Packaging Signal. Science *348*, 917.

Kilzer, J.M., Stracker, T., Beitzel, B., Meek, K., Weitzman, M., and Bushman, F.D. (2003). Roles of host cell factors in circularization of retroviral DNA. Virology *314*, 460–467.

Kipreos, E.T., and Pagano, M. (2000). The F-box protein family. Genome Biol. *1*, 1–7.

Kitawi, R.C., Hunja, C.W., Aman, R., Ogutu, B.R., Muigai, A.W.T., Kokwaro, G.O., and

Ochieng, W. (2017). Partial HIV C2V3 envelope sequence analysis reveals association of coreceptor tropism, envelope glycosylation and viral genotypic variability among Kenyan patients on HAART. Virol. J. *14*.

Kjems, J., Calnan, B.J., Frankel, A.D., and Sharp, P.A. (1992). Specific binding of a basic peptide from HIV-1 Rev. EMBO J. *11*, 1119–1129.

Klarmann, G.J., Yu, H., Chen, X., Dougherty, J.P., and Preston, B.D. (1997). Discontinuous plus-strand DNA synthesis in human immunodeficiency virus type 1-infected cells and in a partially reconstituted cell-free system. J. Virol. *71*, 9259–9269.

Klatzmann, D., Barre-Sinoussi, F., Nugeyre, M., Danquet, C., Vilme, E., Griscelli, C., Brun-Veziret, and Rouzioux, C. (1984). Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. Science (80-.). 59–63.

Koh, Y., Wu, X., Ferris, A.L., Matreyek, K.A., Smith, S.J., Lee, K., Kewalramani, V.N., Hughes, S.H., and Engelman, A. (2013). Differential Effects of Human Immunodeficiency Virus Type 1 Capsid and Cellular Factors Nucleoporin 153 and LEDGF/p75 on the Efficiency and Specificity of Viral DNA Integration. J. Virol. *87*, 648–658.

Kondo, N., and Melikyan, G.B. (2012). Intercellular Adhesion Molecule 1 Promotes HIV-1 Attachment but Not Fusion to Target Cells. PLoS One *7*, 44827–44835.

Konvalinka, J., Kräusslich, H.G., and Müller, B. (2015). Retroviral proteases and their roles in virion maturation. Virology *479–480*, 403–417.

Koot, M., Vos, A.H.V., Keet, R.P.M., Goede, D.E.Y., and Dercksen, M. Wouter Terpstra, Fokke G. Coutinho, Roel A. Miedema, Frank Tersmette, M. (1992). HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. Aids *6*, 49–54.

Korber, A.B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Bhattacharya, T., Lapedes, A., Hahn, B.H., et al. (2000). Timing the Ancestor of the HIV-1 Pandemic Strains Linked references are available on JSTOR for this article : Timing the Ancestor of the HIV-1 Pandemic Strains. Science (80-.). *288*, 1789–1796.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. Br. Med. Bull.

58, 19–42.

Kouyos, R.D., Von Wyl, V., Yerly, S., Rg Bö, J., Taffé, P., Shah, C., Bü, P., Klimkait, T., Weber, R., Hirschel, B., et al. (2010a). Molecular Epidemiology Reveals Long-Term Changes in HIV Type 1 Subtype B Transmission in Switzerland. J. Infect. Dis. *201*, 1488–1497.

Kouyos, R.D., Von Wyl, V., Yerly, S., Böni, J., Taffé, P., Shah, C., Bürgisser, P., Klimkait, T., Weber, R., Hirschel, B., et al. (2010b). Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J. Infect. Dis. *201*, 1488–1497.

Kumi, S., Rutstei, M., Powers, S.E., Fidler, K.A., Miller, W.C., Eron Jr, J.J., and Cohen, M.S. (2013). The Detection and Management of Early HIV Infection: A Clinical and Public Health Emergency. J Acquir Immune Defic Syndr *63*, 187–199.

van der Kuyl, A.C., and Berkhout, B. (2012). The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. Retrovirology *9*, 1–14.

Langford, S.E., Ananworanich, J., and Cooper, D.A. (2007). Predictors of disease progression in HIV infection: A review. AIDS Res. Ther. *4*, 1–14.

Lapadat-tapolsky, M., De Rocquigny, H., Van Gent, D., Roques, B., Plasterk, R., and Darlix, J.L. (1993). Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. Nucleic Acids Res. *21*, 2024.

LaRosa, G., Davide, J., Weinhold, K., and Et. al. (1991). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science (80-.). *251*, 932–935.

Lau, K.A., and Wong, J.J.L. (2013). Current trends of HIV recombination worldwide. Infect. Dis. Rep. *5*, 16–20.

Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A.J. (2003a). Hypermutation of HIV-1 DNA in the Absence of the Vif Protein. Science (80-.). *300*, 1112–1115.

Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A.J. (2003b). Hypermutation of HIV-1 DNA in the Absence of the Vif Protein. Science (80-.). *300*, 1122.

Leda, A.R., Hunter, J., Castro de Oliveira, U., Junqueira de Azevedo, I., Kallas, E.G., Araripe Sucupira, M.C., and Diaz, R.S. (2020). HIV-1 genetic diversity and divergence and its

correlation with disease progression among antiretroviral naïve recently infected individuals. Virology *541*, 13–24.

Lee, K.E., Ambrose, Z., Martin, T.D., Oztop, I., Mulky, A., Julias, J.G., Vandegraaff, N., Baumann, J.G., Wang, R., Yuen, W., et al. (2010). Flexible Use of Nuclear Import Pathways by HIV-1. Cell Host Microbe *7*, 221–233.

Leitner, T., Alaeus, A., Marquina, S., Lilja, E., LIDMAN, K., and ALBERT, J. (1995). Yet Another Subtype of HIV Type 1? AIDS Res. Hum. Retroviruses *11*, 995–997.

Lever, A., Gottlinger, H., Haseltine, W., and Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J. Virol. *63*, 4085–4087.

Levy, J.A. (2009). HIV pathogenesis: 25 years of progress and persistent challenges. Aids *23*, 147–160.

Li, W.-H., Tanimura, M., and Sharp2, P.M. (1998). Rates and Dates of Divergence between AIDS Virus Nucleotide Sequences'. Mol. Biol. Evol. *5*, 313–330.

Li, X., Liu, H., Liu, L., Feng, Y., Kalish, M.L., Ho, S.Y.W., and Shao, Y. (2017). Tracing the epidemic history of HIV-1 CRF01_AE clusters using near-complete genome sequences. Sci. Reports 2017 71 *7*, 1–11.

Liang, B., Luo, M., Scott-Herridge, J., Semeniuk, C., Mendoza, M., Capina, R., Sheardown, B., Ji, H., Kimani, J., Ball, B.T., et al. (2011). A Comparison of Parallel Pyrosequencing and Sanger Clone-Based Sequencing and Its Impact on the Characterization of the Genetic Diversity of HIV-1. PLoS One *6*.

Lightfoote,', M.M., Coligan, J.E., Folks,', T.M., Fauci, A.S., Martin, M.A., and Venkatesan3, S. (1986). Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. J. Virol. *60*, 771–775.

Lihana, R.W., Khamadi, S.A., Lwembe, R.M., Kinyua, J.G., Muriuki, J.K., Lagat, N.J., Okoth, F.A., Makokha, E.P., and Songok, E.M. (2009). HIV-1 subtype and viral tropism determination for evaluating antiretroviral therapy options: an analysis of archived Kenyan blood samples. BMC Infect. Dis. 2009 91 *9*, 1–8.

Lö Chel, H.F., Riemenschneider, M., Frishman, D., and Heider, D. (2018). SCOTCH: subtype A coreceptor tropism classification in HIV-1. Bioinformatics *34*, 2575–2580.

Loetscher, M., Geiser, T., O'Reilly, T., Zwahlen, R., Baggiolini, M., and Moser, B. (1994). Cloning of a human seven-transmembrane domain receptor, LESTR, that is highly expressed in leukocytes. J. Biol. Chem. *269*, 232–237.

Louis, J.M., Wondrak, E.M., Kimmel, A.R., Wingfield, P.T., and Nashed, N.T. (1999). Proteolytic Processing of HIV-1 Protease Precursor, Kinetics and Mechanism*. J. Biol. Chem. *274*, 3437–3442.

Louwagie, J., McCutchan, F.E., Peeters, M., Brennan, T.P., Sanders-Buell, E., Eddy, G.A., van der Groen, G., Fransen, K., Gershy-Damet, G.-M., Deleys, R., et al. (1993). Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS *7*, 769–780.

Low, A.J., Dong, W., Chan, D., Sing, T., Swanstrom, R., Jensen, M., Pillai, S., Good, B., and Harrigan, P.R. (2007). Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. Aids *21*, F21–F24.

Lu, K., Heng, X., and Summers, M.F. (2012). Structural Determinants and Mechanism of HIV-1 Genome Packaging. J. Mol. Biol. *410*, 609–633.

Lucic, B., and Lusic, M. (2016). Connecting HIV-1 integration and transcription: a step toward new treatments. FEBS Lett. *590*, 1927–1939.

Lwembe, R., Lihana, R.W., Ochieng', W., Panikulam, A., Mongoina, C.O., Palakudy, T., Koning, H. de, Ishizaki, A., Kageyama, S., Musoke, R., et al. (2009). Changes in the HIV Type 1 Envelope Gene from Non-Subtype B HIV Type 1-Infected Children in Kenya. AIDS Res. Hum. Retroviruses *25*, 141–147.

Mabeya, S.N., Nyamache, A.K., Ngugi, C., Nyerere, A., and Wekesa, R.L. (2018). Predominance of CCR5 tropism in non-B HIV-1 subtypes circulating in Kisii County, Kenya. East Afr. Med. J. *95*, 1889–1898.

Majello, B., Napolitano, G., Giordano, A., and Lania, L. (1999). Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. ONCOGNE *18*, 4598–4605.

Malim, M.H., and Cullen', B.R. (1993). Rev and the Fate of Pre-mRNA in the Nucleus: Implications for the Regulation of RNA Processing in Eukaryotes. Mol. Cell. Biol. *13*, 6180–6189.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Lett. to Nat. *424*, 99–101.

Mansky, L.M., Preveral, S., Selig, L., Benarous, R., and Benichou, S. (2000). The Interaction of Vpr with Uracil DNA Glycosylase Modulates the Human Immunodeficiency Virus Type 1 In Vivo Mutation Rate. J. Virol. *74*, 7039–7047.

Marchetti, G., Tincati, C., and Silvestri, G. (2013). Microbial translocation in the pathogenesis of HIV infection and AIDS. Clin. Microbiol. Rev. *26*, 2–18.

De Marco, A., Heuser, A.-M., Glass, B., Kräusslich, H.-G., Müller, B., and Briggs, J.A.G. (2012). Role of the SP2 Domain and Its Proteolytic Cleavage in HIV-1 Structural Maturation and Infectivity. J. Virol. *86*, 137008–13716.

Margottin, F., Bour, S., Durand, H., Selig, L., Benichou, S., Richard, V., and Et. al. (1988). ANovel Human WDProtein, h-NL TrCP, that Interacts with HIV-1 Vpu Connects CD4 to the ER Degradation Pathway through an F-Box Motif. Mol. Cell *1*, 565–574.

Marin, M., Rose, K.M., Kozak, S.L., and Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat. Med. *9*, 1398–1403.

Marlink, R., Kanki, P., Thior, I., Travers, K., Eisen, G., Siby, T., Traore, I., Hsieh, C.C., Dia, M.C., Gueye, E.H., et al. (1994). Reduced Rate of Disease Development After HIV-2 Infection as Compared to HIV-1. Science (80-.). *265*, 1587–1590.

Martin-Serrano, J., Zang, T., and Bieniasz, P.D. (2001). HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nature *7*, 1313–1318.

Martin-Serrano, J., Zang, T., and Bieniasz, P.D. (2003). Role of ESCRT-I in Retroviral Budding. J. Virol. *77*, 4794–4804.

Martinez-Mariño, B., Foster, H., Hao, Y., and Levy, J.A. (2007). Differential gene expression in CD8+ cells from HIV-1-infected subjects showing suppression of HIV replication. Virology *362*, 217–225.

Masankwa, G., Magiri, E., Nyanjom, S.G., Hunja, C., Kitawi, R.C., Nzomo, T., Mwatelah, R.S., Kimulwo, M.J., Okendo, J., Oketch, J., et al. (2016). HIV-1 Coreceptor Tropism among Kenyans Under Highly Active Antiretroviral Therapy. J. Biol. *6*, 113–118.

Mattern, T., SCHOLZ, W., FELLER, A.C., FLAD, H. -D, and ULMER, A.J. (1991). Expression of CD26 (Dipeptidyl Peptidase IV) on Resting and Activated Human T-Lymphocytes. Scand. J. Immunol. *33*, 737–748.

Maziere, J., Landureau, J., Giral, P., and Et. al. (1994). Lovastatin inhibits HIV-1 expression in H9 human T lymphocytes cultured in cholesterol-poor medium. Biomed Pharmacother *48*, 63–67.

Mbisa, J.L., Delviks-Frankenberry, K.A., Thomas, J.A., Gorelick, R.J., and Pathak, V.K. (2009). Real-time PCR analysis of HIV-1 replication post-entry events. Methods Mol. Biol. *485*, 55–72.

McDougal, J., KENNEDY, M., SLIGH, J., CORT, S., MAWLE, A., and NICHOLSON, J. (1986). Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. Science (80-.). 382–388.

McElrath, M.J., De Rosa, S.C., Moodie, Z., Dubey, S., Kierstead, L., Janes, H., Defawe, O.D., Carter, D.K., Hural, J., Akondy, R., et al. (2008). HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. Lancet (London, England) *372*, 1894–1905.

Mckinnon, L.R., Nagelkerke, N.J., Kaul, R., Shaw, S.Y., and Capina, R. (2012). HIV-1 Clade D Is Associated with Increased Rates of CD4 Decline in a Kenyan Cohort. PLoS One *7*, 49797.

Mclean, A.R., and Michie, C.A. (1995). In vivo estimates of division and death rates of human T lymphocytes. Proc. Natl. Acad. Sci. U. S. A. *92*, 3707–3711.

McMichael, A.J., Borrow, P., Tomaras, G.D., Goonetilleke, N., and Haynes, B.F. (2010). The immune response during acute HIV-1 infection: Clues for vaccine development. Nat. Rev. Immunol. *10*, 11–23.

Micheal, M.H., Cullen, B.R., Hauber, J., and Bohnlein, S. (1989). Functional dissection of the HIV-1 Rev trans-activator—Derivation of a trans-dominant repressor of Rev function. Cell *58*, 205–214.

Mild, M., Esbjörnsson, J., Fenyö, E.M., and Medstrand, P. (2007). Frequent Intrapatient Recombination between Human Immunodeficiency Virus Type 1 R5 and X4 Envelopes: Implications for Coreceptor Switch. J. Virol. *81*, 3369–3376.

Milich, L., Margolin, B.H., and Swanstrom, R. (1997). Patterns of Amino Acid Variability in NSI-like and SI-like V3 Sequences and a Linked Change in the CD4-Binding Domain of the HIV-1 Env Protein. Virology *239*, 108–118.

Miller, M.D., Wang, B., and Bushman, F.D. (1995). Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. J. Virol. *69*, 3938–3944.

Miller, R.J., Cairns, J.S., Bridges, S., and Sarver, N. (2000). Human Immunodeficiency Virus and AIDS: Insights from Animal Lentiviruses. J. Virol. *74*, 7187–7195.

Miller, W.C., Rosenberg, N.E., Rutstein, S.E., and Powers, K.A. (2010). The Role of Acute and Early HIV Infection in the Sexual Transmission of HIV. Curr. Opin. HIV AIDS *5*, 277.

Mortier, V., Dauwe, K., Vancoillie, L., Staelens, D., Van Wanzeele, F., Vogelaers, D., Vandekerckhove, L., Chalmet, K., and Verhofstede, C. (2013). Frequency and predictors of HIV-1 co-receptor switch in treatment naive patients. PLoS One *8*, 10–14.

Mosier, D.E. (2009). How HIV changes its tropism : evolution and adaptation ?

Mothé, B.R., Horton, H., Carter, D.K., Allen, T.M., Liebl, M.E., Skinner, P., Vogel, T.U., Fuenger, S., Vielhuber, K., Rehrauer, W., et al. (2002). Dominance of CD8 Responses Specific for Epitopes Bound by a Single Major Histocompatibility Complex Class I Molecule during the Acute Phase of Viral Infection. J. Virol. *76*, 875–884.

Moyle, G.J., Wildfire, A., Mandalia, S., Mayer, H., Goodrich, J., Whitcomb, J., and Gazzard, B.G. (2005). Epidemiology and Predictive Factors for Chemokine Receptor Use in HIV-1 Infection. J. Infect. Dis. *191*, 866–873.

Mulinge, M., Lemaire, M., Servais, J.-Y.Y., Rybicki, A., Struck, D., Santos Da Silva, E., Verhofstede, C., Lie, Y., Seguin-Devaux, C., Schmit, J.-C.C., et al. (2013). HIV-1 Tropism Determination Using a Phenotypic Env Recombinant Viral Assay Highlights Overestimation of CXCR4-Usage by Genotypic Prediction Algorithms for CRRF01_AE and CRF02_AG. PLoS One *8*, 60566–60579.

Musey, L., Hughes, J., Schacker, T.W., Shea, T., Corey, L., and McElrath, M.J. (1997). CYTOTOXIC-T-CELL RESPONSES, VIRAL LOAD, AND DISEASE PROGRESSION IN EARLY HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 INFECTION A BSTRACT Background Early in human immunodeficiency vi. N. Engl. J. Med. *337*, 1267–1274.

Muthumani, K., Hwang, D.S., Desai, B.M., Zhang, D., Dayes, N., Green, D.R., and Weiner, D.B. (2002). HIV-1 Vpr Induces Apoptosis through Caspase 9 in T Cells and Peripheral Blood Mononuclear Cells*. J. Biol. Chem. *277*, 37820–37831.

Myers, G., Macinnes, K., and Korber, B. (1992). The Emergence of Simian/Human Immunodeficiency Viruses. AIDS Res. Hum. Retroviruses *8*, 373–381.

Nabatov, A.A., Pollakis, G., Linnemann, T., Kliphius, A., Chalaby, M.I.M., and Paxton, W.A. (2004). Intrapatient Alterations in the Human Immunodeficiency Virus Type 1 gp120 V1V2 and V3 Regions Differentially Modulate Coreceptor Usage , Virus Inhibition by CC / CXC Chemokines , Soluble CD4 , and the b12 and 2G12 Monoclonal Antibodies. J. Virol. *78*, 524–530.

Naganawa, S., Yokoyama, M., Shiino, T., Suzuki, T., and Ishigatsubo, Y. (2008). Net Positive Charge of HIV-1 CRF01_AE V3 Sequence Regulates Viral Sensitivity to Humoral Immunity. PLoS One *3*, 3206.

Naif, H.M. (2013). Pathogenesis of HIV infection. Infect. Dis. Rep. *5*, 26–30.

Nasir, I.A., Babayo, A., Ahmad, A.E., Okechukwu, E.C., and Ojeamiren, I.N. (2015). The roles of CCR5delta32 allele on HIV/AIDS epidemic in native Africans. J. Appl. Virol. *4*, 46.

Navia, M.A., Fitzgerald, P.M.., McKeever, B.M., Herber, W.K., Sigal, I.S., and Springer, J.P. (1988). Three-dimensional structure of aspartyl protease fom HIV-1. Nature *337*, 615–620.

Nduva, G.M., Hassan, A.S., Nazziwa, J., Graham, S.M., Esbjörnsson, J., Sanders, E.J., and

Esbjörnsson, J. (2020). HIV-1 Transmission Patterns Within and Between Risk Groups in Coastal Kenya. Sci. Rep. *10*.

Nichols, W.G., Steel, H.M., Bonny, T., Adkison, K., Curtis, L., Millard, J., Kabeya, K., Clumeck, N., Triangle Park, R., and Carolina, N. (2008). Hepatotoxicity Observed in Clinical Trials of Aplaviroc (GW873140). Antimicrob. Agents Chemother. *52*, 858–865.

Nielsen, M.H., Pedersen, F.S., and Kjems, J. (2005). Molecular strategies to inhibit HIV-1 replication. Retrovirology *2*, 1–20.

Nisole, S., and Saïb, A. (2004). Early steps of retrovirus replicative cycle. Retrovirology *1*, 1–20.

Novikova, M., Adams, L.J., Fontana, J., Gres, A.T., Balasubramaniam, M., Winkler, D.C., Kudchodkar, S.B., Soheilian, F., Sarafianos, S.G., Steven, A.C., et al. (2018). Identification of a structural element in HIV-1 gag required for virus particle assembly and maturation. MBio *9*, 1–18.

O' Brien, S.K., Cao, H., Nathans, R., Ali, A., and Rana, T.M. (2010). P-TEFb Kinase Complex Phosphorylates Histone H1 to Regulate Expression of Cellular and HIV-1 Genes * □ S. J. Biol. Chem. *285*, 29713–29720.

Ohtsuki, T., Tsuda, H., and Morimoto, C. (2000). Good or evil: CD26 and HIV infection. J. Dermatol. Sci. *22*, 152–160.

De Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., Janse Van Rensburg, E., Wensing, A.M.J., Van De Vijver, D.A., et al. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinforma. Appl. NOTE *21*, 3797–3800.

Ott, M., Geyer, M., and Zhou, Q. (2011). The control of HIV transcription: Keeping RNA polymerase II on track. Cell Host Microbe *10*, 426–435.

Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (1996). A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. Proc. Natl. Acad. Sci. U. S. A. *93*, 5572.

Palm, A.A., Esbjörnsson, J., Månsson, F., Kvist, A., Isberg, P.E., Biague, A., Da Silva, Z.J.,

Jansson, M., Norrgren, H., and Medstrand, P. (2014). Faster Progression to AIDS and AIDS-Related Death Among Seroincident Individuals Infected With Recombinant HIV-1 A3/CRF02_AG Compared With Sub-subtype A3. J. Infect. Dis. *209*, 721–728.

Palmer, S., Wiegand, A.P., Maldarelli, F., Bazmi, H., Mican, J.M., Polis, M., Dewar, R.L., Planta, A., Liu, S., Metcalf, J.A., et al. (2003). New Real-Time Reverse Transcriptase-Initiated PCR Assay with Single-Copy Sensitivity for Human Immunodeficiency Virus Type 1 RNA in Plasma. J. Clin. Microbiol. *41*, 4531–4536.

Panos, G., and Nelson, M. (2007). HIV-1 tropism. Biomarkers Med. *1*, 473–481.

Parkin, N.T., Chamorro, M., and Varmus, H.E. (1992). Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. J. Virol. *66*, 5147–5151.

Pastore, C., Nedellec, R., Ramos, A., Pontow, S., Ratner, L., and Mosier, D.E. (2006a). Human Immunodeficiency Virus Type 1 Coreceptor Switching: V1/V2 Gain-of-Fitness Mutations Compensate for V3 Loss-of-Fitness Mutations. J. Virol. *80*, 750–758.

Pastore, C., Nedellec, R., Ramos, A., Pontow, S., Ratner, L., and Mosier, D.E. (2006b). Human Immunodeficiency Virus Type 1 Coreceptor Switching: V1/V2 Gain-of-Fitness Mutations Compensate for V3 Loss-of-Fitness Mutations †. J. Virol. *80*, 750–758.

Paxton, W., Connor, R.I., and Landau, N.R. (1993). Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. J. Virol. *67*, 7229–7237.

Peng, J., Marshall, N.F., and Price, D.H. (1998). Identification of a Cyclin Subunit Required for the Function of Drosophila P-TEFb*. J. Biol. Chem. *273*, 13855–13860.

Perkins, A., Cochrane, A.W., Ruben, S.M., and Rosen, C.A. (1989). Structural and functional characterization of the human immunodeficiency virus rev protein. J. Acquir. Immune Defic. Syndr. *2*, 256–263.

Perlman, M., and Resh, M.D. (2006). Identification of an Intracellular Trafficking and Assembly Pathway for HIV-1 Gag. Traffic *7*, 731–745.

Peterlin, M.B., and Price, D.H. (2006). Controlling the Elongation Phaseof Transcription with P-TEFb. Mol. Cell Biol. *23*, 297–305.

Petersen, A., Cowan[3], S.A., Nielsen[3], J., Fischer[4], T.K., Fonager[4], J., Andreas, P., Cowan, S.A., Jens, N., Fischer, T.K., and Fonager, J. (2018). Surveillance and outbreak report Characterisation of HIV-1 transmission clusters and drug-resistant mutations in Denmark. Euro Surveill *23*, pii=170063.

Pettit, S.C., Everitt, L.E., Choudhury, S., Dunn, B.M., and Kaplan, A.H. (2004). Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism. J. Virol. *78*, 8477–8485.

Pettit, S.C., Clemente, J.C., Jeung, J.A., Dunn, B.M., and Kaplan, A.H. (2005). Ordered Processing of the Human Immunodeficiency Virus Type 1 GagPol Precursor Is Influenced by the Context of the Embedded Viral Protease. J. Virol. *79*, 10601–10607.

Pillai, S., Corbeil, J., Richman, D., and Corbeil, J. (2003). A New Perspective on V3 Phenotype Prediction. AIDS Res. Hum. Retroviruses *19*, 145–149.

Pineda-Peña, A.C., Faria, N.R., Imbrechts, S., Libin, P., Abecasis, A.B., Deforche, K., Gómez-López, A., Camacho, R.J., De Oliveira, T., and Vandamme, A.M. (2013). Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. Infect. Genet. Evol. *19*, 337–348.

Pinter, A., Honnen, W.J., He, Y., Gorny, M.K., Zolla-Pazner, S., and Kayman, S.C. (2004). The V1/V2 Domain of gp120 Is a Global Regulator of the Sensitivity of Primary Human Immunodeficiency Virus Type 1 Isolates to Neutralization by Antibodies Commonly Induced upon Infection. J. Virol. *78*, 5205–5215.

Pohlmann, S., Stolte, N., Münch, J., Haaft, P. Ten, Heeney, J.L., Stahl-Hennig, C., and Kirchhoff, F. (1999). Co-receptor Usage of BOB/GPR15 in Addition to CCR5 Has No Significant Effect on Replication of Simian Immunodeficiency Virus In Vivo. J. Infect. Dis. *180*, 1494–1502.

Pollakis, G., Kang, S., Kliphuis, A., Chalaby, M.I.M., Goudsmit, J., and Paxton, W.A. (2001). N-Linked Glycosylation of the HIV Type-1 gp120 Envelope Glycoprotein as a Major Determinant

of CCR5 and CXCR4 Coreceptor Utilization *. J. Biol. Chem. *276*, 13433–13441.

Poveda, E., Luna, J.D.D., Chueca, N., Garrido, C., Marta, A., Mendoza, C. De, and Garcı, F. (2009). Improvement in the Determination of HIV-1 Tropism Using the V3 Gene Sequence and a Combination of Bioinformatic Tools. J. Med. Virol. *81*, 763–767.

Powers, K.A., Poole, C., Pettifor, A.E., and Cohen, M.S. (2008). Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. Lancet. Infect. Dis. *8*, 553–563.

Prakash, R., Bhattacharjee, P., Blanchard, A., Musyoki, H., Anthony, J., Kimani, J., Gakii, G., Sirengo, M., Muraguri, N., Mziray, E., et al. (2018). Effects of exposure to an intensive HIV-prevention programme on behavioural changes among female sex workers in Nairobi, Kenya. Africa J. AIDS Res. *17*, 99–108.

Prosperi, M.C.F., Bracciale, L., Fabbiani, M., Di Giambenedetto, S., Razzolini, F., Meini, G., Colafigli, M., Marzocchetti, A., Cauda, R., Zazzi, M., et al. (2010). Comparative determination of HIV-1 co-receptor tropism by Enhanced Sensitivity Trofile, gp120 V3-loop RNA and DNA genotyping. Retrovirology *7*, 1–11.

Pruss, D., Bushman, F.D., and Wolffe, A.P. (1994). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. U. S. A. *91*, 5913–5917.

Quinn, T.C. (2007). Circumcision and HIV transmission. Curr. Opin. Infect. Dis. *20*, 33–38.

Quinn, T.C., Wawer, M.J., Sewankambo, N., Serwadda, D., Li, C., Wabwire-Mangen, F., Meehan, M.O., Lutalo, T., and Gray, R.H. (2000). Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N. Engl. J. Med. *342*, 921–929.

Rao, P.K.S. (2009). CCR5 inhibitors: Emerging promising HIV therapeutic strategy. Indian J. Sex. Transm. Dis. AIDS *30*, 1.

Rastogi, S., Agrahari, S., Ganga, S., Hospital, R., Singh, U.P., and Verma, A. (2011). Clinical Stages of HIV. Biolixir *I*, 29–31.

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Souyris, C., Sandres-Sauné, K., Cuzin, L.,

Marchou, B., Massip, P., and Izopet, J. (2008). Correlation between genotypic predictions based on V3 sequences and phenotypic determination of HIV-1 tropism. Aids *22*, F11-16.

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Souyris, C., Encinas, S., Sandres-Sauné, K., Pasquier, C., Marchou, B., Massip, P., et al. (2009). Genotypic prediction of human immunodeficiency virus type 1 CRF02-AG tropism. J Clin Microbiol *47*, 2292–2294.

Raymond, S., Delobel, P., Mavigner, M., Cazabat, M., Souyris, C., Encinas, S., Bruel, P., Sandres-sauné, K., Marchou, B., Massip, P., et al. (2010). Development and performance of a new recombinant virus phenotypic entry assay to determine HIV-1 coreceptor usage. J. Clin. Virol. *47*, 126–130.

Raymond, S., Delobel, P., Chaix, M.L., Cazabat, M., Encinas, S., Bruel, P., Sandres-Sauné, K., Marchou, B., Massip, P., and Izopet, J. (2011). Genotypic prediction of HIV-1 subtype D tropism. Retrovirology *8*, 1–8.

Raymond, S., Nicot, F., Jeanne, N., Delfour, O., Carcenac, R., Lefebvre, C., Cazabat, M., Sauné, K., Delobel, P., and Izopet, J. (2017). Performance comparison of next-generation sequencing platforms for determining HIV-1 coreceptor use. Sci. Rep. *7*, 1–8.

Re, F., Braaten, D., Franke, E.K., and Luban, J. (1995). Human Immunodeficiency Virus Type 1 Vpr Arrests the Cell Cycle in G 2 by Inhibiting the Activation of p34 cdc2-Cyclin B. J. Virol. *69*, 6859–6864.

Recordon-Pinson, P., Soulié, C., Flandre, P., Descamps, D., Lazrek, M., Charpentier, C., Montes, B., Trabaud, M.A., Cottalorda, J., Schneider, V., et al. (2010). Evaluation of the genotypic prediction of HIV-1 coreceptor use versus a phenotypic assay and correlation with the virological response to maraviroc: The ANRS genotropism study. Antimicrob. Agents Chemother. *54*, 3335–3340.

Reeves, J.D., Gallo, S.A., Ahmad, N., Miamidian, J.L., Harvey, P.E., Sharron, M., Pöhlmann, S., Sfakianos, J.N., Derdeyn, C.A., Blumenthal, R., et al. (2002). Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. Proc. Natl. Acad. Sci. U. S. A. *99*, 16249–16254.

Refaeli, Y., Levy, D.N., and Weinert, D.B. (1995). The glucocorticoid receptor type II complex

is a target of the HIV-1 vpr gene product. Med. Sci. *92*, 3621–3625.

Reichert, J.M. (2017). mAbs Antibodies to watch in 2017 Antibodies to watch in 2017. Perspectives (Montclair). *9*, 167–181.

Resch, W., Hoffman, N., and Swanstrom, R. (2001). Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology *288*, 51–62.

Riemenschneider, M., Cashin, K.Y., Budeus, B., Sierra, S., Shirvani-Dastgerdi, E., Bayanolhagh, S., Kaiser, R., Gorry, P.R., and Heider, D. (2016). Genotypic Prediction of Co-receptor Tropism of HIV-1 Subtypes A and C. Sci. Rep. *6*, 1–9.

Robbins, K.E., Kostrikis, L.G., Brown, T.M., Anzala, O., Shin, S., Robbins, K.E.;, Kostrikis, L.G.;, Brown, T.M.;, Anzala, O.;, Shin, S.;, et al. (1999). Genetic Analysis of Human Immunodefiency Virus Type I Strains in Kenya: A Comparison Using Phylogenetic Analysis and a Combinatorial Melting Assay. AIDS Res. Hum. Retroviruses *15*, 329–335.

Robertson, D., Anderson, J., Bradac, J., Carr, J., Foley, B., Funkhouser, R., Gao, F., Hahn, B., Kalish, M., Kuiken, C., et al. (2000). HIV-1 nomenclature proposal. Science (80-.). *288*, 55–56.

Robertson, D.L., Sharp, P.M., Mc Cutchan, F.E., and Hahn, B.H. (1995). Recombination in HIV-1. Nature *374*, 124–126.

Rosenthal, N. (1994). Regulation of Gene Expression. N. Engl. J. Med. *331*, 931–933.

Rousseau, C.M., Learn, G.H., Bhattacharya, T., Nickle, D.C., Heckerman, D., Chetty, S., Brander, C., Goulder, P.J.R., Walker, B.D., Kiepiela, P., et al. (2007). Extensive Intrasubtype Recombination in South African Human Immunodeficiency Virus Type 1 Subtype C Infections. J. Virol. *81*, 4492.

Ruiz, A.P., Ajasin, D.O., Ramasamy, S., Desmarais, V., Eugenin, E.A., and Prasad, V.R. (2019). A Naturally occurring Polymorphism in the HIV-1 Tat Basic Domain Inhibits Uptake by Bystander Cells and Leads to Reduced Neuroinflammation. Sci. Rep. *9*.

Saad, J.S., Miller, J., Tai, J., Kim, A., Ghanam, R.H., and Summers, M.F. (2006). Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc. Natl.

Acad. Sci. U. S. A. *103*, 11364–11369.

Saag, M., Heera, J., Goodrich, J., Dejesus, E., Clumeck, N., Cooper, D., Walmsley, S., Ting, N., Coakley, E., Reeves, J., et al. (2008). Reanalysis of the MERIT Study with the Enhanced Trofile Assay (MERIT-ES). HIV AIDS Rev. *12*, 346–352.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature *382*, 722–726.

Sánchez, V., Masiá, M., Robledano, C., Padilla, S., Ramos, J.M., and Gutiérrez, F. (2010). Performance of genotypic algorithms for predicting HIV-1 tropism measured against the enhanced-sensitivity trofile coreceptor tropism assay. J. Clin. Microbiol. *48*, 4135–4139.

Sander, O., Sing, T., Sommer, I., Low, A.J., Cheung, P.K., Harrigan, P.R., Lengauer, T., and Domingues, F.S. (2007). Structural Descriptors of gp120 V3 Loop for the Prediction of HIV-1 Coreceptor Usage. PLoS Comput. Biol. *3*, 58–65.

Sarafianos, S.G., Das, K., Tantillo, C., Clark Jr, A.D., and Ding, J. (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J. *20*, 1449–1461.

Sarafianos, S.G., Marchand, B., Das, K., Himmel, D., Parniak, M.A., Hughes, S.H., and Arnold, E. (2009). Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol *385*, 693–713.

Saunders, C.J., Mccaffrey, R.A., Zharkikh, I., Kraft, Z., Malenbaum, S.E., Burke, B., Cheng-Mayer, C., and Stamatatos, L. (2005). The V1, V2, and V3 Regions of the Human Immunodeficiency Virus Type 1 Envelope Differentially Affect the Viral Phenotype in an Isolate-Dependent Manner. J. Virol. *79*, 9069–9080.

Scarborough, R.J., and Gatignol, A. (2015). HIV and Ribozymes. Adv. Exp. Med. Biol. *848*, 97–116.

Scarlata, S., and Carter, C. (2003). Role of HIV-1 Gag domains in viral assembly. Biochim. Biophys. Acta - Biomembr. *1614*, 62–72.

Schäfer, A., Bogerd, H.P., and Cullen, B.R. (2004). Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology *328*, 163–168.

Schaller, T., Ocwieja, K.E., Rasaiyaah, J., Price, A.J., Brady, T.L., Roth, S.L., Hué, S., Fletcher, A.J., Lee, K., KewalRamani, V.N., et al. (2011). HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. PLoS Pathog. *7*, e1002439.

Schindler, M., Würfl, S., Benaroch, P., Greenough, T.C., Daniels, R., Easterbrook, P., Brenner, M., Münch, J., and Kirchhoff, F. (2003). Down-Modulation of Mature Major Histocompatibility Complex Class II and Up-Regulation of Invariant Chain Cell Surface Expression Are Well-Conserved Functions of Human and Simian Immunodeficiency Virus nef Alleles. J. Virol. *77*, 10548–10556.

Schmidtmayerova, H., Notrett, H.S.L.M., Nuovo, G., Raabe, T., Flanagant, C.R., Dubrovsky, L., Gendelmanttii, H.E., Cerami, A., Bukrinsky, M., and Sherry, B. (1996). Human immunodeficiency virus type 1 infection alters chemokine f8 peptide expression in human monocytes: Implications for recruitment of leukocytes into brain and lymph nodes. Proc. Natl. Acad. Sci. U. S. A. *93*, 700–704.

Schooley, R.T., Merigan, T.C., Gaut, P., Hirsch, M.S., Holodniy, M., Flynn, T., Liu, S., Byington, R.E., Henochowicz, S., Gubish, E., et al. (1990). Recombinant soluble CD4 therapy in patients with the acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. A phase I-II escalating dosage trial. Ann. Intern. Med. *112*, 247–253.

Schuitemaker, H., Koot, M., Kootstra, N.A., Dercksen, M.W., de Goede, R.E., van Steenwijk, R.P., Lange, J.M., Schattenkerk, J.K., Miedema, F., and Tersmette, M. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. J. Virol. *66*, 1354–1360.

Sechet, M., Roussel, C., Schmit, J.-L., Saroufim, C., Ghomari, K., Merrien, D., Cordier, F., Pik, J.-J., Landgraf, N., Douadi, Y., et al. (2015). X4 Tropic Virus Prediction Is Associated with a Nadir CD4 T-Cell Count below 100 Cells/mm 3. Intervirology *58*, 155–159.

Seclé, E., Garrido, C., Del Mar Gonzá Lez, M., Gonzá Lez-Lahoz, J., de Mendoza, C., Soriano, V., Poveda, E., Seclén, E., Garrido, C., González, M. del M., et al. (2010). High sensitivity of specific genotypic tools for detection of X4 variants in antiretroviral-experienced patients suitable to be treated with CCR5 antagonists. J. Antimicrob. Chemother. *65*, 1486–1492.

Seclén, E., Soriano, V., González, M.M., Gómez, S., Thielen, A., and Poveda, E. (2011). High Concordance between the Position-Specific Scoring Matrix and Geno2pheno Algorithms for Genotypic Interpretation of HIV-1 Tropism: V3 Length as the Major Cause of Disagreement. J. Clin. Microbiol. *49*, 3380.

Seitz, R. (2016). Human Immunodeficiency Virus (HIV). Transfus. Med. Hemotherapy *43*, 203–222.

Seto, M., Aikawa, K., Miyamoto, N., Aramaki, Y., Kanzaki, N., Takashima, K., Kuze, Y., Iizawa, Y., Baba, M., and Shiraishi, M. (2006). Highly potent and orally active CCR5 antagonists as anti-HIV-1 agents: Synthesis and biological activities of 1-benzazocine derivatives containing a sulfoxide moiety. J. Med. Chem. *49*, 2037–2048.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. *1*, 1–22.

Shaw, G.M., and Hunter, E. (2012). HIV Transmission. Cold Spring Harb. Perspect. Med. *2*.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646–650.

Shehu-Xhilaga, M., Crowe, S.M., and Mak, J. (2001). Maintenance of the Gag/Gag-Pol Ratio Is Important for Human Immunodeficiency Virus Type 1 RNA Dimerization and Viral Infectivity. J. Virol. *75*, 1834–1841.

Sheppard W. Hyanes, Celum Coonie, Micheal L. Nelson, O'Brien Stephen, Dean Micheal, Carrington Mary, D.D. (2002). HIV_1_Infection_in_Individuals_With_the.13.pdf. J. Acquir. Immune Defic. Syndr. *29*, 3-7–313.

Shinn, P., Chen, H., Berry, C., Ecker, J.R., Bushman, F., and Jolla, L. (2002). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. Cell *110*, 521–529.

Shioda, T., Oka, S., Ida, S., Nokihara, K., Toriyoshi,', H., Mori, S., Takebe, Y., Kimura, S., Shimada, K., and Nagail, Y. (1994). A Naturally Occurring Single Basic Amino Acid Substitution V3 Region of the Human Immunodeficiency Virus Type Env Protein Alters the Cellular Host Range and Antigenic Structure of the Viru. J. Virol. *68*, 7689–7696.

Siddik, A.B., Haas, A., Rahman, S., Aralaguppe, S.G., Amogne, W., Bader, J., Klimkait, T., and Neogi, U. (2018). Phenotypic co-receptor tropism and Maraviroc sensitivity in HIV-1 subtype C from East Africa OPEN. Sci. Rep. *8*, 2363–2369.

Silliciano, R.F., and Lusic, M. (2017). Nuclear landscape of HIV-1 infection and integration. Reviews *15*, 71–80.

Simon, V., Ho, D.D., and Karim, Q.A. (2006). HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet *368*, 489–504.

Sing, T., Low, A.J., Beerenwinkel, N., Sander, O., Cheung, P.K., Domingues, F.S., Büch, J., Däumer, M., Kaiser, R., Lengauer, T., et al. (2007). Predicting HIV coreceptor usage on the basis of genetic and clinical covariates. Antivir. Ther. *12*, 1097–1106.

Singh, P.K., Plumb, M.R., Ferris, A.L., Iben, J.R., Wu, X., Fadel, H.J., Luke, B.T., Esnault, C., Poeschla, E.M., Hughes, S.H., et al. (2015). LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes Dev. *29*, 2287–2297.

Skrabal, K., Low, A.J., Dong, W., Cheung, P.K., Mammano, F., Richard, P., Skrabal, K., Low, A.J., Dong, W., Sing, T., et al. (2007). Determining Human Immunodeficiency Virus Coreceptor Use in a Clinical Setting : Degree of Correlation between Two Phenotypic Assays and a Bioinformatic Model Determining Human Immunodeficiency Virus Coreceptor Use in a Clinical Setting : Degree of Correla. J. Clin. Microbiol. *45*, 279–284.

Skripkin, E., Paillart, J.-C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994). Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. Proc. Natl. Acad. Sci. U. S. A. *91*, 4945–4949.

Soulié, C., Derache, A., Aimé, C., Marcelin, A.G., Carcelain, G., Simon, A., Katlama, C., and Calvez, V. (2008). Comparison of two genotypic algorithms to determine HIV-1 tropism. HIV Med. *9*, 1–5.

Soulié, C., Fourati, S., Lambert-Niclot, S., Malet, I., Wirden, M., Tubiana, R., Valantin, M.A., Katlama, C., Calvez, V., and Marcelin, A.G. (2010). Factors associated with proviral DNA HIV-1 tropism in antiretroviral therapy-treated patients with fully suppressed plasma HIV viral load: Implications for the clinical use of CCR5 antagonists. J. Antimicrob. Chemother. *65*, 749–751.

Soulié, C., Fofana, D.B., Boukli, N., Sayon, S., Lambert-Niclot, S., Wirden, M., Simon, A., Katlama, C., Calvez, V., Girard, P.M., et al. (2016). Performance of genotypic algorithms for predicting tropism of HIV-1CRF02_AG subtype. J. Clin. Virol. *76*, 51–54.

Stade, K., Ford, C.S., Guthrie, C., and Weis, K. (1997). Exportin 1 (Crm1p) is an essential nuclear export factor. Cell *90*, 1041–1050.

Stanfield, R., Cabezas, E., Satterthwait, A., Stura, E., Profy, A., and Wilson, I. (1999). Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing Fabs. Structure *7*, 131–138.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol. Biol. Evol. *37*, 1237–1239.

Sterjovski, J., Churchill, M.J., Ellett, A., Gray, L.R., Roche, M.J., Dunfee, R.L., Purcell, D.F.J., Saksena, N., Wang, B., Sonza, S., et al. (2007). Asn 362 in gp120 contributes to enhanced fusogenicity by CCR5-restricted HIV-1 envelope glycoprotein variants from patients with AIDS. Retrovirology *4*, 89.

Sterjovski, J., Roche, M., Churchill, M.J., Ellett, A., Farrugia, W., Gray, L.R., Cowley, D., Poumbourios, P., Lee, B., Wesselingh, S.L., et al. (2010). An altered and more efficient mechanism of CCR5 engagement contributes to macrophage tropism of CCR5-using HIV-1 envelopes. Virology *404*, 269–278.

Stove, V., Van De Walle, I., Naessens, E., Coene, E., Stove, C., Plum, J., and Verhasselt, B. (2005). Human Immunodeficiency Virus Nef Induces Rapid Internalization of the T-Cell Coreceptor CD8. J. Virol. *79*, 11422–11433.

Strack, B., Calistri, A., Craig, S., Popova, E., and Göttlinger, H.G. (2003). AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell *114*, 689–699.

Strebel, K., Daugherty, D., Clouse, K., Cohen, D., Folks, T.M., and Martin, M. (1987). e HIV

“A” (sor) gene product is essential for virus infectivity. Nature *328*, 738–730.

Sun, Z., Ouyang, J., Zhao, B., An, M., Wang, L., Ding, H., and Han, X. (2020). Natural polymorphisms in HIV-1 CRF01_AE strain and profile of acquired drug resistance mutations in a long-term combination treatment cohort in northeastern China. BMC Infect. Dis. *20*, 1–10.

Suphaphiphat, P., Essex, M., and Lee, T.H. (2007). Mutations in the V3 stem versus the V3 crown and C4 region have different effects on the binding and fusion steps of human immunodeficiency virus type 1 gp120 interaction with the CCR5 coreceptor. Virology *360*, 182–190.

Swamy, M.N., Wu, H., and Shankar, P. (2016). Recent advances in RNAi-based strategies for Therapy and Prevention of HIV-1/AIDS Graphical Abstract HHS Public Access. Adv Drug Deliv Rev *103*, 174–186.

Swanstrom, R., and Coffin, J. (2012). HIV-1 Pathogenesis : The Virus. Perspect. Med. *2*.

Swigut, T., Shohdy, N., and Skowronski, J. (2001). Mechanism for down-regulation of CD28 by Nef Tomek. EMBO J. *20*, 209–212.

Swingler, S., Brichacek, B., Jacque, J.-M., Ulich, C., Zhou, J., and Stevenson, M. (2003). HIV-1 Nef intersects the macrophage CD40L signalling pathway to promote resting-cell infection Simon. Nature *424*, 213–219.

Tagat, J.R., McCombie, S.W., Nazareno, D., Labroli, M.A., Xiao, Y., Steensma, R.W., Strizki, J.M., Baroudy, B.M., Cox, K., Lachowicz, J., et al. (2004). Piperazine-based CCR5 antagonists as HIV-1 inhibitors. IV. Discovery of 1-[(4,6-dimethyl-5-pyrimidinyl)carbonyl]- 4-[4-[2-methoxy-1(R)-4-(trifluoromethyl)phenyl]ethyl-3(S)-methyl-1-piperazinyl]- 4-methylpiperidine (Sch-417690/Sch-D), a potent, highly selective, and orally bioavailable CCR5 antagonist. J. Med. Chem. *47*, 2405–2408.

Tago, A., McKinnon, L.R., Wanjiru, T., Muriuki, F., Munyao, J., Gakii, G., Akolo, M., Kariri, A., Reed, N., Shaw, S.Y., et al. (2021). Declines in HIV prevalence in female sex workers accessing an HIV treatment and prevention programme in Nairobi, Kenya over a 10-year period. AIDS *35*, 317–324.

Takeuchi, Y., Akutsu, M., Murayama, K., Shimizu, N., and Hoshino, H. (1991). Host range

mutant of human immunodeficiency virus type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. J. Virol. *65*, 1710–1718.

Tang, S., Murakami, T., Agresta, B.E., Campbell, S., Freed, E.O., and Levin, J.G. (2001). Human Immunodeficiency Virus Type 1 N-Terminal Capsid Mutants That Exhibit Aberrant Core Morphology and Are Blocked in Initiation of Reverse Transcription in Infected Cells. J. Virol. *75*, 9357–9366.

Taylor, B.S., and Hammer, S.M. (2008). The challenge of HIV-1 subtype diversity. N Engl J Med *359*, 1965–1966.

Thomas, D.C., Voronin, Y.A., Nikolenko, G.N., Chen, J., Hu, W.-S., and Pathak, V.K. (2007). Determination of the Ex Vivo Rates of Human Immunodeficiency Virus Type 1 Reverse Transcription by Using Novel Strand-Specific Amplification Analysis. J. Virol. *81*, 4798–4807.

Thomas, J.A., Gagliardi, T.D., Alvord, W.G., Lubomirski, M., Bosche, W.J., and Gorelick, R.J. (2006). Human immunodeficiency virus type 1 nucleocapsid zinc-finger mutations cause defects in reverse transcription and integration. Virology *353*, 41–51.

Tovanabutra, S., Sanders, E.J., Graham, S.M., Mwangome, M., Peshu, N., Mcclelland, R.S., Muhaari, A., Crossler, J., Price, M.A., Gilmour, J., et al. (2010). Evaluation of HIV Type 1 Strains in Men Having Sex with Men and in Female Sex Workers in Mombasa, Kenya. AIDS Res. Hum. Retroviruses *26*, 123–131.

Tremblay, C., Hardy, I., Lalonnde, R., Trottier, B., and Tsarevsky, I. (2013). HIV-1 tropism testing and clinical management of CCR5 antagonists: Quebec review and recommendations. Can J Infect Dis Med Microbiol 2013;24(4):202-208. Infect. Dis. Med. Microbiol. *24*, 202–208.

Triques, K., Bourgeois, A., Vidal, N., Mpoudi-Ngole, E., Delaporte;, E., and Peetersi, M. (2000). Near-Full-Length Genome Sequencing of Divergent African HIV Type 1 Subtype F Viruses Leads to the to the Identification of a New HIV Type 1 Subtype Designated K. AIDS Res. Hum. Retroviruses *16*, 139–151.

Tristem, M., Marshall1, C., Karpas, A., and Hill2, F. (1992). Evolution of the primate lentiviruses: evidence from vpx and vpr. EMBO J. *1*, 3405–341.

Trouplin, V., Salvatori, F., Cappello, F., Obry, V., Brelot, A., Heveker, N., Alizon, M., Scarlatti,

G., Clavel, F., and Mammano, F. (2001). Determination of Coreceptor Usage of Human Immunodeficiency Virus Type 1 from Patient Plasma Samples by Using a Recombinant Phenotypic Assay. J. Virol. *75*, 251–259.

Tsukamoto, T., Langford, D.T., Thayer, S.A., Ajasin, D., and Eugenin, E.A. (2020). HIV-1 Tat: Role in Bystander Toxicity. Front. Cell. Infect. Microbiol. *10*.

Turner, B.G., and Summers, M.F. (1999). Structural biology of HIV. J. Mol. Biol. *285*, 1–32.

Turville, S.G., Cameron, P.U., Handley, A., Lin, G., Pöhlmann, S., Doms, R.W., and Cunningham, A.L. (2002). Diversity of receptors binding HIV on dendritic cell subsets. Nat. Immunol. *3*, 975–983.

Tyagi, M., Rusnati, M., Presta, M., and Giacca, M. (2001). Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. J. Biol. Chem. *276*, 3254–3261.

Vandekerckhove, L.P.R., Wensing, A.M.J., Kaiser, R., Brun-Vézinet, F., Clotet, B., De Luca, A., Dressler, S., Garcia, F., Geretti, A.M., Klimkait, T., et al. (2011). European guidelines on the clinical management of HIV-1 tropism testing. Lancet Infect. Dis. *11*, 394–407.

Vandenbroucke, I., Van Marck, H., Mostmans, W., Van Eygen, V., Rondelez, E., Thys, K., Van Baelen, K., Fransen, K., Vaira, D., Kabeya, K., et al. (2010). HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. AIDS Res. Ther. *7*, 3–8.

Vangelista, L., Secchi, M., and Lusso, P. (2008). Rational design of novel HIV-1 entry inhibitors by RANTES engineering. Vaccine *26*, 3008–3015.

Verhofstede, C., Vandekerckhove, L., Eygen, V. Van, Vogelaers, D., and Stuyver, L. (2009). CXCR4-Using HIV Type 1 Variants Are More Commonly Found in Peripheral Blood Mononuclear Cell DNA Than in Plasma RNA. J. Acquir. Immune Defic. Syndr. *50*, 126–136.

Verhofstede, C., Brudney, D., Reynaerts, J., Vaira, D., Fransen, K., De Bel, A., Seguin-Devaux, C., De Wit, S., Vandekerckhove, L., and Geretti, A.-M.M. (2011). Concordance between HIV-1 genotypic coreceptor tropism predictions based on plasma RNA and proviral DNA. HIV Med. *12*, 544–552.

De Vincenzi, I. (1994). A Longitudinal Study of Human Immunodeficiency Virus Transmission by Heterosexual Partners. N. Engl. J. Med. *331*, 341–346.

VonSchwedler, U.K., Stuchell, M., Müller, B., Ward, D.M., Chung, H.Y., Morita, E., Wang, H.E., Davis, T., He, G.P., Cimbora, D.M., et al. (2003). The protein network of HIV budding. Cell *114*, 701–713.

Voronin, Y., Chohan, B., Emerman, M., and Overbaugh, J. (2007). Primary Isolates of Human Immunodeficiency Virus Type 1 Are Usually Dominated by the Major Variants Found in Blood. J. Virol. *81*, 10232–10241.

Wade, J., Sterjovski, J., Gray, L., Roche, M., Chiavaroli, L., Ellett, A., Jakobsen, M.R., Cowley, D., da Fonseca Pereira, C., Saksena, N., et al. (2010). Enhanced CD4+ cellular apoptosis by CCR5-restricted HIV-1 envelope glycoprotein variants from patients with progressive HIV-1 infection. Virology *396*, 246–255.

Wang, J.-K., Kiyokawa, E., Verdin, E., and Trono, D. (2000). The Nef protein of HIV-1 associates with rafts and primes T cells for activation. PNAS *97*, 394–399.

Wanjiru, R., Nyariki, E., Babu, H., Lwingi, I., Liku, J., Jama, Z., Kung, M., Ngurukiri, P., Maisha Fiti Study Champions, T., Nyamweya, C., et al. (2022). Beaten but not down! Exploring resilience among female sex workers (FSWs) in Nairobi, Kenya. BMC Public Health *22*, 1–12.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature *460*, 711–716.

Wei, X.M., Xu, H.F., Cheng, X. Di, Bu, N., and Zhou, H.Z. (2016). Position 22 of the V3 loop is associated with HIV infectivity. Arch. Virol. 2016 1623 *162*, 637–643.

Weller, S.C. (1993). A meta-analysis of condom effectiveness in reducing sexually transmitted HIV. Soc. Sci. Med. *36*, 1635–1644.

West, C.M., Van Der Wel, H., Sassi, S., and Gaucher, E.A. (2004). Cytoplasmic glycosylation of protein-hydroxyproline and its relationship to other glycosylation pathways. Biochem. Biophys. Acta *1673*, 29–44.

Westby, M., and Van Der Ryst, E. (2010). Review CCR5 antagonists: host-targeted antiviral agents for the treatment of HIV infection, 4 years on. Antivir. Chem. Chemother. *20*, 179–192.

Westby, M., Lewis, M., Whitcomb, J., Youle, M., Pozniak, A.L., James, I.T., Jenkins, T.M., Perros, M., and van der Ryst, E. (2006). Emergence of CXCR4-Using Human Immunodeficiency Virus Type 1 (HIV-1) Variants in a Minority of HIV-1-Infected Patients following Treatment with the CCR5 Antagonist Maraviroc Is from a Pretreatment CXCR4-Using Virus Reservoir. J. Virol. *80*, 4909–4920.

Whitcomb, J.M., Huang, W., Fransen, S., Limoli, K., Toma, J., Wrin, T., Chappey, C., Kiss, L.D.B., Paxinos, E.E., and Petropoulos, C.J. (2007a). Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. Antimicrob. Agents Chemother. *51*, 566–575.

Whitcomb, J.M., Huang, W., Fransen, S., Limoli, K., Toma, J., Wrin, T., Chappey, C., Kiss, L.D.B., Paxinos, E.E., and Petropoulos, C.J. (2007b). Development and Characterization of a Novel Single-Cycle Recombinant-Virus Assay To Determine Human Immunodeficiency Virus Type 1 Coreceptor Tropism □. *51*, 566–575.

Wilen, C.B., Tilton, J.C., and Doms, R.W. (2012a). HIV: Cell binding and entry. Cold Spring Harb. Perspect. Med. *2*.

Wilen, C.B., Tilton, J.C., and Doms, R.W. (2012b). HIV: Cell Binding and Entry. Perspect. Med. *2*, 1–5.

Wilkin, T.J., and Gulick, R.M. (2012). CCR5 Antagonism in HIV Infection: Current Concepts and Future Opportunities. Annu. Rev. Med. *63*, 81–93.

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. (2008). High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States. PLoS Biol *6*, 96.

Wiskerchen, M., and Muesing, M.A. (1995). Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. J. Virol. *69*, 376–386.

Wlodawer, A., Miller, M., Jaskolski, M., and Baldwin, E. (1989). Conserved Folding in

Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease . Sci. New Ser. *245*, 616–621.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.-J., Kabongo, J.-M.M., Kalengayi, R.M., Marck, E. Van, et al. (2008). Direct Evidence of Extensive Diversity of HIV-1 in Kinshasa by 1960. Nature *455*, 661–664.

Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. Science (80-.). *28*.

Wyatt, R., Thali, M., Tilley, S., Pinter, A., Posner, M., Ho, D., Robinson, J., and Sodroski, J. (1992). Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to a component of the CD4 binding site in the fourth conserved region. J. Virol. *66*, 6997–7004.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol. Cell Biol. *21*, 227–237.

Yamaguchi, J., Vallari, A., Mcarthur, C., Sthreshley, L., Cloherty, G.A., Berg, M.G., and Rodgers, M.A. (2019). Complete Genome Sequence of CG-0018a-01 Establishes HIV-1 Subtype L. J. Acquir. Immune Defic. Syndr. *83*, 319–322.

Yebra, G., de Mulder, M., Martín, L., Pérez-Cachafeiro, S., Rodríguez, C., Labarga, P., García, F., Tural, C., Jaén, À., Navarro, G., et al. (2011). Sensitivity of seven HIV subtyping tools differs among subtypes/recombinants in the Spanish cohort of naïve HIV-infected patients (CoRIS). Antiviral Res. *89*, 19–25.

Yokoyama, M., Naganawa, S., Yoshimura, K., Matsushita, S., and Sato, H. (2012). Structural Dynamics of HIV-1 Envelope Gp120 Outer Domain with V3 Loop. PLoS One *7*.

Yu, H., Jetzt, A.E., Ron, Y., Preston, B.D., and Dougherty, J.P. (1998). The nature of human immunodeficiency virus type 1 strand transfers. J. Biol. Chem. *273*, 28384–28391.

Yu, Q., König, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat. Struct. Mol. Biol. *11*, 435–442.

Yuan, H., Xie, Y.-M., and Chen, I.S.Y. (2003). Depletion of Wee-1 Kinase Is Necessary for both Human Immunodeficiency Virus Type 1 Vpr- and Gamma Irradiation-Induced Apoptosis. J. Virol. *77*, 2063–2070.

Zack, J.A., Haislip, A.M., Krogstad, P., and Chen, I.S. (1992). Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle. J. Virol. *66*, 1717–1725.

Zapp, M.L., Hope, T.J., Parslowt, T.G., and Green, M.R. (1991). Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: A dual function for an arginine-rich binding motif (mRNA export/pre-mRNA splicing/RNA binding protein/virus expression). Proc. Natl. Acad. Sci. U. S. A. *88*, 7734–7738.

Zaunders, J.J., Hazenberg, M.D., and Schuitemaker, H. (2002). Cell turnover and cell tropism in HIV-1 Miles P. *10*, 275–278.

Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., and Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. Cell *101*, 173–185.

Zhang, F., Dou, Z., Ma, Y., Zhang, Y., Zhao, Y., Zhao, D., Zhou, S., Bulterys, M., Zhu, H., and Chen, R.Y. (2011). Effect of earlier initiation of antiretroviral treatment and increased treatment coverage on HIV-related mortality in China: a national observational cohort study. Lancet. Infect. Dis. *11*, 516–524.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Lett. to Nat. *242*, 94–97.

Zhang, L., He, T., Talal, A., Wang, G., Frankel, S.S., and Ho, D.D. (1998). In Vivo Distribution of the Human Immunodeficiency Virus/Simian Immunodeficiency Virus Coreceptors: CXCR4, CCR3, and CCR5. J. Virol. *72*, 5035–5045.

Zhao, Y., Cao, J., O’gorman, M.R.G., Yu, M., and Yogev, R. (1996). Effect of Human Immunodeficiency Virus Type 1 Protein R (vpr) Gene Expression on Basic Cellular Function of Fission Yeast Schizosaccharomyces pombe. J. Virol. *70*, 5821–5826.

Zheng, Y.-H., Plemenitas, A., Fielding, C.J., and Matija Peterlin, B. (2003). Nef increases the

synthesis of and transports cholesterol to lipid rafts and HIV-1 progeny virions. PNAS *100*, 8460–8465.

Zhong, J., Saxena, A., Borroni, E.M., Vangelista, L., and Vento, S. (2018). The Expanding Therapeutic Perspective of CCR5 Blockade. Front. Immunol *8*, a1981.

Zhou, Q., Tiandao, L., and Price, D.H. (2012). RNA Polymerase II Elongation Control. Annu. Rev. Biochem. *81*, 119–143.

Zhu, P., Chertova, E., Bess, J., Lifson, J.D., Arthur, L.O., Liu, J., Taylor, K.A., and Roux, K.H. (2003). Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proc. Natl. Acad. Sci. U. S. A. *100*, 15812–15817.

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B., and Price, D.H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. Genes Dev. *11*, 2622–2632.

Zhuang, J., Jetzt, A.E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B.D., and Dougherty, J.P. (2002). Human Immunodeficiency Virus Type 1 Recombination: Rate, Fidelity, and Putative Hot Spots. J. Virol. *76*, 11273–11282.

Zolla-pazner, S., Gomy, M., and Nyambi, P. (1999). The implications of antigenic diversity for vaccine development. Immunol. Lett. *66*, 159–164.

Zolla-Pazner, S., Sharpe Cohen, S., Boyd, D., Kong, X.-P., Seaman, M., Nussenzweig, M., Klein, F., Overbaugh, J., Totrov, M., Zolla-Pazner, C.S., et al. (2016). Structure/Function Studies Involving the V3 Region of the HIV-1 Envelope Delineate Multiple Factors That Affect Neutralization Sensitivity. J. Virol. *90*, 636–649.