LUND UNIVERSITY

Dynamics of HIV-1 infection within and between hosts

Nazziwa, Jamirah

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Nazziwa, J. (2022). *Dynamics of HIV-1 infection within and between hosts*. [Doctoral Thesis (compilation), Department of Translational Medicine]. Lund University, Faculty of Medicine.

Total number of authors:
1

Creative Commons License:
CC BY-NC-ND

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Dynamics of HIV-1 infection within and between hosts

JAMIRAH NAZZIWA | DEPARTMENT OF TRANSLATIONAL MEDICINE
FACULTY OF MEDICINE | LUND UNIVERSITY

Dynamics of HIV-1 infection within and between hosts

Jamirah Nazziwa

Department of Translational Medicine

LUND UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Agardh Lecture Hall, CRC, Jan Waldenströms gata 35, Malmö.
Wednesday, June 15th, 2022 at 13:00.

Faculty opponent
Professor Fernando Gonzalez Candelas,
Institute for Integrative Systems Biology, University of Valencia, Spain

Supervisor
Associate Professor Joakim Esbjörnsson, Lund University, Sweden

Co-supervisors
Professor Sarah Rowland-Jones, University of Oxford, United Kingdom
Professor Philippe Lemey, Katholieke Universiteit Leuven, Belgium
Professor Johan Malmström, Lund University, Sweden
Dr Amin S Hassan, KEMRI-Wellcome Trust Research Program, Kenya

Organization LUND UNIVERSITY Faculty of Medicine Department of Translational Medicine	**Document name** DOCTORAL DISSERTATION **Date of issue** June 15, 2022
Author Jamirah Nazziwa	**Sponsoring organization**

Title Dynamics of HIV-1 infection within and between hosts

Abstract

A deep understanding of HIV-1 transmission, evolution, and virus-host interactions during disease progression is needed to fast-track the design of HIV-1 intervention strategies. However, HIV-1 transmission and evolutionary dynamics in regions with the highest HIV-1 burden (such as in sub-Saharan Africa, SSA), are not well understood. In **paper I**, we characterised the HIV-1 epidemic in Nigeria using HIV-1 pol sequences (n=1442). The circulating recombinant forms CRF02_AG, CRF43_02G, and subtype G were the main circulating strains. We estimated five major HIV-1 introduction events – associated with a previous civil war in Nigeria. Within-country epidemics first originated and expanded within large urban centres, before migrating to smaller rural areas. A follow-up literature review (**paper II**) on the nature of the epidemic in SSA revealed some localized epidemics in specific communities (with limited HIV-1 transmission between neighboring communities) – and that human migration linked to economic activities (including mining and fishing) may have contributed to increased HIV-1 transmission in SSA.

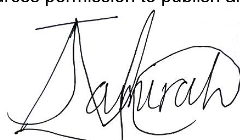
It is well known that most HIV-1 transmissions occur during the acute HIV-1 infection (AHI) phase – potentially linked to a high viral load. However, inter-host associations between AHI symptoms, viral load and disease progression are not well understood. In **paper III**, we explored the associations between innate immune responses during AHI, and AHI symptoms using pre-infection and post-infection plasma samples from HIV-1 infected individuals (n=55) from five African countries and Sweden. We demonstrated that individuals with strong innate immune responses had increased odds of showing symptoms during AHI, regardless of HIV-1 subtype, age or risk group. In addition, interferon gamma-induced protein 10 (IP-10) was independently associated with AHI symptoms and could be a marker for stronger innate immunity. In **paper IV**, we quantified temporal (pre-infection and post-infection) changes in protein expression in plasma from matched HIV-1 infected study participants (n=54). We found individual-specific longitudinal protein expression profiles after HIV-1 infection. We also identified some proteins (n=20) associated with AHI symptoms, and others (n=37) associated with virus control and disease progression. Although these findings provide further insights into the virus-host interactions associated with disease progression, it is well-established that such interactions may differ between adults and infants due to varying immune systems. Therefore, **in paper V**, we aimed to characterize within-infant HIV-1 evolution after perinatal HIV-1 transmission. Sequentially collected clonal HIV-1 sequences from 14 treatment naïve infants and their mothers were phylogenetically analysed. Overall, 78% of the infants were infected with a single transmitted founder virus. Immune-driven selection was apparent at three months post-infection. Majority of the observed $CD8^+$ T-cell escape mutations were transmitted from the mothers and did not revert during the first year. Work in this thesis deepens current understanding of HIV-1 transmission and evolution (intra- and inter-host), shaped by the immune system and viral-hosts mechanisms that regulate HIV-1 pathogenesis.

Key words: HIV-1, phylogenetics, acute HIV-1 infection, acute retroviral syndrome, innate immune responses, plasma proteome, intra-host HIV-1 evolution, mother to child HIV-1 transmission, disease progression

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title 1652-8220 Doctoral Dissertation Series 2022:93		**ISBN** 978-91-8021-254-0
Recipient's notes	**Number of pages** 104	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Date 2022-05-10

Dynamics of HIV-1 infection within and between hosts

Jamirah Nazziwa

Lund 2022

Department of Translational Medicine, Systems Virology Group

LUND
UNIVERSITY

Front cover photo by Zosia Rostomian and NIAID - Scanning electron micrograph of HIV-1 particles (yellow) emerging from an infected T-cell (blue)

Back cover photo by Dr Zsófia Ilona Szojka

Copyright pp 1- 104 Jamirah Nazziwa 2022

Front cover photo © 2020 The Regents of the University of California, Lawrence Berkeley National Laboratory

Paper I © 2020 The Authors. Published by Springer Nature: Scientific reports, 10, 3468. (CC BY 4.0)

Paper II © 2021 The Authors. Published by MDPI: Viruses,13(6), 1174. (CC BY 4.0)

Paper III © 2021 The Authors. Published by Oxford University Press: Journal of Clinical Infectious Diseases, 73, 5, 832-841. (CC BY NC ND 4.0)

Paper IV © by the Authors (Manuscript submitted)

Paper V © by the Authors (Manuscript unpublished)

This work was supported by: The Faculty of Medicine, Lund University, Sweden; Swedish Research Council; Swedish Society for Medical Research; The Royal Physiographic Society; The Crafoord Foundation; and Knut and Alice Wallenberg Foundation – NBIS Bioinformatics long term support (WABI)

Systems Virology, Department of Translation Medicine
Faculty of Medicine, Lund University, Sweden

ISBN 978-91-8021-254-0
ISSN 1652-8220
Lund University, Faculty of Medicine Doctoral Dissertation Series 2022:93

Printed in Sweden by Media-Tryck, Lund University, Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

To Isaac Anguria.

In memory of Dr HARR FREEYA NJAI (Died 2014)

"It takes two flints to make a fire." (Louisa May Alcott)

Table of Contents

List of papers 9

List of figures 11

List of tables 12

Abbreviations 13

Context of this thesis 14

Overview of this thesis 16

Popular summary 17

Introduction 19

HIV discovery 19

HIV structure, genome, origin and types 20

HIV-1 epidemiology 24

HIV-1 replication and treatment 26

HIV-1 recombination 32

HIV-1 immune responses 33

Cellular components of innate immune responses to HIV-1 33

Soluble factors in innate immune responses to HIV-1 37

Adaptive immune responses 40

Humoral responses 43

Immune responses in infants 44

HIV-1 Pathogenesis: Natural history of infection in adults 45

Mucosal barrier 45

Transmitted/founder virus 46

Eclipse phase of acute HIV-1 infection 46

Systemic phase/peak viremia 47

Latent infection 47

Viral set point 48

$CD4^+$T-cell dynamics 49

HIV-1 Pathogenesis: Natural history of infection in infants 50

HIV disease progression....52
Classification of HIV-1 disease progression....52
Viral factors associated with HIV-1 disease progression....53
Host factors associated with HIV-1 disease progression....53
Use of phylogenetics in HIV-1 research....54
HIV-1 phylogenetics between hosts....56
HIV-1 phylogenetics within hosts....56
HIV-1 plasma proteomics....60
Characteristics of the plasma proteome....60
Detection of proteins in plasma....61
Comparison of depleted and neat plasma....66
Aims of this doctoral dissertation 67
Specific Aims....67
Materials and methods.... 69
Cohorts....69
Study design....72
Definitions....72
Laboratory procedures (methods)....73
Statistical analyses....77
Ethical considerations....79
Results and discussion.... 81
Molecular phylogenetic description of the HIV-1 epidemic in Nigeria and sub-Saharan Africa....81
Dynamics of early innate immune responses and plasma proteome before and during hyper-acute HIV-1 infection....82
Infant intra-host HIV-1 evolution and CTL responses during acute HIV-1...85
Conclusions, limitations and future perspectives.... 89
Acknowledgements.... 91
References.... 95

List of papers

This thesis is based on five papers, referred to with roman numerals (I-V).

I. **Nazziwa J**, Faria N.R, Chaplin B, Rawizza H, Kanki P, Dakum P, Abimiku A, Charurat M, Ndembi N and Esbjörnsson J. **Characterisation of HIV-1 Molecular Epidemiology in Nigeria: Origin, Diversity, Demography and Geographic Spread**. *Scientific Reports* 10, 3468 (2020).

II. Nduva G.M*, **Nazziwa J***, Hassan A.S, Sanders E, and Esbjörnsson J. **The Role of Phylogenetics in Discerning HIV-1 Mixing among Vulnerable Populations and Geographic Regions in Sub-Saharan Africa: A Systematic Review**. *Viruses*. 13(6):1174 (2021). **Shared first authorship*

III. Hassan A.S, Hare J, Gounder K, **Nazziwa J**, Karlson S, Olsson L, Streatfield C, Kamali A, Karita E, Kilembe W, Price M.A, Borrow P, Björkman P, Kaleebu P, Allen S, Hunter E, Ndung'u T, Gilmour J, Rowland-Jones S, Esbjörnsson J* and Sanders E.J*. **A Stronger Innate Immune Response During Hyperacute HIV-1 Infection is associated with acute retroviral syndrome.** *Clinical Infectious Diseases*, ciab139 (2021). **Shared senior authorship*

IV. **Nazziwa J**, Freyhult E, Hong M, Årman F, Hare J, Gounder K, Rezeli M, Mohanty T, Kjellström S, Kamali A, Karita E, Kilembe W, Price M.A, Kaleebu P, Allen S, Hunter E, Ndung'u T, Gilmour J, Rowland-Jones S, Sanders E, Malmström J, Hassan A.S*, and Esbjörnsson J*. **Quantifying the dynamics of blood plasma proteome in acute HIV-1 infection**. (2022) (*Manuscript unpublished*). **Shared senior authorship*

V. **Nazziwa J**, Andrews S, Hou M.H, Bruhn C.W, Miguel A. Garcia-Knight M.A, Slyker J, Hill S, Payne B.L, Mbori-Ngacha D, Lemey P, John-Stewart G, Rowland-Jones S*, and Esbjörnsson J*. **Higher HIV-1 evolutionary rate is associated with cytotoxic T lymphocyte escape mutations in infants** (2022) (*Manuscript submitted*). **Shared senior authorship*

All publications have been reprinted with permission from the publishers.

The following papers are not included in this dissertation but are of relevance to the field and have been generated during my doctoral education:

A. Nduva G.M, Hassan A.S, **Nazziwa J**, Graham S.M, Esbjörnsson J, and Sanders E.J. **HIV-1 Transmission Patterns Within and Between Risk Groups in Coastal Kenya**. *Scientific Reports* 10, 6775 (2020).

B. Boswell M. T, **Nazziwa J**, Kuroki K, Palm A, Karlson S, Månsson F, Biague A, da Silva Z. J, Onyango C.O, de Silva T.I, Jaye A, Norrgren H, Medstrand P, Jansson M, Maenaka K, Rowland-Jones S. L, Esbjörnsson J, the SWEGUB CORE group. **Intrahost evolution of the HIV-2 capsid correlates with progression to AIDS.** *MedRxiv* (2022).

List of figures

Figure 1. HIV structure. 20
Figure 2. HIV-1 genome 21
Figure 3. Origin of HIV-1 in the human population 22
Figure 4. Phylogenetic tree for HIV-1 groups, subtypes, and sub-subtypes............ 23
Figure 5. Adult HIV-1 prevalence and subtypes in Africa............ 25
Figure 6. Global CRF/subtype distribution............ 25
Figure 7. HIV-1 replication cycle............ 29
Figure 8. HIV-1 recombination............ 33
Figure 9. Dendritic cells............ 35
Figure 10. NK cell interactions with some of the inhibitory and activating receptors. 36
Figure 11. Antiviral effector CD4+ T-cells and functions of cytokines produced...... 41
Figure 12. Adult HIV-1 pathogenesis, and acute HIV-1 viral load dynamics. 49
Figure 13. Infant HIV-1 viral load dynamics in comparison to adults............ 51
Figure 14. Classification of disease progression. 52
Figure 15. The general phylogenetic inference process 54
Figure 16. The hierarchical phylogenetic model............ 59
Figure 17. Label-free proteomics: discovery strategies............ 64
Figure 18. Paper I – Summary of analytical procedures............ 73
Figure 19. Paper IV – Summary of plasma proteomics laboratory procedures 75
Figure 20. Paper V – Summary of analytical procedures............ 76

List of tables

Table 1. First HIV-1 cases discovered in Africa and Sweden 20

Table 2. Time to the Most Recent Common Ancestor(tMRCA) for HIV-1 subtypes or CRFs in the study population analysed in each of the separate studies..................... 26

Table 3. Cytokines produced by innate immune cells. 38

Table 4. Chemokines, their receptors and roles 39

Table 5. Use of phylogenetics in HIV-1 research 56

Table 6. Previous HIV-1 plasma proteomics research... 65

Table 7. Summary of study design and study participants included in the PhD projects. 72

Table 8: Summary of the statistical methods used in this thesis 77

Abbreviations

AHI	acute HIV-1 infection
AIDS	acquired immunodeficiency syndrome
ART	antiretroviral therapy
CD4	cluster of differentiation 4
CD8	cluster of differentiation 8
CRF	circulating recombinant form
CTL	cytotoxic T-lymphocyte
DIA	data independent acquisition
DNA	deoxyribonucleic acid
FGN1	fibrinogen-like protein 1
GRN1	progranulin
HIV	human immunodeficiency virus 1
HLA	human leukocyte antigen
IL	interleukin
IP-10	interferon gamma-induced protein 10
MHC	major histocompatibility complex
MS	mass spectrometry
RNA	ribonucleic acid
SIV	simian immunodeficiency virus
SWATH-MS	sequential window acquisition of all theoretical mass spectra
tMRCA	time to the most recent common ancestor
TNF	tumour necrosis factor
URF	unique recombinant form
Vif	viral infectivity factor
Vpr	viral protein R
Vpu	viral protein U
Vpx	viral protein X
WF	von Willebrand factor

Context of this thesis

This PhD research was conducted at the Systems Virology group, Department of Translational Medicine, Faculty of Medicine, Lund University. Samples were collected from six African countries (Kenya, Nigeria, Rwanda, South Africa, Uganda, and Zambia) as well as from Sweden. Close contacts and collaborations have been formed with the Systems Virology group and the researchers from these African countries.

At the outset of this project in 2017, there was a paucity of knowledge about the molecular epidemiology of HIV-1 in Nigeria on a national scale. We researched and published a research article on the topic with my main supervisor (Assoc Prof Joakim Esbjörnsson) and other researchers from the Institute of Human Virology Nigeria and Harvard University. In addition, I cooperated with a PhD student in the team who was researching the same topic, and as co-first authors, we produced a critical review article addressing the limitations and degree of usage of phylogenetics in HIV-1 prevention in sub-Saharan Africa.

As a large proportion of HIV-1 transmissions occur during acute HIV-1 infection, we delineated and published on the early innate immune responses during AHI, as well as how these responses relate to the acute retroviral syndrome, in collaboration with Prof Eduard Sanders, Dr Amin S Hassan and other investigators from the International AIDS Vaccine Initiative and Sweden. As a sub-study, we collaborated with Prof Thumbi Ndung'u from South Africa, and Prof Johan Malmström from the Department of Clinical Sciences, Lund University, Sweden to conduct a longitudinal matched study quantifying the plasma proteome before and throughout hyperacute HIV-1 infection (two weeks and one-month post-infection). In this study, we elucidated virus-host mechanisms involved in the regulation of HIV-1 disease progression. This is a particularly unique cohort since samples collected from the study participants before HIV-1 infection were available. To the best of my knowledge, this is the first study that has evaluated protein expression before and after HIV-1 infection using a longitudinal approach.

Data on acute HIV-1 infection in infants is scarce, owing in part to a scarcity of ART-naive samples. Furthermore, while intra-host evolution of HIV-1 in adults has been linked to disease progression, such studies in infants are lacking. In collaboration with Prof Sarah L Rowland-Jones from University of Oxford, we initiated a study to better understand vertical HIV-1 transmission and how infant immune selection pressure shapes the intra-host evolutionary dynamics of HIV-1.

With the assistance of my supervisors, I was involved in the laboratory work, data analyses, and wrote or contributed to the drafting of papers for all the research in this PhD – project. In light of the COVID-19 pandemic constraints, I am happy that I still managed to attend conferences and seminars, supervise one master's student, support my fellow PhD students, apply for external funds, and participate as a tutor in the Biomedicine programme at the Faculty of Medicine, Lund University.

The findings of this thesis represent the first comprehensive assessment of the Nigerian HIV-1 molecular epidemic, as well as significant contributions to our understanding of virus-host interactions during acute HIV-1 infection in both adults and infants.

Overview of this thesis

Paper I	
Aim	To investigate the molecular epidemiology of HIV-1 in Nigeria
Study design	Cross-sectional
Samples	366 previously unpublished HIV-1 *pol* sequences + 1076 LANL *pol* sequences
Site of collection	Abuja (Nigeria) and Los Alamos database
Analysis	Phylogenetics
Paper II	
Aim	To provide an overview of the contribution of phylogenetic inference in dissecting HIV-1 mixing between geographic areas and key populations
Study design	Systematic literature review
Samples	2365 articles (64 full-text reviews)
Site of collection	PubMed
Analysis	Systematic review
Paper III	
Aim	To elucidate associations between innate immune responses during hyperacute HIV-1 infection and acute retroviral syndrome.
Study design	Prospective, longitudinal, matched (two time points)
Samples	55 samples
Site of collection	Rwanda, Uganda, Zambia, Kenya and Sweden
Analysis	Acute retroviral syndrome, HIV-1 subtyping, innate immune responses and statistics
Paper IV	
Aim	To characterize the blood plasma proteome before, during and after acute HIV-1 infection
Study design	Prospective, longitudinal, matched
Samples	157 (54 patients, three-time points)
Site of collection	Kenya, Rwanda, Zambia, South Africa
Analysis	DIA-SWATH mass spectrophotometry, proteomic bioinformatics and statistics
Paper V	
Aim	To determine the *gag* and *nef* intra-host HIV-1 diversity and evolution within the first 15 months of life in infants with different rates of disease progression
Study design	Prospective, longitudinal, matched pairs
Samples	1210 *gag,* and 1264 *nef* clonal sequences from14 infants, and 14 mothers
Site of collection	Nairobi, Kenya
Analysis	Intra-host phylogenetics

Popular summary

HIV-1 is one of the deadliest viruses in human history, having killed between 27 and 48 million individuals since the epidemic began in the 1980s. By the end of 2020, around 38 million people were living with AIDS, with approximately 73% receiving antiretroviral medication. Despite treatment expansion, 1.5 million people were newly infected with HIV-1 and 680,000 were diagnosed with AIDS-related illnesses, with children accounting for 15% of the total. HIV-1 transmissions are most common in the initial few weeks following infection, when the virus load in the blood is highest and most persons are ignorant of their serostatus. Because HIV-1 attacks and slowly destroys immune system cells, and it constantly mutates to weaken or conceal from these cells' action, HIV-1 can be present for three to ten years before an infected individual develops AIDS if not treated.

We were able to investigate when the virus was introduced and how it propagated within Nigeria by looking at how HIV-1 has evolved in Nigeria, the country with the highest number of AIDS-related deaths in the world. We found three main forms of HIV-1 circulating in Nigeria: CRF02_AG, CRF43_02G, and subtype D, which were introduced as five significant sub-epidemics in the 1960s and 1970s. According to HIV-1 migration analysis, HIV-1 first emerged and spread in urban areas such as Abuja and Lagos before spreading to smaller communities. A systematic review of other studies on the topic revealed similar movement patterns in other Western African countries, with cross-border transmissions; nevertheless, certain epidemics in Eastern and Southern Africa are localised, with little indication of mixing with surrounding populations.

We were able to explore why some people experience symptoms during the first two weeks after infection while others do not by looking at how individuals responded to early innate immune responses in the first two weeks after infection. We discovered that those who have a stronger innate immune response are more likely to experience acute infection symptoms than those who have a lesser response. IP-10, for example, is 14 times greater and independently associated with these symptoms, and can be utilised to identify people with stronger immune responses. We also discovered that each

individual has a specific protein expression profile and that 21 per cent of the plasma proteome significantly changed before and during the first month of acute HIV-1 infection. Among the top proteins are vWF, FGN1, and GRN1, which have roles in immune activation and disease. We also identified proteins that differed considerably between individuals who progressed slowly and quickly, as well as those who had or did not have acute retroviral syndrome. We attempted to discover the mechanisms by which the virus interacts with the host during acute infection and how they interact to control the rate of disease progression.

Finally, we investigated how the virus evolves within an infant during his or her first year of life. Within an untreated individual, HIV-1 generates roughly a billion virus particles every two-three days, but only one or a handful of these virus variants are transmitted from the donor to the recipient. To understand the dynamics of these transmitted viruses, studies of these transmitted variants have mostly been conducted in adults. When we studied the HIV-1 transmission dynamics in infants, we discovered that the infection was caused by a single variant in 78 per cent of the infants, infection was caused by a single variant transmitted from the mother. The virus began to escape from the action of immune cells at three months, and the mutations responsible for this escape were transmitted from the mother and remained throughout the first year. Furthermore, viruses with escape mutations evolved faster than those without. This work sheds light on how the infant immune selection pressure influences how HIV-1 evolves within infants.

Introduction

HIV discovery

Signs of a newly emerging infectious disease among homosexuals in the United States were reported 41 years ago[1]. A few years later, various African and European countries began reporting their first cases of an illness with similar symptoms (**Table 1**). This resulted in an extraordinary mobilisation of clinicians and researchers to understand the etiological agent of the disease and battle it[1]. One of the first European teams was the Montagnier laboratory at the Pasteur Institute in France, and their goal was to determine whether the new disease was caused by the first human retrovirus, human T-cell leukaemia virus (HTLV), which Gallo discovered in 1981[2, 3].

Based on the clinical data, cultures were done from a lymph node biopsy of a patient with unexplained lymphadenopathy who was original thought to be the source of new virus, and both reverse transcriptase activity and RNA were detected in the cell culture supernatant[4]. There were similarities in the reverse transcriptase of the newly discovered virus with that of the HTLV virus, but no cross-reactions with the p24, hence the new virus was dubbed "Lymphadenopathy Associated Virus (LAV)"[4]. When the first complete genome sequence of LAV was published in 1985, it was discovered that LAV was unique from HTLV but had a genomic organization similar to lentivirus called Visna Maedi (an ovine lentivirus). Further studies revealed that LAV had a selective tropism for helper T-lymphocytes expressing CD4 molecule[5, 6]. A few months after the discovery of LAV, a new drug Antimoniotungstate (HPA 34) was developed. Unfortunately, downstream studies showed that HPA34 was ineffective[7].

To present, the Food and drug administration (FDA) has approved 32 generic drugs for the treatment of LAV, now known as Human immunodeficiency virus (HIV) and 75% of 37.7 million persons living with HIV are on treatment[8]. However, each year, over one million individuals become infected and over half a million die as a result of AIDS-related illnesses [8]. I believe that more basic science research is needed to find a solution and eliminate the HIV-1 epidemic. For example, it has been revealed that there are various variants of this virus, which I will explore further in the subsequent section.

Table 1. First HIV-1 cases discovered in Africa and Sweden.
The focus is on countries where samples in this thesis were collected

County	Year	Location	Risk group
Kenya[9]	1984	Nairobi	MSM
Nigeria[10]	1985	Lagos	FSW
Rwanda[11]	1983	Kigali	HET (MP)
South Africa[12]	1982	Pretoria	MSM
Sweden	1982	Stockholm	MSM
Uganda[13]	1982	Rakai/Masaka	HET (MP)
Zambia	1984	Lusaka	na

MSM, men who have sex with men; FSW, female sex worker; HET (MP), heterosexual with multiple partners; na, risk group not known.

HIV structure, genome, origin and types

HIV was categorised as a member of the genus Lentivirus in the family *Retroviridae* based on the reverse transcriptase activity as described in the preceding section[14]. In humans, it is the causative agent of acquired immunodeficiency syndrome (AIDS). HIV is an enveloped virus with two homologous genomic sequences, each of which consists of 9.7 kilobases of linear, non-segmented, single-stranded, positive-sense RNA. As illustrated in **Fig. 1**, these two RNA strands are encased by a conical capsid core surrounded by a matrix.

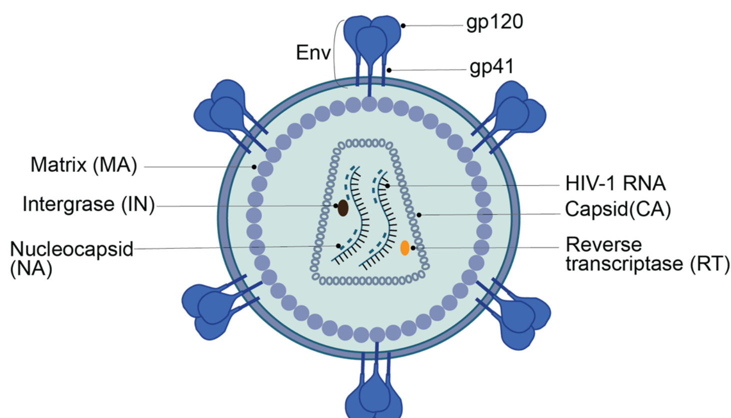


Figure 1. HIV structure.

The genome is made up of nine genes that code for fifteen viral proteins[15]. The key HIV genes are *gag* (codes for structural proteins – Matrix, Capsid, and Nucleocapsid),

pol (codes for enzymes – Protease, Reverse transcriptase, and Integrase), and *env* (codes for envelope proteins – Gp120, Gp41) (**Fig. 2**). The remaining genes encode regulatory (Tat, Rev) and accessory (Vif, Vpr, Vpu/Vpx, Nef), which each play a unique role in HIV replication and interaction with the host proteins.

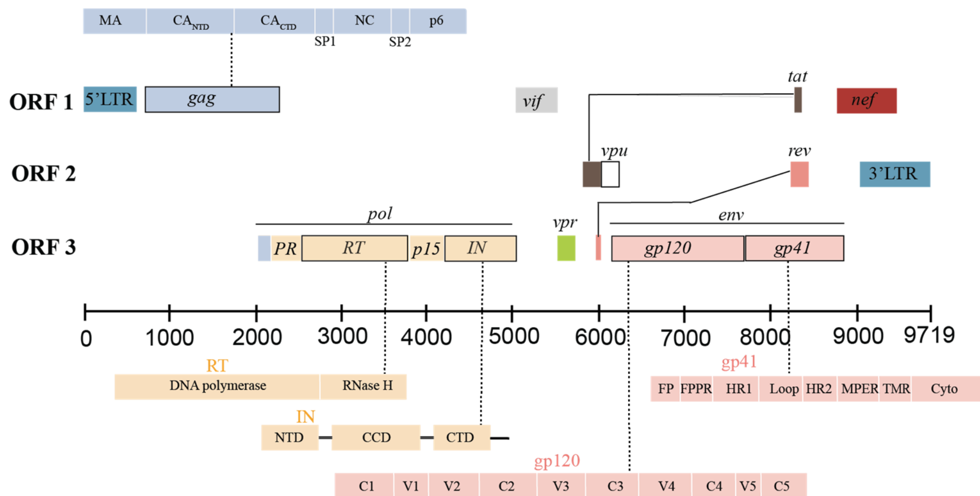


Figure 2. HIV-1 genome.
This figure depicts the HIV-1 genes in various colours. The genome has nine open reading frames (shown by rectangles) that code for 15 proteins. The structural genes *gag*, *env*, and *pol* are further subdivided. The Gag polyprotein has four domains: Matrix (MA), Capsid (CA_{NTD}) and C-terminal (CA_{CTD}), nucleocapsid (NC), and P6, as well as two short spacer peptides SP1, SP2. The viral enzymes Protease (PR), reverse transcriptase (RT) (consisting of DNA polymerase and RNase H enzymes), and Integrase (IN) are all included in the Pol protein (consists of three domains i.e. N-terminal domain, NTD; catalytic core domain, CCD and C-terminal domain, CTD). Env is produced as a Gp160 precursor glycoprotein that cleaves into the CD4 receptor-binding subunit Gp120 and the fusion protein subunit Gp41. The Gp120 protein is made up of five conserved sections (C1 to C5) and five variable regions (V1 to V5). The Gp41 comprise the N-terminal fusion peptide (FP), a Fusion peptide proximal region (FPPR), the Heptad repeat region 1 (HR1), a loop region followed by HR2, the Membrane-proximal external region (MPER), the Transmembrane region (TMR), and a Cytoplasmic domain (Cyto). Abbreviations: ORF, open reading frame; LTR, long terminal repeat, PR, protease; RT, reverse transcriptase; IN, integrase; *vif*, viral infectivity protein; *vpr*, virus protein r; *vpu*, virus protein unique; *nef*, negativity regulatory factor; *rev*, regulator of expression of viral protein; *tat*, trans-activator of transcription.

The HIV pandemic is characterised by substantial genetic diversity induced by a variety of reasons, including repeated zoonotic transmissions into the human population, as well as high rates of viral evolution and recombination[16]. These factors have resulted in two recognised types of HIV, namely HIV-1 and HIV-2, which have a nucleotide similarity of approximately 55%[17]. These two viruses are also linked to the different zoonotic origins, with HIV-1 originating from the SIV of chimpanzees[18], and HIV-2

originating from SIV of sooty mangabeys[19]. SIV was first discovered in Asian macaques kept captive in the US, and this SIV was causing HIV/AIDS-like symptoms in these macaques[20]. However, Asian macaques (SIVmac) did not carry SIV in the wild, whereas other non-human primate species such as chimpanzees (SIVcpz), gorillas (SIVgor), African green monkey (SIVagm), sooty mangabeys (SIVsm), and Sykes monkey (SIVsyk) did[21, 22]. SIV in the macaques was eventually discovered to be a result of cross-species transmissions from contaminated blood and tissue inoculations from sooty mangabeys[23]. Further studies found that SIV was mostly non-pathogenic to natural hosts and been present in these primates for over 30,000 years[24]. SIVsm, on the other hand, has the *vpx* gene found in HIV-2 whilst SIVcpz and SIVgor carry *vpu* genes found in HIV-1. These genomic relationships provided the first evidence that HIV emerged in humans from three primate species (chimpanzees, gorillas, and sooty mangabeys).

From phylogenetic studies, it was discovered that monkeys from central Africa have been infected with SIV for a long time, and that these monkeys likely transmitted SIV to chimpanzees[24]. The SIVcpz was then transmitted to gorillas and humans resulting in the first HIV-1 group **M**(major) epidemic[22]. The second independent SIVcpz transmission to humans resulted in HIV-1 group **N** (new) which was first identified in 1998, with only 13 documented cases[22]. The SIVcpz transmitted to gorillas was also transmitted to humans, resulting in the discovery of HIV-1 group **O** (outlier) in 1990 and HIV-1 group **P** in 2009[22] (**Fig. 3**).

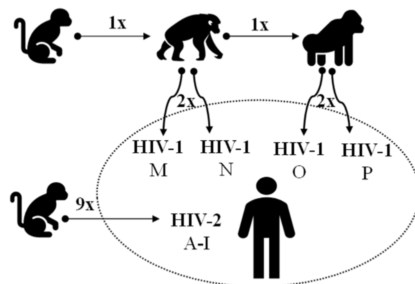


Figure 3. Origin of HIV-1 in the human population.
Adopted with permission from Ref[24].

Figure 4 depicts a phylogenetic tree demonstrating the relatedness of these taxa. SIVsm was also directly transmitted from sooty mangabeys to humans in West Africa via nine distinct zoonotic transmissions, resulting in the nine HIV-2 groups (A-I) primarily restricted in West Africa (**Fig. 3**).

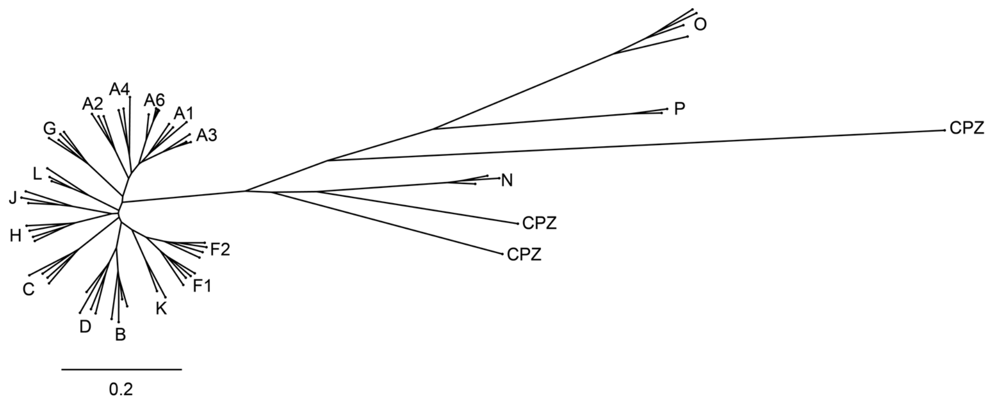


Figure 4. Phylogenetic tree for HIV-1 groups, subtypes, and sub-subtypes.
Group M is the most diverse HIV-1 group with different subtypes (A-D, F-H, J-L) and sub-subtypes for A (A1-A6) and F (F1-F2).

HIV is a pandemic, whereof HIV-1 accounts for approximately 98% and HIV-2 for approximately 2% of the pandemic[25]. More than 98% of the human HIV-1 is of Group M, which has been divided into 10 subtypes (A-D, F-H, J-L) with sub-sub types in A, F (**Fig. 4**), and as of May 2022, there are 118 circulating recombinant forms (CRFs, https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html)[22]. Genetic variation within a subtype can range from 15 to 20% depending on the genomic region, whereas the genetic variance between subtypes ranges from 25 to 35%[26]. Furthermore, the HIV-1 substitution rates between these subtypes and CRFs have been observed to differ, with subtype C having the highest rate at $3.5 \cdot \times 10^{-3}$ substitutions per site per year (s/s/y), and CRF02_AG having the lowest rate of $1.3 \cdot \times 10^{-3}$(s/s/y)[27] (**Table 2**). The diversity of HIV-1 genotypes varies dramatically from one strain to another and could be explained by an explosive within-host proliferation and a large, and still growing, number of infected persons that lead to very large population sizes. It has been observed that this diversity also varies by geographic location[28].

Groups A and B are the most common of the nine HIV-2 groups, and one HIV-2 CRF has been described – the HIV-2 CRF01_AB [29]. To establish a new CRF categorization for HIV, full genome sequences from at least three non-transmission related individuals are required, according to current HIV nomenclature guidelines[30].

HIV-1 epidemiology

HIV-1 group M was introduced into the human population between 1909 and 1930, with Kinshasa (DRC) serving as the epicentre[22, 31-35]. Since then, distinct HIV-1 group M subtypes have propagated and been observed in different countries, resulting in a global pandemic. In 2020, 37.7 million individuals were infected with HIV-1 (including 1.7 million children), 680,000 people died from AIDS-related illnesses, and 1.5 million people became infected[36]. Furthermore, 55% of all people with HIV-1 live in Eastern and Southern Africa, 12% in Western and Central Africa, 15% in Asia and the Pacific, and 5% in Western and Central Europe and North America. The remaining 12% is dispersed across areas with low HIV-1 prevalence such as the Caribbean, Eastern Europe, Central Asia, Latin America, Middle East, and North Africa (https://aidsinfo.unaids.org/). In Africa, the prevalence of HIV-1 is highest in Southern Africa, followed by Eastern Africa (**Fig. 5**).

HIV-1 subtype C accounts for 47% of all HIV-1 infections worldwide and is primarily found in Southern Africa, Ethiopia and India. Subtype A, which is mainly found in Eastern Africa but also in Eastern Europe and Central Asia, is responsible for 10% of HIV-1 all infections. Subtype B accounts for 12% of all HIV-1 infections and is primarily found in the Americas, Oceania, and Central and Western Europe. Recombinants between these subtypes account for 22.8% of all global HIV-1 infections, and various studies have demonstrated that the global fractions of URFs have steadily increased over time (**Fig. 6**). These global subtype differences have consequences for vaccine development antiretroviral therapy response and laboratory tests. The time to the most recent common ancestor (tMRCA) for the subtypes circulating in the countries where samples were collected is shown in Table 2. [27].

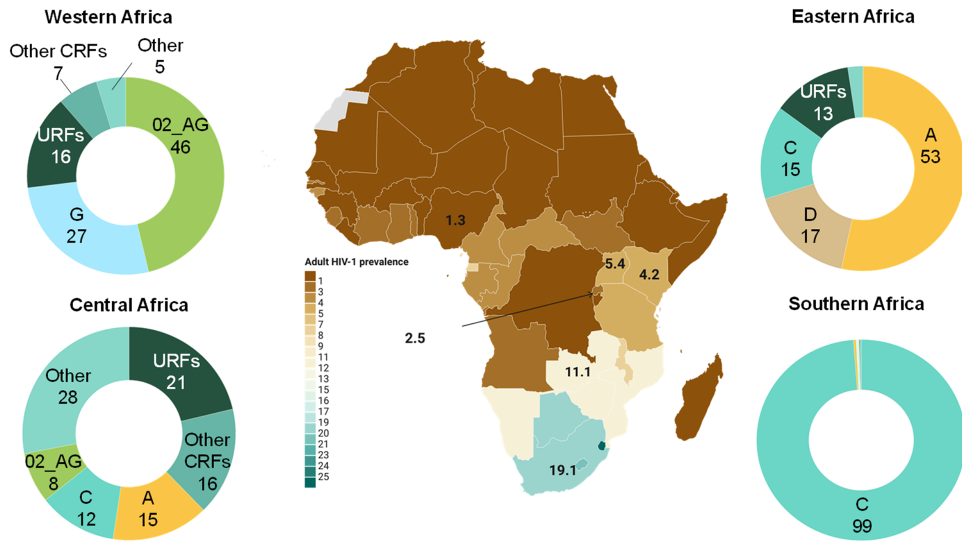


Figure 5. Adult HIV-1 prevalence and subtypes in Africa.
The focus is on countries where samples in this thesis were collected. Source: 2021 UNAIDS report (aidsinfo.uanids.org) and Ref[37]. Map was created with Datawrapper.

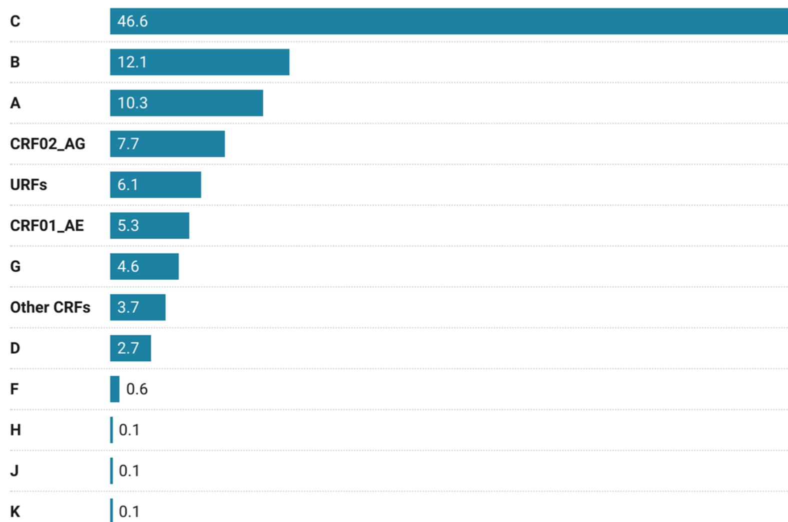


Figure 6. Global CRF/subtype distribution.
Bar length corresponds to the percentage of the globally infected population that is infected by the particular subtype. Source: 2021 UNAIDS report (aidsinfo.uanids.org) and Ref[37]. Map was created with Datawrapper

Table 2. Time to the Most Recent Common Ancestor(tMRCA) for HIV-1 subtypes or CRFs in the study population analysed in each of the separate studies

Estimates (median and 95% Highest Posterior Density; lower and upper limit) for the tMRCA, and the substitution rates (× 10^{-3}) of each subtype/ CRF based on the full genome for A1 – 02_AG, *pol* for 43_02G and *env* C2V3

Subtype/CRF	tMRCA (years)	Nucleotide substitution rate (x 10^{-3} s/s/y)
A1[27]	1953 (1944 – 1960)	3-0 (2.5 – 3.5)
B[27]	1957 (1952 – 1962)	2.8 (2.7 – 2.2)
C[27]	1965 (1957 – 1971)	**3.5 (2.9 – 4.2)**
D[27]	1956 (1927 – 1961)	1.9 (1.3 – 2.6)
G[27]	1969 (1961 – 1974)	2.4 (1.9 – 2.9)
02_AG[27]	1948 (1913 – 1969)	**1.3 (0-7 – 1.8)**
43_02G[38]	1971 (1953 – 1983)	2.7 (1.6 – 3.7)
Group M[31]	**1920 (1909 – 1930)**	**3.3**

HIV-1 replication and treatment

HIV-1 can spreads from person to person via contaminated breast milk, blood, semen or vaginal secretions. When HIV-1 enters the body, it spreads as a cell-free particle or cell-associated virus through target cells. HIV-1 replicate within an activated $CD4^+$ T-cell in 1-2 days, followed by programmed cell death of all infected $CD4^+$ T-cells(life span of 2.2 days)[39]. Below is an outline summarising the current knowledge about the HIV-1 replication cycle:

The early phase of replication

Cell binding, entry and fusion: The CD4 receptor is the major entry receptor for HIV-1 and its density influences the efficiency of HIV-1 infectivity[40]. Cells with low CD4 expression require high levels of virus titers to allow entry[41]. HIV-1 attach to the host cell through the envelope gene (reviewed in Ref [42]). Several specific (α4β7 integrin – the gut mucosal homing receptor for peripheral T cells, DC-SIGN – a dendritic cell-specific HIV-1-binding protein) and non-specific attachment factors (heparin sulphate proteoglycans) can augment HIV-1 infection by bringing the HIV-1 envelope protein Gp120 close to the CD4 receptor[43-45]. The C1-C3 region (which contains the CD4 binding site) of the Env Gp120 glycoprotein binds to the D1 domain of CD4, causing V1/V2 and then V3 rearrangements and the creation of a bridging sheet[46]. The relocated V3 and bridging sheet are crucial in exposing the binding site for the

chemokine co-receptor, CCR5 or CXCR4 CD4, as well as activating Env Gp41 glycoprotein. The co-receptor binds to the V3 loop, exposing and extending the Gp41 fusion peptide so that it can insert into the target cell's plasma membrane[47, 48]. The Gp120 subunits then dissociate, allowing the Gp41 HR1 and HR2 helices to connect (**Fig. 7A**). This closes the gap between the host and viral membranes, allowing membrane fusion and release of the mature core into the cytosol[48].

HIV-1 has been classified based on co-receptor usage/tropism, such as R5-tropic (CCR5) non-syncytium inducing and X4-tropic (CXCR4) syncytium inducing virus, as well as dual or mixed tropism where viruses can bind to either coreceptor[47, 49]. The tropism of HIV-1 has been linked to transmission and pathogenicity. R5 tropic viruses are preferentially transmitted and are the most common circulating strains in early HIV-1 infection whereas X4-tropic viruses are the opposite[50].

Maraviroc, a CCR5 antagonist drug that inhibits viral fusion by blocking the CCR5 coreceptor, is presently used to treat HIV[51]. Enfuvirtide, a fusion inhibitor that binds to HR1 in the Gp41 subunit, blocking the conformational changes essential for fusion, has also been approved by FDA to treat HIV-1[52]. Other drugs approved to prevent entry include 1) fostemsavir, an attachment inhibitor that binds to the Gp120 Env protein, preventing the virus from attaching to the CD4 receptors[53]; and 2) ibalizumab, a post-attachment inhibitor that binds to CD4 receptors on T-cells, preventing the HIV-1 Gp120 protein from changing shape to engage with co-receptors[54].

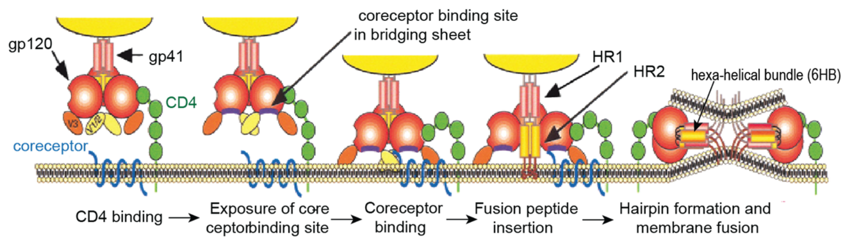
Reverse transcription: Once the virus has fused with the host cell, the virus's two single-stranded RNA genomes are transformed into linear double-stranded DNA. This action is thought to occur within the HIV-1 virion and is catalysed by the virus- pol encoded enzyme reverse transcriptase (RT). This enzyme is a heterodimer composed of two enzyme domains: DNA polymerase, which can copy either DNA or RNA template and RNaseH, which degrades RNA only after the synthesis of RNA-DNA duplex[55, 56]. RT requires a template (plus stranded genomic RNA), a primer (Lys3 tRNA) and dNTPs for the synthesis of the first DNA strand (minus strand) [57]. Because the viral capsid stays intact during fusion, HIV imports nucleotides required for DNA synthesis through dynamic pores on the capsid[58]. During each round of viral DNA synthesis,

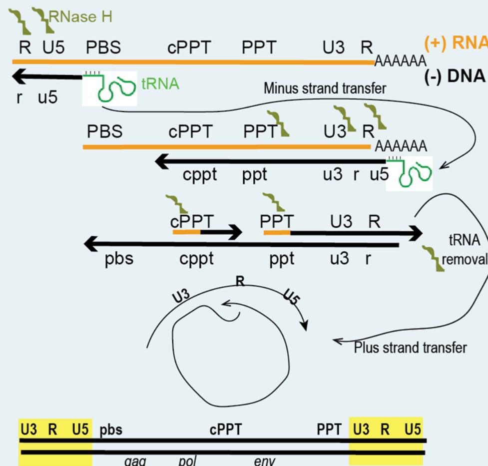
the minus strand is generated first, followed by two mechanistic replicative switches that join and duplicate the (+) stranded genomic RNA sequences. **Fig. 7B** depicts a summary of these steps. Following completion of DNA synthesis, the resulting linear double-stranded viral DNA is translocated in the host cell nucleus, where it is integrated into the host genome by integrase (now becomes provirus or proviral DNA).

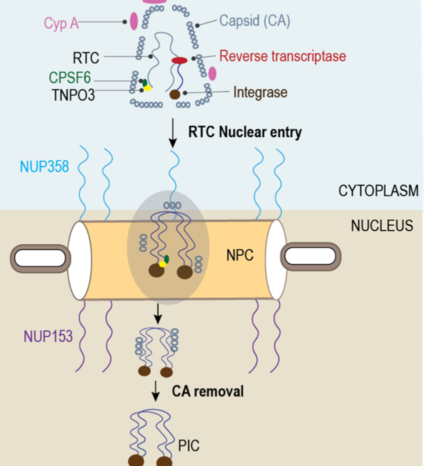
FDA has approved two classes of antiretroviral drugs that target RT polymerase activity: **1**) Nucleoside/nucleotide analogue RT inhibitors (NRTIs), which are chain terminators (lack the 3′–OH) or structural diverse analogues of the natural substrates of DNA synthesis (dNTPs) FDA approved NRTIs are emtricitabine (FTC), lamivudine(3TC), zidovudine (AZT), didanosine(ddI), tenofovir (TDF), stavudine(d4T) and abacavir. **2**) Non-nucleoside reverse transcriptase inhibitors(NNRTIs) which attach a hydrophobic pocket next to the active site of polymerase rendering the RT enzyme inactive[56]. NNRTIs that have been approved by FDA include rilpivirine (RPV), etravirine (ETR), and doravirine (DOR) (3rd generation); efavirenz (EFV) (2nd generation); nevirapine (NVP) and delavirdine (DLV) (1st generation).

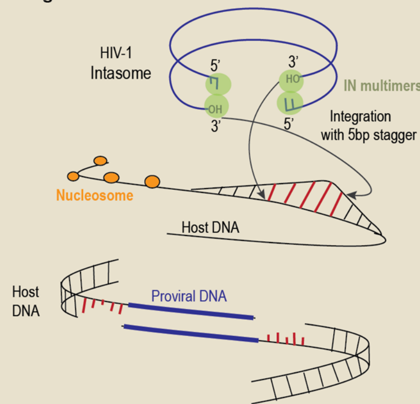
Uncoating and nuclear entry: Because the whole HIV-1 cone-shaped mature capsid (CA) core is too big to penetrate the nuclear pore, it must detach from the rest of the reverse transcription complex (RTC), a process known as uncoating (reviewed in Ref [59]). The CA first interacts with the host factor cyclophilin A (CypA) to promote correct uncoating kinetics while also protecting the viral DNA genome until nuclear entry[60]. The host factor cleavage and polyadenylation specificity factor 6 (CPS6) subsequently attach to the RTC, acting as a shuttle between the cytoplasm and the nucleus[61]. Transportin 3 (TPNO3) generally transports CFSP6 to the nucleus, implying a role in CA uncoating and allowing the CA to connect directly to the nuclear pore complex (NUP153, NUP358)[62]. The interaction of these host proteins promotes CA uncoating and the transport of the pre-integration complex (PIC) through nucleus pores before viral DNA integration into host genomic DNA[63] (**Fig. 7C**).

Although peptide-based antivirals targeting the CA's several binding sites have been developed, none have been authorised for human use (reviewed in Ref[64]).









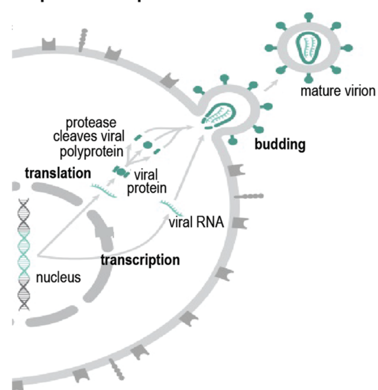


Figure 7. HIV-1 replication cycle.

A. A schematic diagram of HIV-1 entry, binding and fusion. After the interaction of the HIV-1 Envelope with the host receptors, Gp120 and Gp41 undergo conformational changes resulting in the exposure of HR1 and HR2 helices of Gp41 and the formation of a pre-hairpin intermediate. The HR2 helix loops back and interacts with the groove between HR1 helices forming a hex-helical bundle (6HB), bringing the viral and cellular membranes to close proximity for fusion. **B**. Process of HIV-1 reverse transcription: 1. tRNA-

primed (-) DNA synthesis; 2. (-) Strand DNA transfer; 3. PPT-primed (+) DNA synthesis; 4. (+) stand DNA transfer; 5. Bidirectional DNA synthesis. **C**. HIV-1 uncoating is linked to reverse transcription and requires host cell factors CypA which protects the viral DNA genome until nuclear entry, which is facilitated by CA binding of host cell proteins TNPO3, CPSF6, and components of the nuclear pore complex (NUP153 and NUP358).**D**. Depiction of the integration of HIV proviral DNA into the host genome.E. Late phase of HIV-1 replication Abbreviations HR1, heptad repeat 1; HR2, heptad repeat 2; PPT, polypurine tract; cPPT, central polypurine tract, PBS, primer binding site PIC, pre-integration complex; NPC, nuclear pore complex; NUP153, nucleoporin 153; NUP358, nucleoporin 358; RTC, reverse transcription complex. Figure adopted with permission from Refs.[65][66][59, 67] and E is By Thomas Splettstoesser [CC BY-SA 3.0], via Wikimedia Commons. A, D are open acess CC 4.0.

Integration: One of the most important properties of retroviruses is the incorporation of viral DNA into the host genome. This process is dependent on the enzymatic activity of viral integrase (IN) and involves many host cellular factors that influence the integration site selection (reviewed in Ref [68]). Integrase is encoded by the *pol* gene and is consists of three structural domains, including CTD and NTD domains, which are responsible for coordinating the interaction of DNA and chromatin binding, and the CCD domain, which is required for catalysis of 3' ends of the HIV-1 DNA (reviewed in Ref [67], **Fig. 2**).

Following PIC nuclear import at the NPC, HIV-1 IN multimers are positioned at HIV-1-LTR (the ends of the HIV-1 DNA) generating an intasome. IN then cleaves two nucleotides (T, G) from the 3' of each viral DNA resulting in staggered ends with a chemically reactive hydroxyl group. CTDs bind in the expanded major groove of the host cell/ target DNA to create a target capture complex (TCC)[69]. The TCC subsequently cuts five nucleotides of the host DNA resulting in a strand transfer complex (reviewed in Ref [70]). IN attaches to the target DNA using the staggering 3'-OH ends. The complementary strands of viral DNA connect with the 5'phosphates of the target DNA, allowing the HIV-1 DNA to be transferred to host cell DNA(reviewed in Ref [67]). IN duplicates the five base pairs of the host sequences directly flanking provirus's 5' and 3' ends (viral DNA integrated into host DNA, **Fig. 7D**). The proviral genome can be transcribed, translated and assembled into new virions (Late phase of viral replication). However, the integrated genome can also undergo transcriptional silencing resulting in the formation of a latent viral reservoir, a key impediment to HIV eradication[71]. Integration does not happen at random in the host genome (reviewed in Ref [67]). However, it has been demonstrated that effective integration must take place at certain sites, i.e. 1) target regions that have a high gene density, are transcriptionally

active, and highly spliced e.g. BACH2, STATSB, MKL2 genes of T-cells; 2) near the NPC (nuclear membrane)[72].

Several drugs targeting integration have been developed (reviewed in Ref [68]), however, the FDA has approved Raltegravir (RAL) and Dolutegavir (DTG) as integrase strand transfer inhibitors (INSTIs) for use in humans. In most African countries, one INSTI in combination with two NRTIs is currently used to treat HIV-1.

The late phase of replication (from gene expression to maturation) (**Fig.** 7**E**)

Transcription: The viral promoter located in the U3 region of the 5' LTR, initiates proviral transcription, with the integrated virus as a template. The viral transactivator protein (Tat) attaches to the trans-acting response (TAR) element, which is a specific sequence in the R region of the 5' LTR that promotes transcriptional processivity and RNA elongation. The cellular factor pTEFb is required for this effect. Tat enables the efficient synthesis of full-length HIV transcripts, and alternative splicing generates over 25 distinct mRNAs in three size classes. Unspliced RNA (9 kb) functioning as genomic RNA or Gag and Gag-Pol precursors, single spliced (4 kb) RNA encoding Vif, Vpr, Vpu, and Env, and fully spliced (2 kb) RNA expressing Tat, Rev, and Nef are examples. The viral Rev protein, which interacts with the Rev responsive element (RRE) in the viral RNA and the cellular export factor Crm1 to connect these viral RNAs to the export machinery, transports unspliced and partially spliced mRNAs from the nucleus to the cytoplasm.

Translation and assembly: Tat, Rev, and Nef proteins are produced fully spliced viral RNAs. The Nef protein inhibits CD4 and MHC class I surface receptor molecules, making infected cells less detectable to immune systems. Slowly, Nef manipulates cells to make them more efficient producers of fully infectious viral particles. The synthesis of Tat and Rev results in the production of full-length unspliced mRNA expressing the Gag and Gag-Pol precursors, which are then processed to important structural and enzymatic proteins. Simultaneously, the Vif, Vpr, Vpu, and Env proteins are produced from single-spliced viral RNAs. Assembly is a sequential process that begins with Gag-Pol precursors, then moves on to the Env glycoproteins, and last, the two copies of viral RNA are recruited. The accumulation of viral proteins and RNA at the plasma

membrane causes the development of a membrane-coated spherical particle (reviewed in Ref [73]).

Budding. The late domain of Gag's P6 part, and the cellular Tsg101 protein both play roles in the discharge of offspring virions from infected cells into the circulation. Notably, this late stage of the viral replication cycle is also targeted by a restriction factor: tetherin (BST-2) tethers mature and infectious viruses to the cell surface and is inhibited by the HIV-1 Vpu and other primate lentiviruses Nef or Env proteins (reviewed in Ref [74]).

Maturation: The HIV-1 particles are released in an immature and non-infectious form as Gag and Gag-Pol precursors. During or shortly after budding, the viral protease becomes activated and cleaves the Gag and Gag-Pol precursors into their mature final components. As a consequence, the configuration of the proteins is reorganized to generate the conical inner core and to render the virus infectious (reviewed in Ref. [73]).

Drugs that block this last step of the viral life cycle by inhibiting the viral protease are a component of effective ART. While HIV can still replicate in the presence of protease inhibitors, the resulting virions are immature and unable to infect new cells. These drugs include: Atazanavir (ATV), Darunavir (DRV), lopinavir (LPV)/ritonavir (RTV).

HIV-1 recombination

Studies suggest that HIV-1 genetic recombination results from the alternate use of two RNA templates during reverse transcription i.e. during the minus or plus-strand DNA synthesis (reviewed in Ref [75]). When a $CD4^+$ T-cell is infected with two strains of HIV-1 at the same time, the RNA transcript from each of these strains can be incorporated into a single heterozygous virion. When this virion subsequently infects a new cell and template switching happens during reverse transcription, a recombinant retroviral DNA sequence is produced, and all subsequent progeny virions have this recombinant genotype[76]. **Fig. 8A** depicts HIV-1 recombination between subtype G and CRF 02AG, which results in CRF43 02AG, as shown in **Fig. 8B**.

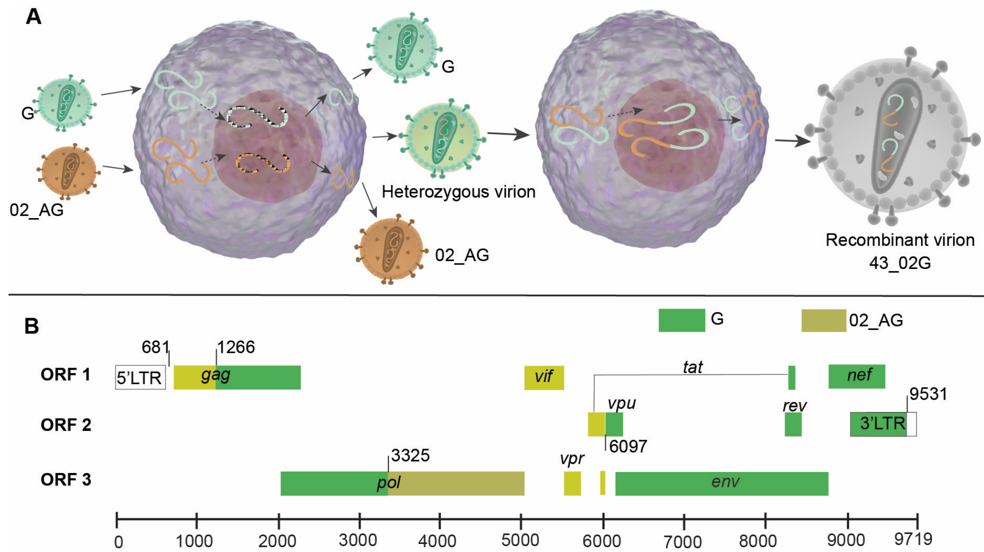


Figure 8. HIV-1 recombination.
A. Mechanism of HIV-1 recombination. **B** is a schematic representation of the HIV-1 circulating recombinant form CRF43_02G contains genetic segments from HIV subtypes G (the darker, green segments) and CRF02_AG (the lighter, army-green segments). These recombination breakpoints can be detected using phylogenetic software when sufficient sequence data is available, as indicated in the Los Alamos CRF Database (http://www.hiv.lanl.gov/).

HIV-1 immune responses

The human immune system is divided into two parts: the innate and adaptive immune systems [77]. When a foreign pathogen enters the body, immune cells detect the infection and initiate responses that kill as well as develop immunologic memory of the infection. The first line of defence, the innate immune cells, will initiate, amplify and direct subsequent long-lasting adaptive responses (provided by T lymphocytes) via a variety of contact-dependent signals with the infected cells as well as cytokine-mediated signals (**Figs. 9, 10**).

Cellular components of innate immune responses to HIV-1

Because innate reactions lack immunologic memory, they typically remain constant no matter how frequently the antigen is presented[78]. Innate immune cells detect pathogen-associated molecular patterns (PAMPs) through pathogen recognition receptors (PRR). Specifically, HIV-1 nucleic acids found in the cytosol of infected

CD4[+] T-cell (cell-associated virus) and the endosomes of phagocytosing cells (monocytes (Mo), macrophages (MØ), and dendritic cells (DC))[79]. Following this contact, these immune cells release cytokine and interferons which createS an antiviral state that suppresses infection but also induces adaptive immune responses.

Dendritic cells

Dendritic cells (DCs) are haematopoietic cells that are found in all tissues, including blood and lymphoid organs[80]. In peripheral tissues, DCs are found in an immature stage specialized in the capture of antigens. DCs undergo a complex maturation process into antigen-presenting cells in response to microbes before migrating to the local draining lymph node [78, 81]. DCs can develop into either plasmacytoid DCs (pDCs), or conventional DCs(cDCs).

In pDCs, the cell-free virus is phagocytosed and then recognised by pattern recognition receptors Toll-like receptors, TL7 and TL9 (**Fig. 9**), which bind single-stranded RNA and DNA, respectively, resulting in massive production of Type I (α, β) and III (λ) interferons and proinflammatory cytokines[82, 83]. These interferons exert their antiviral effect by stimulating the expression of interferon-stimulated genes (ISGs), e.g. APOBEC3, TRIM5, tetherin that synergistically interferes with replication and inhibits different stages of the viral replication cycle[84, 85] It is important to note that because they express CD4/CCR5/CXCR4, pDCs can be targets of HIV-1 infection.

In addition to generating cytokines, cDCs present antigens on MHC-II class molecules on their surfaces that bind to the T-cell receptor in naïve T-cells triggering T-cells to respond to infection (**Fig. 9**). Furthermore, HIV-1 does not replicate in cDCs type 2 due to the highly active SAMHDI restriction factor (degrades dNTPs in cells). HIV-2 encodes the Vpx protein, which degrades SAMHDI, resulting in effective reverse transcription of HIV-2 in cDC2, which likely explains why HIV-2 is more effectively controlled than HIV-1. Activated cDCs release pro-inflammatory cytokines such as IL-1, IL-6, IL-12, IL-23 and TNF (functions are summarised in **table 3**)[81].

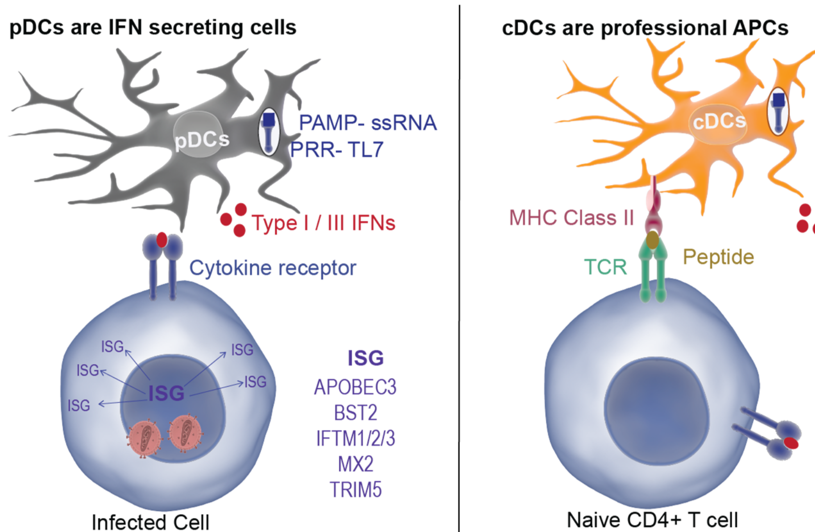


Figure 9. Dendritic cells

The resulting IFN response from pDCs induces the expression of ISGs like MX2 that prevents viral DNA nuclear entry. rhTRIM5α and TRIMCyp accelerate uncoating, where premature uncoating leads to reduced reverse transcription and little nuclear import of viral DNA. Abbreviations: ISG, interferon-stimulated genes; pDCs, plasmacytoid dendritic cells; cDCs, conventional dendritic cells; TCR, T-Cell receptor, MHC, major histocompatibility complex; PAMP; pathogen-associated molecular patterns; PRR, pathogen recognition receptor; TLR-7, Toll-like receptor 7; APOBEC3G, apolipoprotein B mRNA editing enzyme, catalytic subunit 3G; BST; Tetherin or bone marrow stromal antigen 2; IFITM1/2/3; Interferon Induced Transmembrane Protein 1/2/3; MX1, murine myxovirus resistance 1; TRIM, tripartite motif.

Natural killer cells

Natural killer cells are innate lymphocytes that kill infected cells and mediate immune regulatory function through cytokine production[86]. NK cells recognise infected cells either through interactions with an intricate network of inhibitory and activating receptors for MHC class I molecules, classical co-stimulatory ligands, and cytokines or via interaction with antibodies bound to the target cell surface (also known as antibody-dependent cellular cytotoxicity (ADCC))[87]. Healthy NK cells express specific MHC class 1 molecules that bind to the inhibitory receptor of NK cells, preventing them from being destroyed. MHC-1 is downregulated in infected cells, resulting in the activation of NK cells (**Fig. 10**).

NK cells produce cytotoxic granules, which destroy the infected cell[88]. NK cells can also identify infected cells that express antigens on their surface that antibodies can

attach to. The FC part of the antibody will then attach to the CD16 receptor on NK cells, causing ADCC and cell death.

There are 14 different polymorphic killer cell immunoglobulin-like receptors (KIRs) receptors on the surface of NK cells that bind to the HLA class I molecules of the target cells ($CD8^+$ T-cells), determining the threshold of NK activation and thus, potential viral clearance. This is an important complementary mechanism to $CD8^+$ T-cell responses that depend on MHC-class-I-recognition[86, 89-91]. After activation, NK cells release various cytokines including interferon-γ (IFN-γ), tumour necrosis factor-α (TNF-α), granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-1, IL-12, 1L21, 1L-15, IL-18, IL-1 and chemokines (CCL1, CCL2, macrophage inflammatory protein (MIP)-1α (CCL3), MIP-1β (CCL4), CCL5, and CXCL8) that can modulate the function of other innate and adaptive immune cells[91, 92].

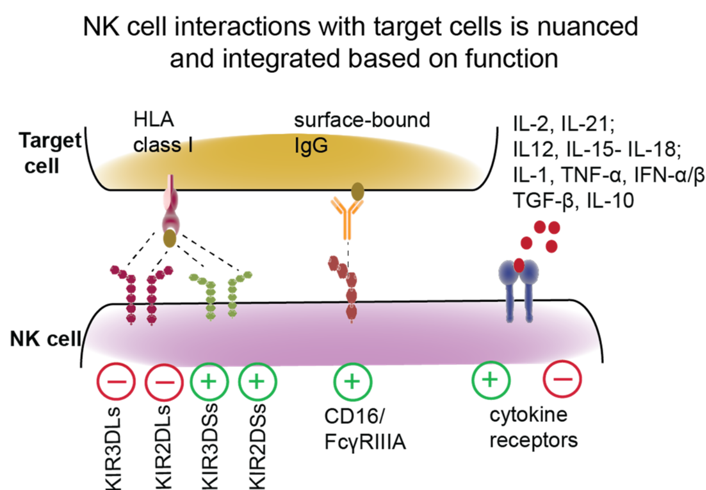


Figure 10. NK cell interactions with some of the inhibitory and activating receptors. Abbreviations: HLA, human leukocyte antigen; IL, interleukin; TNF, tumour necrosis factor; IFN, interferon; TGF, transforming growth factor; KIR3DLs, Killer Cell Immunoglobulin Like Receptor, Three Ig Domains and Long Cytoplasmic Tail; KIR3DS, Killer Cell Immunoglobulin Like Receptor, Three Ig Domains and Short Cytoplasmic Tail; FcγIIIA, Fragment crystallisable gamma receptor III A.

Macrophages

Macrophages (MØs) are terminally differentiated, non-dividing cells that are generated from circulating or blood-borne monocytes[78]. They are a unique population of phagocytes found in lungs (alveolar MØ), brain (microglia), liver (Kupffer cells) that

phagocytose pathogens and act as APCs[93]. HIV-1 uninfected MØs trigger antibody specific responses from B cells by MHC-II antigen presentation to $CD4^+$ T-cells, and they also aid to optimize the anti-HIV $CD8^+$ cytotoxic T-cells (CTL) responses by cross-presentation of HIV antigens via MHC-I thereby contributing to HIV-1 control *in vivo*[94, 95]. IFN-γ secreted by helper T-cells can boost MØ activity and they are also capable of chemotaxis (a process of being attracted and displaced to a particular location by specific molecules – chemokines)[96].

When exposed to inflammatory stimuli, MØs release cytokines such as tumour necrosis factor (TNF), IL-1, IL-6, IL-8, and IL-12; which can induce vascular permeability, recruitment of inflammatory cells, and production of acute inflammatory response proteins[96]. Because of their ability to infiltrate various organs including the brain, MØs have been identified as key cells responsible for mother-to-child-transmission of HIV-1 via breast milk, as well as contributing to the spread of HIV-1 within patients[97].

Soluble factors in innate immune responses to HIV-1

Cytokines

This is a large set of proteins or cell signalling messengers that aid the initiation of an immune response to an infection. Cytokines are 1) primarily produced by macrophages and lymphocytes; 2) functionally redundant (share similar functions); 3) pleiotropic (act on many different cell types- IFN); 4) paracrine (if the action is directed to a neighbouring cell but is in the place of release – IL-6); 5) autocrine (if the action is on the cell that secretes it, e.g. TNF from macrophages is an autocrine regulator of TLR-induced inflammatory signalling[98]); and 6) endocrine (if it reaches remote regions of the body-transforming growth factor-beta (TGF-β), and monocyte colony-stimulating factor (M-CSF))[96]. Each cytokine binds to a specific cell surface receptor, triggering the production of other cytokines or initiating a cell signalling cascade that can either negatively or positively regulate many genes and their transcription factors, affecting cell function (**Table 3**).

Table 3. Cytokines produced by innate immune cells.

Name	Type	T-cell stimulated	Function
Cytokines produced by macrophages[96]			
TNF	Proinflammatory		Elevates the synthesis of C-reactive protein and other mediators
IL-1β	Proinflammatory		Stimulates the production of acute-phase proteins from the liver
IL-6	Proinflammatory/ anti-inflammatory	CTL	Promotes differentiation of B cells into plasma cells, activates CTL
Il-12	Proinflammatory	Th-1	Synergises with TNF in stimulating IFN-γ production
Il-18	Proinflammatory		Inducer of IFN-γ production Synergises with IL-12 to activate T-cells, NK cells
IL-23	Proinflammatory		IFN-γ inducer and T-cell activator, IL-23 augments IL-10 release and induces IL-17 synthesis by activated naïve T-cells
IL-27	Proinflammatory/ anti-inflammatory	Th-1	Inhibits the differentiation of Th-17 cells
IL-10	Anti-inflammatory		Suppression of macrophage activation and production of TNF, IL-1β, IL-6, IL-8, IL-12, and GM-CSF
TGF-β	Anti-inflammatory		Potent suppressor of both Th-1 and Th-2 cells Tones down inflammatory effects of TNF, IL-1β, IL-2, IL-12
Cytokines produced by dendritic cells[81]			
TNF-α	Proinflammatory		elevating the synthesis of C-reactive protein and other mediators
IL-1	Proinflammatory		stimulates the production of acute-phase proteins from the liver
IL-6	Proinflammatory/ anti-inflammatory	CTL	promotes differentiation of B cells into plasma cells, activates CTL
Il-12	Proinflammatory	Th-1	It synergizes with TNF in stimulating IFN-γ production
IFN-α	Proinflammatory		T and B cell differentiation
IFN-β	Proinflammatory		

Abbreviations: IL, interleukin; TNF, tumour necrosis factor; IFN, interferon; GM-CSF, granulocyte-macrophage colony-stimulating factor; Th-2 , helper T-lymhocytes; CTL, cytotoxic $CD8^+$ T-cells.

Chemokines

Chemokines are small chemotactic cytokines with molecular weights around 8–10 kDa that regulate immune cell migration[99]. Chemokines direct immune cells to sites of need, offer signals for cell differentiation and stimulate angiogenesis, cell growth and death [100]. Proinflammatory cytokines such as TNF, IL-6, and IL-1β frequently stimulate chemokine release[96]. Chemokines are currently classified into four families based on the arrangement of cysteine residues, i.e. 1) **CC** (first two cysteine residues are adjacent to each other); 2) **CXC** (first two cysteine residues have an amino acid between them), **XC** (lacks cysteine one), and **CX_3C** (exhibits three amino acids between the first two cysteines), which is followed by L (for ligand) and then a number[100]. Chemokines work by activating specific G protein-coupled receptors causing inflammatory and non-inflammatory cells to migrate[101]. About 40 chemokines and 20

chemokine receptors have now been identified in humans[99]. Along with immune activation after HIV infection, a wide spectrum of chemokines and other cytokines are elevated, resulting in the so-called "cytokine storm," which begins within 1-3 days of hyperacute HIV-1 infection[102, 103]. **Table 4** is a list of chemokines and chemokine receptors that have been linked to HIV-1 replication and disease progression.

Table 4. Chemokines, their receptors and roles.

Chemokines released by macrophages[96, 104]			
Chemokine	Other names	Receptor[99]	HIV-1 role
CXCL1	MIP-2α	CXCR2	Recruit neutrophils and hematopoietic stem cells
CXCL2	MIP-2α	CXCR2	
CXCL8	IL8	CXCR1, CXCR2	
CXCL9	MIG	CXCR3	Strong T-cell chemoattractant to the site of inflammation, mediates cell recruitment necessary for inflammation and repair of tissue damage.
CXCL10	IP-10	CXCR3	Inflammatory, attracts T-cells, NK cells, dendritic cells (DCs)significantly elevated during the early stages of AHI, and associated with disease progression[105, 106]
CXCL11	IP-9, I-TAC	CXCR3	Mediates T-cell recruitment, more potently than CXCL9 and CXCL10
CCL5	RANTES	CCR1, CCR3, CCR5	Inflammatory chemoattractant for T-cells, basophils, eosinophils, and dendritic cells to the site of inflammation
Chemokines released by secondary lymphoid tissue[99, 104]			
CCL19	ELC, MIP-3β	CCR7	T-cell and DC homing to LN, Elevated during AHI, chronic infection even on treatment
CCL21	SLC	CCR7	Regulate the homing of naïve and central memory T-cells
CCL20	MIP-3α	CCR6	Has a strong chemotactic effect on immature dendritic cells (DCs) and lymphocytes, CCL20 was upregulated during inflammation; Th-17 responses; B cell and DC homing to gut-associated lymphoid tissue
CCL2	MCP-1	CCR2	Inflammatory monocyte trafficking
CCL4	MIP-1β	CCR5	T-Cell – DC interactions
CCL11	Eotaxin-1	CCR3	Eosinophil and basophil migration
CCL26	Eotaxin-3	CCR3, CX3CR1	Eosinophil and basophil migration
CCL17	TARC	CCR4	Th-2 responses, Th-2 cell migration, Treg, lung and skin-homing
CCL3	MIP-1α	CCR1, CCR5	Macrophage and NK cell migration
CCL22	MDC	CCR4	Th-2 response, Th-2 cell migration, Treg migration
CCL13	MCP-4	CCR2, CCR3, CCR5	Th-2 response

Other soluble factors involved in the innate responses include 1) complement – which generates immunologically active substances such as C3b, which enhances phagocytosis and then C3a, C4a, C5a, which cause release inflammatory mediators, and then membrane attack complex (C5b, C6, C8, C9), which leads to target cell death via cell membrane perforation[77]. 2) Acute-phase proteins (C-reactive protein, serum amyloid A protein, proteinase inhibitors, and coagulation proteins), whose plasma levels rapidly increase in response to infection and inflammation, and whose role is to enhance resistance to infection and promote repair of damaged tissues[77].

Adaptive immune responses

T lymphocytes, which mediate cellular immunity, and B lymphocytes, which mediate humoral immunity, provide adaptive immunity, which works in tandem with the innate immune system. Through the T-cell receptor (TCR), T-cells recognise HIV-1 short peptides presented by the MHC molecule on the antigen-presenting cells (APCs) (DCs, activated macrophages, B cells). MHC molecules are classified into two types: class I and class II. HLA-A, B, and C are the three main types of class I molecules, whereas HLA-DP, DQ, and DR are the three main types of class II molecules[77]. T-cells that express "cluster of differentiation" CD4 and CD8 molecules on the surface have been discovered as being critical in immune responses and T-cell development. CD4 binds to MHC class II molecule, whereas $CD8^+$ binds to MHC class I molecule. $CD4^+$ T-cells typically function as helper T-cells, recognising antigens presented by MHC class II molecules, whereas $CD8^+$ T-cells are usually cytotoxic, recognising antigens presented by MHC class I molecules.

$CD4^+$ T-cell responses to HIV-1

When the HIV-1 peptides are recognised, the naïve $CD4^+$ T-cells are activated, resulting in proliferation and differentiation into specialised effector cells. This lineage-specific differentiation is influenced by the cytokine levels in the priming environment. The effector $CD4^+$ helper T-cells migrate to infected tissue characterised by their cytokine-producing potential. Furthermore, they are separated into different cell subsets based on the type of assistance they provide (summarised in **Fig. 11**). These $CD4^+$ T-cells appear to be especially critical in HIV-1 infection for maintaining and

encouraging the formation of effector and memory CD8[+] T-cell populations by augmenting APC-mediated production of cytokines (IL-1, IL-6, TNF, IL-15) [107, 108].

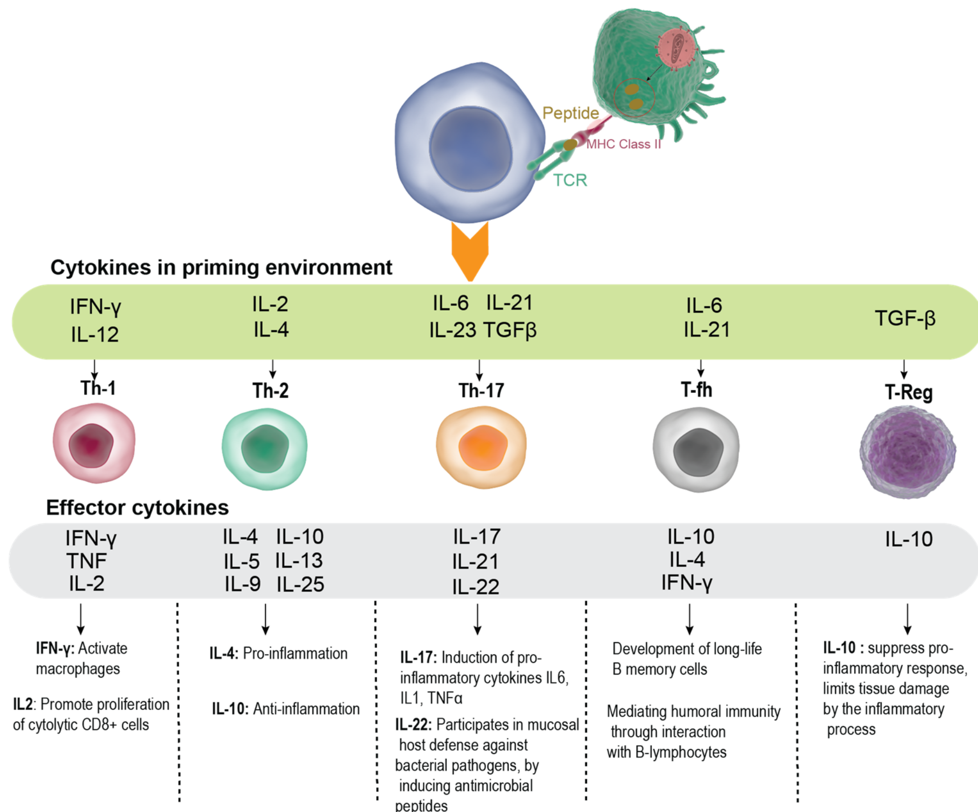


Figure 11. Antiviral effector CD4[+] T-cells and functions of cytokines produced.
The steps in generating primary antiviral T cell responses include 1) uptake of viral antigens by antigen-presenting cells (APCs) in infected tissue; 2) recognition of antigens on activated APCs by naïve T-cells during viral infection predominately results in the generation of Th-1 cells owing to the presence of type I interferons (IFNs) and interleukin-12 (IL-12). However, Th-17, Th-2 and regulatory T (TReg), follicular T (Tfh) cell populations are also generated to some degree. Abbreviations: TCR, T cell receptor; TGFβ, transforming growth factor-β; TNF, tumour necrosis factor; IL, Interleukin.

CD8[+] T-cells

During HIV-1 infection, CD8[+] T-cells recognise infected cells (are activated) via three signals provided by the APCs: 1) TCR recognition of the HIV- antigen peptide (8-10 amino acids) presented by the MHC class I molecules; 2) binding of a co-stimulatory molecule (CD80 or CD86) from APCs to CD28 molecule on T-cell; and 3) cytokine cell signalling in response to HIV-1 such as IL-12 and IFN[109]. CD8[+] T-cells undergo

clonal expansion, differentiation, generation of numerous effector cells, and migration to the peripheral tissues in response to recognition/activation[110].

The effector mechanisms of CD8 T-cells can be classified as: 1) lytic (cytolytic) – this depends on perforin and granzymes from lytic granules and the degranulation capacity of the $CD8^+$ T-cells that collaborate to induce the death of target cells[111]; and 2) non-lytic (cytokine and chemokine production) – cytokines secreted by $CD8^+$ T-cells include IFN-γ (antiviral), TNF-α (pro-inflammatory), IL-2 T-cell survival. Chemokines include macrophage inflammatory protein (MIP)-1α (CCL3), MIP-1β (CCL4), and regulated upon activation, normal T-cell expressed and secreted (RANTES; CCL5), which bind to chemokine receptors and prevent HIV-1 binding and entry[112]. When $CD8^+$ T-cells are activated, they differentiate into several phenotypes: Effector, effector memory, central memory and naïve. And several subpopulations have been described based on the cytokine they produce, namely: (i) Th1-like cytokine pattern (Tc1) cells have a strong cytotoxic function and produce high levels of IFN-g and TNF-α; (ii) Th2-like cytokine pattern (Tc2) cells produce IL-4, IL-5, IL-6, and IL-10 and have the lower cytotoxic ability; and (iii) Tc17 cells produce high levels of IL-17 but no granzyme (reviewed in Ref [113]).

Several studies have shown that $CD8^+$ T-cells play an important role in the control of the virus as evidenced by (i) a decline of viremia only after the emergence of HIV-specific $CD8^+$ T-cells and an increase in viral load after $CD8^+$ T-cell depletion[114-116]; (ii) associations between the frequency and/or functional capacity of HIV-specific $CD8^+$ T-cells and limited viral replication and/or disease non-progression in HIV-infected patients[117, 118]; (iii) emergence of viral escape mutations to evade the immunological pressure exerted by HIV-specific $CD8^+$ T-cells [119]; iv) certain MHC or HLA class I molecules presenting peptides, recognized by $CD8^+$ T-cells, such as HLA-B57 and HLA-B27 are associated with viral control, whereas HLA-B35 is associated with rapid progression to AIDS[120].

HIV-1 has devised several mechanisms to evade or escape the $CD8^+$ T-cell response, including 1) at the mucosal portal of entry, where early HIV-1-specific $CD8^+$ T-cell responses have been shown to appear too late and insufficient to control viral

replication[114]; 2) Because of HIV's high mutation rate, the virus has been able to develop escape mutations in CTL epitopes early in infection, which prevent $CD8^+$ T-cell recognition[121]; 3) The ability to downregulate MHC-I (HLA class I A and B alleles) expression from infected cells, hence reducing recognition of infected cells by $CD8^+$ T-cells; 4) HIV-1 also impairs normal $CD8^+$ T-cell signalling by changing the pattern of cytokine production and cellular receptor engagement. As a result of the incorrect T-cell receptor (TcR) stimulation, these cells become anergic; 5) HIV can reduce the circulating pool of effector and memory $CD8^+$ T-cells that are capable of combating viral infection by altering the activity of $CD4^+$ T-cells and APCs that are essential for appropriate $CD8^+$ T-cell maturation. As a result, $CD8^+$ T-cell function is disrupted; 6) $CD8^+$ T-cells become exhausted after being exposed to a large number of HIV antigens during chronic HIV infection. They produce immunological checkpoint proteins, which degrade function and diminish survival capacity[122]; and finally, 7) $CD4^+$ T-cells harbouring a defective provirus, can still express some viral proteins diverting $CD8^+$ T-cells and contributing to their exhaustion.

Humoral responses

This includes immunological responses including B-cells, antibodies, and especially broadly neutralising antibodies (bNAbs). To prevent infection of target cells, antibodies can neutralise the virus by blocking $CD4^+$ molecule binding to the virus's Gp160 spike or preventing viral-plasma cell membrane fusion (reviewed in Ref [123]). However, once a cell is infected, antibodies can bind to its surface and recruit NK cells via its Fc component, resulting in cell death. Around 2-3 years after HIV-1 infection, approximately 20% of infected individuals develop cross-reactive neutralising antibodies (capable of recognising thousands of different viruses) and approximately 1% develop broadly neutralising antibodies, which have been shown to protect macaques from SIV infection (reviewed in Ref[124]). Several viro-epidemiological covariates, including viral diversity, period of infection, viral load, and ethnicity, have been linked to the development of bNAbs (reviewed in Ref [125]).

Several variables have contributed to the difficulties in producing bNAbs, but the primary reason is the direct and indirect effect of HIV on B-cell responsiveness. Specifically, HIV targets follicular helper T-cells, which play critical roles in the

formation of B cells. Furthermore, the virus can avoid antibody detection by using a glycan shield on gp160, producing non-functional spikes, and expressing Gp160 at a low density (reviewed in Ref[126]).

Immune responses in infants

Neonatal dendritic cells and monocytes are less polyfunctional and produce fewer type-1 IFNs than adult cells[127]. In infant innate immune cells, IL-12 and IFN-γ (Th1-polarizing cytokines) are produced at a lower rate than IL-10, IL-6, and IL-23 (Th2- and Th17-inducing cytokines)[127] (reviewed in Ref[128]). This immune response protects against extracellular bacterial and fungal pathogens while avoiding potentially harmful pro-inflammatory Th-1 cell responses, albeit at the expense of efficient antiviral immune responses[128][129].

Infant NK cells have similar numbers to adults, but differ in the expression patterns of inhibitory and activating cell surface receptors; have lower cytotoxic activity against HIV-1 infected cells as reported in cord blood; and have lower ADCC when compared to adult NK cells [130-133]. Neonatal NK cells, on the other hand, appear to be more successful than adult NK cells at suppressing the replication of CCR5-tropic HIV-1 strains[134].

HIV-specific $CD8^+$ T-cell responses can be detected from birth, although they are initially insufficient to control viremia, in contrast to the decline in viral load after acute infection observed in adults[135-137]. Furthermore, because an infant inherits 50% of the HLA alleles from the mother, the transmitted virus can be partially pre-adapted to the shared HLA alleles, preventing the recipient child from mounting an effective immune response against the escaped epitopes and failing to contain viral replication[138]. Nef and Env epitopes of HIV-1 are the primary targets of $CD8^+$ T-cells in infants with a rising proportion of Gag-specific responses only later in life [139, 140]. Furthermore, the specificity, magnitude and effectiveness of a $CD8^+$ T-cell response to control viral replication are determined by its quality, as measured by the ability to degranulate and produce different effector cytokines at the same time, but this effect is reduced in infants younger than two years [117, 140, 141].

The relatively low levels of $CD4^+$ T-cell responses detected in most infants under one year is the most noticeable difference between HIV infection in infants and adults [136, 140]. However, $CD4^+$ T-cell responses detected at 3–6 months of age appear to be essential in the T-cell-mediated control of HIV[140]. There is a selective impairment of $CD4^+$ T-cell responses in early childhood with a shift from Th-1 toward Th-2 type responses [142].

HIV-1 Pathogenesis: Natural history of infection in adults

Early events in HIV-1 infection influence disease progression in such a way that a more robust viral control during AHI resulting in a lower viral set-point level, is associated with a slower disease progression[143]. In this section, I will discuss the early events in sexual transmission of HIV-1 from a host and viral standpoint (summarised in **Fig. 12**).

Mucosal barrier

An individual can become infected with HIV-1 mostly via mucosal surfaces during sexual intercourse (anal-penile or penile-vaginal) and from the mother (during gestation, labour or breastfeeding). Infection can also occur through injection using a tainted needle. Most global HIV-1 infections are transmitted through sexual contact, and the rate of transmission varies depending on the structural conformation of the genital mucosa (epithelium)[144, 145]. The anal-rectal monolayer epithelium has the highest probability of transmission (0.3-5 %) followed by a non-keratinised female (0.05-0.5%) and male (<0.1%) epithelium with oral mucosa (0.01%) being the least[146].

If micro-lacerations occur during sexual intercourse, HIV-1 (as a free virus or cell-associated virus) will come in contact with mono or pluristratified epithelium (of the vagina, endocervix, ectocervix, inner foreskin and urethra on the penis, rectal mucosa) that is in contact with the mucus[147]. It will cross the epithelium within 24 hours, reaching the lamina propria where it contacts target cells (cells expressing CD4 receptor and co-receptor CCR5, CXCR4 – pDCs, cDCs, monocyte-derived macrophages, Langerhans cells and $CD4^+$T-cells)[147, 148]. In the lamina propria, the R5 viruses (which use CC-chemokine receptor 5 – CCR5 coreceptors) form the infected founder

population, most likely because genital CD4+ T-cells exhibit high levels of CC-chemokine receptor 5 (CCR5) are rapidly infected by HIV-1[149].

Transmitted/founder virus

Following sexual intercourse, the donor who possesses a diverse viral population in the transmission fluid transmits only one or a few viruses to the recipient – referred to as the transmitted/founder virus, which creates the infected founder population[150]. Several viral characteristics describing the T/F have been established, including: i) fewer N-linked glycosylation sites in *env* of subtype A, C, D T/F viruses[151, 152]; ii) T/F viruses are more resistant to inhibition by interferon-α (IFN-α)[153, 154]; iii) T/F viruses have a higher env protein concentration, which increases the probability of viral attachment[153]; iv) enhanced binding of T/F viruses to dendritic cells, allowing for efficient transport from the epithelial surface to lymphoid tissues; and v) preference for CCR5 and CD4+ tropism. All these factors may have contributed to the transmission bottleneck found during AHI. In heterosexual transmission (penile-vaginal), approximately 80% of infections are established by a single T/F meaning that 20% of transmissions can be from more than one variant[155]. The same frequency has been observed in vertical or mother to child transmission[156]. However, in anal-penile transmission (MSM), the incidence of multiple transmitted founder viruses might reach 40%[157].

Eclipse phase of acute HIV-1 infection

Following the establishment of the T/F virus population, these viruses typically propagate locally but at un-detectable levels (50–100 copies of viral RNA per millilitre of plasma) for about 10 days (**Fig. 12B**). As a result of the innate immune response, acute-phase proteins, pro-inflammatory cytokines and chemokines are elevated in plasma during this viral expansion and before the onset of plasma viremia[106, 158]. IFN-α, IFN-γ, IL-1RA, IL-2, IL-6, IL-8, IL-12, IL-15, IL-18, IL-22, IP-10, TNF-α, and soluble CD14 are elevated beyond pre- infection levels during the eclipse phase in AHI leading to a term called cytokine storm and the level or speed of elevation of these cytokines vary[102, 159].

Systemic phase/peak viremia

At the end of the eclipse phase, the virus continues to disseminate the infection from the site of infection (lamina propria) to lymphatic drainage, establishing a self-propagating infection in the genital draining lymph nodes. The infection then spreads hematogenously to establish systemic infection through the secondary lymphatic tissues particularly the gut-associated lymphatic tissue-GALT, which contains a high number of activated $CD4^+CCR5^+$ memory T-cells; as well as in the spleen and peripheral lymph nodes[160]. The HIV-1 RNA become detectable in blood with PCR from 10-15 days following infection, also known as Fiebig stage I[161]. Fiebig stages are a classification system used to diagnose early HIV-1 infection based on test results for markers of virus and the immune response[161](**Fig. 12B**). Virus p24 antigens can be detected in plasma using ELISA as early as 15 days post-infection – Fiebig staging II.

While HIV-1 is replicating in the GALT and other secondary lymphoid tissues, plasma viremia will increase exponentially to peak, reaching viral loads of $>1 \times 10^8$ viral copies/ml of blood (6.7 log10 copies/ml) at around 21-28 days [162, 163]. During this time, 40–90% of the patients may present with symptoms of acute infection – acute retroviral syndrome (fever, headache, myalgia, fatigue, anorexia, pharyngitis, diarrhoea, night sweats, skin rash, lymphadenopathy and oral ulcers)[163-165]. In addition, approximately 80% of the $CD4^+$ T-cells in the GALT are depleted, resulting in low $CD4^+$ T-cell counts during peak viremia[160]. This reduction mostly affects activated $CD4^+$ effector memory T-cells that co-express CCR5[166]. ELISA or sensitive anti-HIV test kits can detect HIV-1 specific antibodies as early as 20 days after infection – Fiebig staging III. Some of the cytokines and chemokines (IL-6, IL-22, IP10) that were elevated at the eclipse phase begin to decrease back to pre-infection levels at peak viremia while others remain elevated (CXCL13)[102, 158].

Latent infection

HIV-1 must integrate into the genome of the cells it infects. The activity, longevity and fate of the integrated virus, are greatly influenced by the metabolic activation state and life-span of the infected host cell. In other words, $CD4^+$ T-cells can serve as both an active or latent viral reservoir. If integration occurs in an activated $CD4^+$ T-cells (which

have been in contact with APCs), viral replication is rapid, very effective and cytopathic with a half-life of 24 hours[167]. However, if integration occurs in a quiescent/resting $CD4^+$ T-cells (characterised by low metabolic, transcriptional and, translational activities, as well as a lack of spontaneous proliferation), the cells will contain the viral DNA but the viral expression is disrupted at the transcription level, resulting in the formation of a reservoir of latent HIV-1 infection, a significant barrier to HIV-1 cure or effective treatment[168]. This is because if ART is halted, infection from these latent viruses can return. A quiescent T-cell can have a half-life of 44 months[167]. By 21-28 days; the lymphatic tissue reservoir, where HIV is produced and persists in latently infected cells, has already been established[166, 169, 170].

Viral set point

After reaching peak viremia at 21-28 days, virus replication declines to a viral set point of about 30,000 copies per millilitre[163, 171]. Set point viral load (spVL) refers to the steady-state viremia that continues after the resolution of acute viremia is established at a median of 31 days (range 18-42 days) after the onset of viremia and is correlated to the peak viremia and downslope (**Fig. 12A**) [163]. The spVL is maintained by a balance between virus turnover and the immune responses, in the absence of antiretroviral drug therapy (ART). Virus-specific $CD8^+$ cytotoxic T-lymphocyte (CTL) activity and NK cells have been linked to the control of viremia in primary HIV-1 infection; and these CTL responses are preceded by increased innate NK cells before seroconversion (Fiebig stage III) or as viremia approaches its peak [171, 172]. However, under the pressure of CTL responses, numerous escape mutants are selected to evade $CD8^+$ T-cell recognition, and these are temporally associated with loss of immune control of infection.

Virus diversification occurs during this decrease in viral load (following a peak in CTL response). Indeed, at peak viremia (21-28 days), the founder virus population is homogenous (diversity is below 0.5%) demonstrating that there is no immune–driven escape mutations at peak viremia[155, 173]. However, under the strain of CTL responses, numerous escape mutants are selected to evade $CD8^+$ T-cell recognition, and these are temporally associated with loss of immune control of infection.

CD4⁺T-cell dynamics

During the natural history of HIV-1, the dynamics of CD4⁺T-cells go through three stages. The first stage, which occurs at peak viremia, is characterised by a largely reversible drop in CD4⁺ T-cells, which is coupled with GALT depletion of these cells. Following this, effector CD4⁺ T-cells are redistributed to the periphery. The second stage, an asymptomatic phase with no indications or symptoms of infection for several years to a decade or more, results in a modest but progressive drop in the CD4⁺ T-cell population, as well as continuous weak viral replication (**Fig. 12B**).

Finally, in the third stage, CD4⁺ T-cells rapidly decline, resulting in severe immune deficiency (acquired immune deficiency syndrome - AIDS) based on CD4⁺ T-cell levels below 200 copies/ml[174]. This is critical because the total number of CD4⁺ T-cells in a normal young adult is estimated to be 22×10^{11} cells[166]. This stage is characterized by the reactivation of several pathogens (fungi, parasites, CMV, EBV) as well as the appearance of tumour-associated oncogenic viruses. The typical survival time of an ART-naive HIV-1 positive adult is 7-12 years, and it usually takes six years or more to acquire AIDS, depending on age and host characteristics mentioned in the next chapter on disease progression[175, 176].

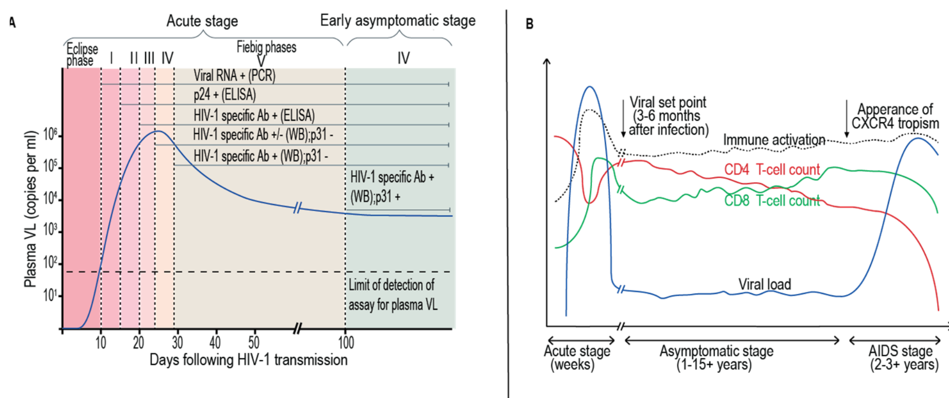


Figure 12. Adult HIV-1 pathogenesis, and acute HIV-1 viral load dynamics.
A. Relative levels of HIV RNA or viral load (blue), CD4⁺ T-cells (red), CD8⁺ T-cells (green), immune activation (dotted lines) and children (solid lines) in the years following the acquisition of HIV-1 **B**. In approximately 100 days following infection, the plasma viral load begins to plateau. These first 100 days after infection can be split into clinical stages based on a stepwise increase in positivity for the detection of HIV-1 antigens and HIV-1-specific antibodies in diagnostic assays (Fiebig stages I–VI are assigned to patients based on a sequential gain in positive HIV-1 clinical diagnostic assays (viral RNA measured by PCR, p24 and p31 viral antigens measured by enzyme-linked immunosorbent assay (ELISA), HIV-1-

specific antibody detected by ELISA and HIV-1-specific antibodies detected by western blot))[161]. The eclipse phase occurs between the moment of infection and the first detection of viral RNA in the plasma. Plasma virus levels then increase exponentially, peaking at 21–28 days after infection, followed by a slower decrease in plasma viral RNA levels. The figure was adopted with permission from Joakim Esbjörnsson and Ref. [159].T

HIV-1 Pathogenesis: Natural history of infection in infants

High maternal viral loads and placental inflammation are the main risk factors for HIV-1 transmission to her baby throughout pregnancy (*in utero*), labour (intrapartum), breast feeding (postpartum)[177-179]. As with sexual HIV-1 transmission, the rate of vertical transmission in an ART-naïve mother varies depending on how the infant got infected, with intrapartum (10–20%) being the most prevalent route, followed by postpartum (5–15%), and *in utero* the least efficient route (5–10%). When exposed to maternal secretions (breast milk) and blood during labour, birth and breastfeeding; HIV-1 will infiltrate the infant's upper gastrointestinal tract (GIT) and other mucosal surfaces. The upper GIT is lined by mono or pluristratified non-keratinized epithelium, a site of HIV-1 invasion and fetal intestines have the higher levels of $CD4^{+}CCR5^{+}$ T-cells than in cord (virtually absent) or neonatal blood[180, 181]. *In utero* transmission of HIV-1 typically occurs by infection of trophoblasts and transcytosis in the placenta, which generally occurs in the third trimester[182, 183]. However, this transmission is predominately cell-associated where maternal lymphocytes enter the trophoblasts, which are naturally non-permissive to cell-free HIV-1[184, 185].

As with adult sexual transmission, the virus population in an infected infant is typically more homogeneous than that of the transmitting mother, implying a transmission bottleneck. Approximately 75% of vertical transmissions are initiated by single T/F in [156, 186, 187]. Notably, infant T/F viruses uniformly represent a minor variant of the maternal viral population Env variants, indicating that maternal antibodies capable of preventing infant virus transmission also need to target minor circulating variants[156].

During the first month following HIV-1 acquisition, the plasma VL in an infected infant is below detection levels just like in adult HIV-1 infection. Plasma viremia in infants is about 5.5-6.5 log10 copies/ml depending on the time of HIV-1 acquisition, which is the same range as in adults[163, 188, 189]. At two months following infection, the

viremia was observed to be persistently high up to five years old when the set point of about 10,000 copies was achieved (**Fig. 13**)[190]. As previously stated, set point viral load in adults is normally achieved within two months following HIV-1 transmission. This difference in levels and timing of spVL in infants is because the early neonatal immune response establishes an immunological tolerogenic environment – that is anti-inflammatory (elevated IL-10, IL6, IL23) and thus avoiding proinflammatory responses (a feature of antiviral cell-mediated immunity in adults). In adults, the rapid decline in viral load during acute infection from several million copies/ml to a spVL 3 logs lower is temporally associated with the appearance of HIV-specific CD8$^+$ T-cells (reviewed in Ref [191]). The CD8$^+$ T-cell responses seem to be insufficient to control viremia, as there is typically no rapid decline in viral load in paediatric HIV infection as opposed to the decline in viral load after acute infection in adults[191]

In the absence of ART, >50% of HIV-1 infected children die by two years of age[192], whereas the median survival time is 11 years for ART-naive HIV-positive adults[193]. In conclusion, differences between infant and adult immune responses are not due to an infant's inability to respond (‘a baby immune system’) but rather reflect age-specific patterns of response. In other words, infants are not incapable of responding to stimuli. Instead, their responses are qualitatively and quantitatively different.

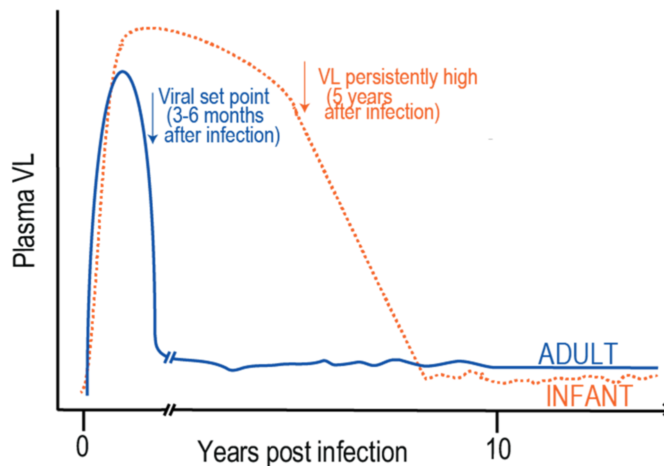


Figure 13. Infant HIV-1 viral load dynamics in comparison to adults.
A schematic representation of the relative plasma levels of HIV following adult infection (blue line) and paediatric infection (orange dotted line) in those children who progress slowly and maintain high CD4$^+$ T-cell counts despite initial high viraemia. Figure was adopted with permission from Ref.[129].

HIV disease progression

Why some people develop AIDS quickly while others remain healthy for years without therapy is a fundamental subject in HIV research. From an evolutionary perspective, there are viral factors that affect the virulence of infection and have an impact that lasts for years in infected individuals and across transmission events.

Classification of HIV-1 disease progression

Several terms have been coined to describe how a person responds to HIV-1 infection based on how the virus is controlled (measured by viral load) and the number or rate of $CD4^+$ T-cell depletion (measured by counting these cells). Based on viral control, patients are grouped into four categories, namely: 1) elite controllers (EC); 2) HIV controllers (HIC); 3) viremic controllers (VC); and 4) non-controllers (NC) (**Fig. 14.**). Based on $CD4^+$ T-cell counts or clinical-immunological classification, individuals are classified as: 1) long-term non-progressors (LNTP); 2) slow progressors (SP); and 3) rapid progressors (RP) (as specified in **Fig. 14**, and reviewed in Ref. [194])

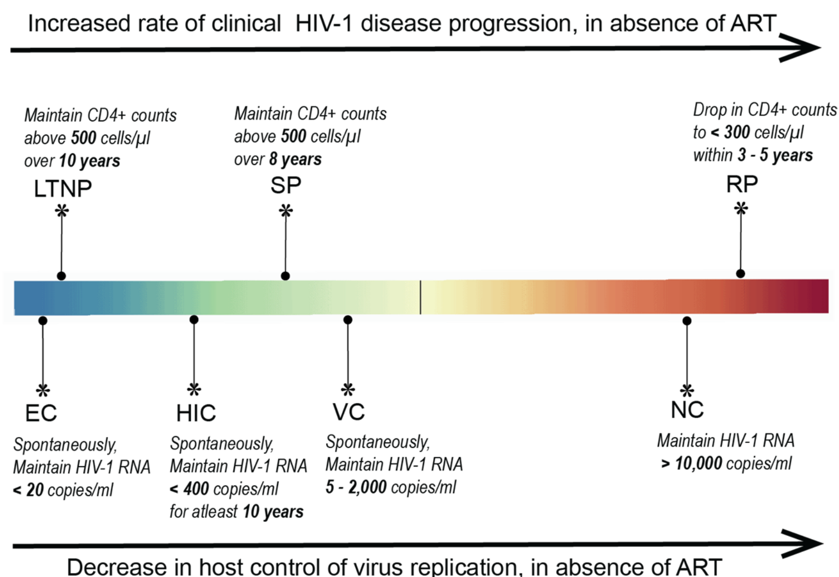


Figure 14. Classification of disease progression.
Terms used to classify ART naïve HIV-1 patients based on levels of $CD4^+$ counts (upper part) and viral load (lower part) over a specified period of time. Abbreviations: LNTP, long-term nonprogressors; SP, slow progressors; RP, rapid progressors, EC, Elite controller; HC, HIV controllers; VC, viremic controllers; NC, non-controller. Figure was adopted with permission from Ref[194].

Viral factors associated with HIV-1 disease progression

The density of HIV-1 in someone's blood is known as viral load, and it stabilises approximately 31 days after peak viremia to a spVL, which ranges from <20 to 10^6 virions/ml of blood plasma and is associated with virulence and infectiousness [163, 195, 196]. Several viral factors have been linked to differences in spVL amongst patients (affecting the rate of disease progression) including: 1) Infecting subtype, where higher SPLV and faster progression to death were reported in patients with subtype D compared to those with non-subtype D infection[197]; 2) Mutations in *vpu* and *vif* have also been related with slow disease progression (reviewed in Ref [198]). Infection by HIV-1 with truncated Nef has been demonstrated to contribute to low-level viral replication and non-pathogenicity[199]; 3) Viruses with V2 loop extension have been associated with slow or no disease progression[200]; 4) People infected with HIV that had a high replicative capacity were more likely to have rapid rates of $CD4^+$ T-cell decline, independent of plasma viral load[201]; 5) Escape mutations in HIV-1 Gag, a relatively conserved protein with structurally critical domains, correlate with improved clinical markers of disease progression[202].

Host factors associated with HIV-1 disease progression

Host genetic factors are substantial contributors to the inter-individual variation observed in response to HIV-1 infection and have been associated with resistance to HIV infection, as well as the rate of disease progression and the likelihood of viral transmission in exposed individuals. These are some examples: 1) Individuals who are positive for certain HLA class I molecules and/or have specific killer immunoglobulin-like receptors (KIR) genotypes have strong virus control, implying that cells that express KIR are important in the control of HIV-1 disease[120, 203, 204]. HLA-B*5701 and HLA-B*2705 alleles are associated with decreased rates of disease progression in Caucasians[205]. Individuals with the HLA-B*57 alelle are less likely to present with symptomatic acute HIV infection and are more likely to have broader and stronger T-cell responses against HIV-1 peptides[206]; 2) Individuals with NK cell receptor genes KIR3DS1/KIR3DL1 and HLA-Bw4-801 have a slower progression to AIDS, a lower viral load and a slower decline of $CD4^+$ T-cells[207]; 3) The first reported variant of the

CCR5 gene was a 32-base-pair deletion (CCR5Δ32), encoding a CCR5 molecule that is not expressed on the cell surface and is a basis for some people's resistance to HIV-1 infection requires CCR5 alone as a coreceptor. However, homozygotes for CCR5Δ32 can occasionally be infected with HIV isolates that employ different coreceptors. Cell-surface expression of CCR5 is reduced in heterozygotes (CCR5Δ32/WT), which is assumed to be the reason for delayed progression to AIDS in these subjects[208].

Use of phylogenetics in HIV-1 research

The HIV-1 population in the recipient is similar to that in the donor, however, because HIV mutates quickly, each infected person's virus evolves differently over time, accumulating mutations as it spreads (leaving genetic footprints). As a result, the degree of similarity between viruses from different people is related to the likelihood of infection from a common source. This can be visualised in a phylogenetic tree by sequencing and analysing the HIV genomes of infected people. The study of the evolutionary history and genetic relationships among transmitted HIV strains is known as HIV phylogenetics, and a phylogenetic tree is an illustration of those relationships and transmission dynamics between the analysed sequences. The reliability of the conclusion is determined by both the assumptions used in the statistical analysis (parametric, non-parametric, or Bayesian) and the data or sequences available for analysis. **Fig. 15** summarises the phylogenetic inference process.

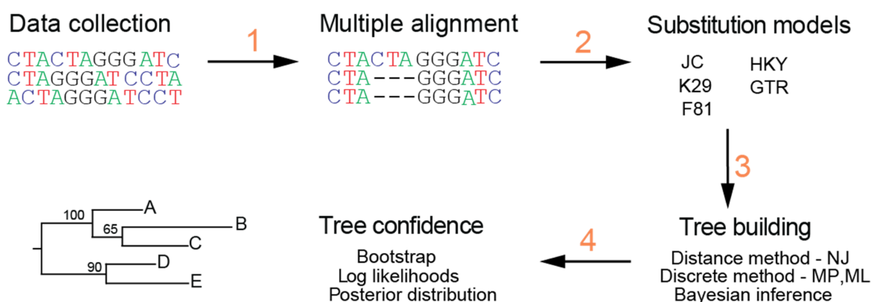


Figure 15. The general phylogenetic inference process.

The most fundamental assumption used in generating phylogenetic trees is that similar viral gene sequences share a common ancestor. 1). HIV-1 sequences must be properly aligned so that similar regions can be compared in the same alignment column. Alignments are typically determined using specific scoring algorithms, e.g Clustal. 2). The substitution model makes fundamental assumptions concerning base composition, rate, and frequency of base substitutions between sites, as well as the form of base substitutions, i.e. transitions vs transversions. The general time-reversible (GTR) model, for example, provides for distinct probability for each conceivable substitution. All substitutions are reversible, which means that substitution from x to y has the same probability as substitution from y to x. 3) There are two main approaches to determining the evolutionary relationship between the various sequences that would eventually comprise the final tree: distance and discrete (or character-based) methods. 4) Statistical approaches such as non-parametric bootstrapping are used to determine whether the reconstructed phylogeny is a reliable hypothesis for the given sequence data.

By coupling sequence data with geographic location and date of collection, a map that shows all sampled HIV-1 infections in a location can be created, as well as a phylogenetic tree that indicates pattern clustering. By examining whether clusters exist and where the bulk of infections occur, HIV transmission chains or distinct sub-epidemics can be identified. If this research is implemented in real-time, healthcare personnel may be able to rapidly target HIV prevention efforts to high-risk individuals and geographic areas. Phylogenetic transmission dynamics can be used to determine whether interventions are effective, whether a specific risk activity is driving the spread or growth of an epidemic in a particular location, and why various lineages behave differently. Using phylogenetics, researchers are gaining a better understanding of how and where HIV spreads, allowing public health officials to rapidly target their treatment and preventions efforts to curb the epidemic. HIV phylogenetics can be classified into three broad terms: Molecular epidemiology, phylodynamics, and phylogeography. **Table 5** summarises HIV-1 research in these three categories and outlines tools used to measure several parameters.

HIV-1 phylogenetics between hosts

Table 5. Use of phylogenetics in HIV-1 research

What?	Examples in HIV-1 research	How are these parameters measured?
Molecular epidemiology		
Quantifying the risk of spread and transmission: Time, source and place	1. Origin and evolutionary rate of HIV[27] 2. Classification, subtyping of HIV[22] 3. Direction of transmission 4. Identify characteristics of those at risk of infection 5. Transmission clusters[209] 6. Detection of recombination[210] 7. Sites under positive and negative selection[211]	1. Multiple sequence alignment 2. Phylogeny reconstruction *-Models of nucleotide substitution* *-Model testing* 3. Method of inferring phylogeny *-Distance or parametric or Bayesian* 4. Tree confidence 5. Recombination analysis 6. Cluster analysis 7. Selection analysis
Phylodynamics		
Use of phylogenetic, coalescent and statistic models to make inference about the spatio-temporal dynamics	1. Time to most recent comon ancenstor of HIV[31] 2. Impact of prevention intervention strategies[212] 3.Effective population size[213] 4. Spatio-temporal dynamics of the epidemic[38] 5. Impact of external factors in HIV-1 spread[214]	1. Time stamped data 2. Sampling strategy 3. Molecular clock 4. Bayesian method of infering phylogeny 5. Population genetics/coalescent theory *- Skygrid models*
Phylogeography		
Impact of HIV-1 migration on spread and identification of geographic hubs of transmission	1. Infer ancestral state locations for sequences sampled under the discrete or continuous context 2. Source – sink dynamics[215] 3. Virus movement[216] 4- Links between site or geographic location	1. Bayesian framework 2. Phylogeographic diffusion models BSSVS 3.Markov jumps

HIV-1 phylogenetics within hosts

HIV-1 is a fast-evolving virus due to the low fidelity of reverse transcriptase (an error rate of 3-4x10^{-5} mutations per nucleotide per cycle), rapid generation time (1.5-2.6 days with ~10^{10}-10^{12} new virion per day) and rapid recombination rate (3-9 RNA template crossovers per genome per round of replication)[39, 217-219]. As a result, the virus population within an HIV-1 seropositive individual is measurably evolving, which means that DNA sequences taken at different points in time will display a statistically significant number of genetic differences[220]. In other words, intra-host or within-host

evolution is an observable microevolutionary process that can be studied by analysing viral sequences sampled longitudinally from an infected host[221].

The genetic variability within an infected individual, which can reach up to 5% at the nucleotide level, can be summarised by a single statistic – the pairwise diversity, which is the average pairwise genetic (nucleotide or amino acid) distance within sequences sampled at a given time point[222]. Divergence estimate, which is the movement of the genetic variants away from the infecting variant(s) or the average genetic distance from the most recent common ancestor, is sometimes paired with genetic diversity estimates. Genetic variability can also be described in detail by within-host phylogeny or dendrogram to indicate how the sampled HIV-1 sequences are related to each other or how transmissions occurred from the donor to the recipient (monophyletic or polyphyletic), or how sampled variants from more than one patient are epidemiologically linked. However, phylogenetic transmission reconstruction can be misleading sometimes due to incomplete lineage sorting, which results in disordered transmission events, such as it may seem like the infant infected the mother. This is determined by the level of diversity and the amount of time that has passed between two transmission events (reviewed in Ref. [222]).

Examples of intra-host HIV-1 evolutionary studies

Within-host relationships are robust when at least 20 HIV variants are sequenced from each patient per timepoint (reviewed in Ref. [222]). With this sample size, one can be 95% confident that a missed variant comprised <15 % of the virus population[155].

Patterns of intra-host virus diversity and divergence over time

1. HIV-1 is associated with a significant population bottleneck, where a new infection is established by a single virion[50, 173, 223].
2. The number of transmitted viruses appears to be dependent on the transmission route[224]. In MSM, multiple transmission variants are more frequent than in vertical transmission[157]. This could be due to anatomic barriers and immune selection mediated by early CTL responses[173].
3. The asymptomatic stage of natural HIV-1 infection is characterised by a steady increase in diversity and divergence, whereas in the symptomatic phase, a decrease in diversity and stabilisation of divergence has been observed[225, 226].

This is attributable to a decrease in target cells sustaining viral replication and less effective selection of the virus[225, 226].

Relationship between intra-host virus evolutionary rate, immune response and disease progression

1. Patients with higher polyfunctional CD8+ T-cells and thus low classified as low risk of progression, had a lower evolutionary rate than in patients with a higher risk of progression[227].
2. Synonymous substitution rates predicted HIV-1 disease progression because they were expected to be neutral and proportional to the underlying viral replication rate[228].
3. There was an inverse relationship between the rate of viral evolution and disease progression. However, there were discrepancies based on the number of patients studied, and gene region analysed (reviewed in Ref. [221]).
4. A high intra host evolutionary rate has been associated with co-receptor use[229].

Reconstruction of the virus transmission history

1. Within-host evolution resolves epidemiological linkage recovered known epidemiological linkages, predicted direction of transmission and revealed common transmission of multiple variants[230, 231].

Identify how HIV-1 spreads within a host

1. By dissecting the evolutionary patterns that may arise from different transmission network types, one can accurately estimate the infected hosts in a growing epidemic[232, 233]. Considering within-host diversity, can accurately estimate the number of infected hosts in growing HIV epidemics and improve proper inference about the underlying epidemic interaction network.
2. Intra-patient recombination between R5 and X4 viruses occurs frequently which may in part be responsible for the coreceptor switch (X4 viruses) later stages of infection[234].

Hierarchical Phylogenetic models (HPM)

Given that the phylogenetic signal observed within individuals may be low at certain time points, hierarchical phylogenetic models can be used. These models allow the borrowing of information from one individual to another, thus providing more precise within-individual level estimates[235, 236]. To identify the differences in evolutionary processes between patient groups e.g. slow vs. fast progressors, fixed effects are incorporated with the HPM. This model still allows the expected evolutionary parameter within individuals to vary, and thus providing a well-designed hypothesis testing framework (**Fig. 16**).

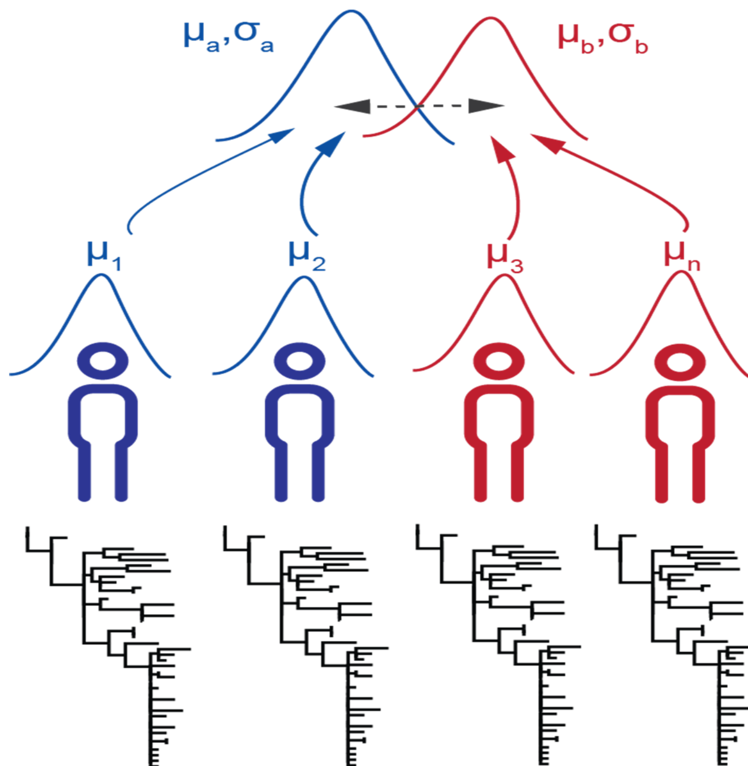


Figure 16. The hierarchical phylogenetic model.
In fixed-effects models, each individual is grouped into two categories for example slow progressors (blue) and fast progressors (red). The individual intra-host evolutionary rate is estimated for all available sequences from different timepoints. An assumption is that the evolutionary rate for each patient is drawn from an underlying normal distribution, where the mean and variance of the underlying prior distribution are unknown and estimated simultaneously along with all sequence data. When the sequence data for a patient is sparse, pooling information allows for more precise individual-patient parameter estimates. In addition, this methodology can be used to estimate the overall rate of nucleotide substitution, the nonsynonymous/synonymous substitution rate ratio (dN/dS), and codon substitution rates within patients or groups.

HIV-1 plasma proteomics

Characteristics of the plasma proteome

In humans, the blood circulatory system is the primary link connecting body organs. As a result, it can be used to provide information about the health status of key organs. The blood also transports metabolic products from organs and cells, while at the same time delivering oxygen and nutrients. Hence, blood can be viewed as a circulating representation of both physiological and pathological processes. One of the key functions of blood is to transport components of the immune system that are part of host defence mechanisms. Its cellular components (erythrocytes, thrombocytes and lymphocytes) are suspended in a liquid called plasma[237]. This comprehensiveness, combined with the ease of access, repeatability, and availability of large biobanks for retrospective analyses of human blood/plasma, as well the vast medical laboratory infrastructure already in place for its analysis, ensures that plasma will remain the preferred diagnostic material for the foreseeable future. Proteins in plasma have been studied since before we even knew what genes were[238].

Up to 5,800 proteins of the total 20,090 human protein-coding genes have been detected in plasma, and there are some compelling arguments that all these human proteins might be present in plasma to some extent given that blood flows through practically all organs of our body[238, 239]. The plasma proteome, or protein index, is distinct from the plasma protein (proteins that carry out their functions in circulation). The proteins constituting the blood plasma proteome are classified into three different classes: **1)The classical or functional plasma proteins** (proteins secreted by solid tissues and act in plasma)(dynamic range = 5×10^4-5×10^{10} pg/ml)[240]. This category covers the most abundant proteins with a physiological functional role in blood, which are mostly secreted by the liver and intestines, have a molecular mass greater than the kidney filtration cut-off (about 45 kDa) and have extended residence time in plasma. The classical proteins include human serum albumin (HSA) with a normal concentration range of 35-50 x 10^9 pg/ml constituting the 55% of the protein mass because its daily synthesis of ~12g in the liver with a lifetime of about 21 days[241] The other classical proteins include apolipoproteins,(lipid transport and homeostasis proteins); acute phase proteins (innate immune proteins); coagulation cascade proteins;

and immunoglobulins (about 10 million different sequences of antibodies are thought to circulate in a normal adult). **2) the tissue leakage proteins** (dynamic range =10^2-10^6 pg/ml) do not have a functional role in plasma but a leak from damaged cells and tissues, particularly in disease conditions, or are released from infectious organisms[238, 240]. This comprises proteins such enzymes aminotransferases which are used for the diagnosis of liver diseases, as well as low-level, tissue-specific isoforms of proteins such as cardiac troponins. Most of the important diagnostic markers are found in this group. **3) Signal proteins** (dynamic range =5-1000 pg/ml) and these include; "long-distance" receptor ligands – hormones; "local" receptor ligands – cytokines and other short distance mediators of cellular responses that are upregulated when needed. These proteins have a relatively short residence time in plasma due to having molecular weight under the kidney filtration cutoff [238].

Given that the second-largest component in plasma is due to tissue leakage[242], it is possible that all 20,090 human protein-coding genes and their splice variants, post-translational modification and cleavage products are present in plasma, making it the most comprehensive and largest version of the human proteome. Most measurement methods will collapse all the protein form into a simple form thus simplifying this complexity. Despite its extensive research and therapeutic value, human plasma remains mostly uncharacterized, or is a challenging protein-containing sample to define. This is due to the large proportion of albumin (55%), the wide dynamic range (variation in concentration of different components of about 12 orders of magnitude) in the abundance of other proteins, and the tremendous heterogeneity of its predominant glycoproteins (most plasma proteins are heavily glycosylated)[243].

Detection of proteins in plasma

Many different technologies are currently deployed to generate protein measurements in serum or plasma, ranging from highly optimized single-protein assays (enzymatic assay or immunoassay, electrophoresis) commonly used in clinical laboratories to exploratory and more flexible mass-spectrometry-(MS)-based workflows, multiplex immune assays and affinity-based assays used as part of research workflows. However, the very wide range of analyte concentrations that must be detected and measured

simultaneously in the absence of means to amplify them, as there is no PCR equivalent for proteins. MS is currently the only technology that can identify these proteins in a versatile and unbiased manner, as opposed to the predefined content offered by affinity-based approaches or immunoaffinity capture.

Mass spectrometry (MS)-based proteomics

MS is a technique for determining the mass-to-charge ratio (m/z) of one or more molecules in a sample. For example, it is feasible to detect up to 500 proteins in un-depleted or neat plasma in a single liquid chromatography-tandem mass spectrometry (LC-MS/MS) run and proteins can increase up to 4,000 depending on how plasma is pre-processed before an LC-MS/MS run[244-246]. Proteomic approaches have been divided into top-down (proteins are analysed as a whole) and bottom-up (proteins are digested into peptides) LC-MS/MS[243]. The top-down approach has a significant disadvantage in that MS analysis of whole proteins does not directly provide a sequence-based identification (there are many proteins with close to a given mass), and thus the protein peaks discovered as markers are not strictly speaking identified without significant additional effort.

Bottom-up proteomics employs two approaches: data-dependent acquisition (DDA) and data-independent acquisition (DIA). DDA selects only certain peptides within a narrow range on m/z signal intensity whereas DIA does not target specific peptides but rather fragments ad analyses all peptides (unbiased)[247, 248]. The Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH-MS) is a specific variant of DIA[249]. Every peptide in the protein digest is comprehensively and repeatedly sampled side-by-side with a wide isolation window in this method, and each peptide can be extracted over time (retention time). Peptide query parameters can be used to retrieve information about specific peptides that are detectable in the data based on ion intensity, elution time, and charge. These query parameters are acquired from a spectral library that is generated from fractionation of the samples of interest and a DDA analysis[249].

In the majority of proteomics experiments regardless of the labelling approach, the extracted proteins are first digested into peptides, usually by trypsin, to create molecular species that are more manageable in both sample handling and MS analysis. Enzymatic

digestion of a whole proteome yields thousands of peptides; this sample complexity is not incompatible with the MS analysis. As a result, the first step in the MS proteomics workflow is frequently aimed at reducing the sample complexity through sample prefractionation. The peptide population is fractionated based on its physicochemical properties, such as charge, isoelectric point, hydrophobicity or combinations of these. These pre-fractionated peptides are then introduced to the ultra-high-performance liquid chromatography (uHPLC) system for an additional separation step to reduce complexity even further. Ideally, all peptides eluted from the LC are queried by the mass spectrometer, but in everyday practice, many peptides elute simultaneously and compete for efficient ionisation: highly abundant species might suppress the ionization of co-eluting less abundant species, inhibiting their MS detection. Following ionisation, peptide precursor ions are introduced into the mass spectrometer, which accurately records their mass-to-charge (m/z) ratio. Single precursors are selected (based on observed intensity) and subjected to MS to generate distinctive fragment ions for the selected precursor in DDA, whereas in DIA, the selection is based on an isolation window (**Fig. 17**). In SWATH-MS mode, typically a single precursor ion (MS1) spectrum is recorded, followed by a series of fragment ion (MS2) spectra with wide precursor isolation windows (for example 25 m/z). Through repeated cycling of consecutive precursor isolation windows over a defined mass range, a comprehensive data set is recorded, which includes continuous information on all detectable fragments and precursor ions. Hence, extracted ion chromatograms can be generated on MS2 as well as MS1 levels[250] (**Fig. 17**). Using search algorithms implemented in software like Spectronaut,[251] the combination of precursor m/z and its fragment ions is then matched to known peptide sequences from large protein databases like UniProt[252]. Finally, data is quantified (either relatively or absolutely). These protein abundances are then evaluated and visualized in the context of the biological question under study.

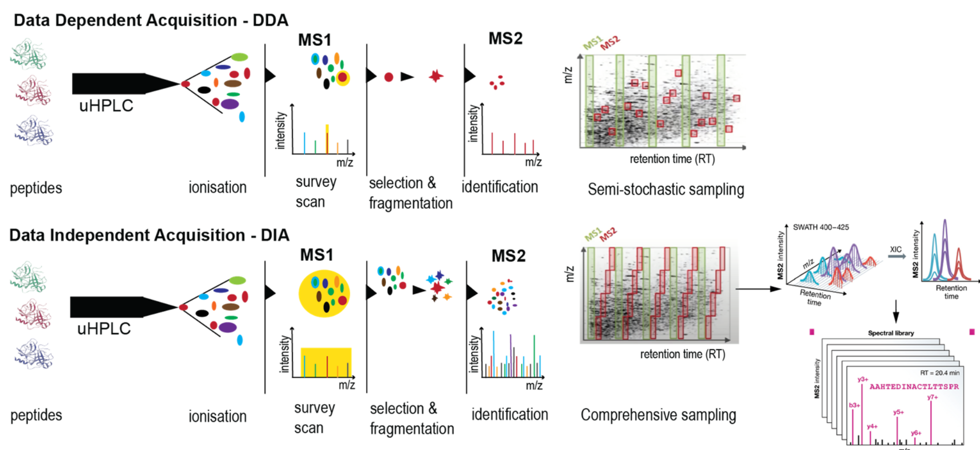


Figure 17. Label-free proteomics: discovery strategies.
In DDA, the MS2 spectra are acquired for specific precursor ions, where all fragments detected come from one single peptide precursor ion resulting in a semi-stochastic sampling of the peptide population of a sample. In DIA, MS-2 scans with a wide isolation window are acquired for example in this figure the isolation window of 400-425 is used. All ions in this window are fragmented together, allowing MS2 fragmentation for all precursors in a sample across the entire retention time range. However, the MS2 data is complex necessitating the use of specialised software like Spectronaut to deconvolute these MS2 spectra and, quantify the peptides. Part of this figure is adopted with permission from Ref.[250]

Plasma proteins detected by mass spectrometry

About 4,000 plasma proteins and their concentrations have been quantified by mass spectrometry[239, 242]. Only 730 of these are actively secreted to blood or locally released in the blood, with the majority being intracellular proteins that leak into the bloodstream that occurs naturally due to cell apoptosis or renewal, or may appear during sample preparation[253].

Only eight publications have investigated and quantitated the plasma proteins during HIV-1 infection, according to a comprehensive search of indexed journals in PubMed. These are primarily from chronically infected individuals in cross-sectional studies comparing HIV-1 disease progression groups. **Table 6** summarises the findings of this review.

Table 6. Previous HIV-1 plasma proteomics research

First Author	Year	N	Group	HIV stage	Method	P	Country	Main message
Tarancón-Diez [254]	2021	20	Early vs. late ART	acute	MS	188	Spain	Protein signature shows a different proinflammatory state associated with a late cART introduction
Al-Mozaini [255]	2021	8	HIV-1 vs. HIV-2	chronic	MS	314	Saudi Arabia	HIV-specific/related protein expression changes identified that could be used in early diagnosis and prognosis of HIV
Zhang[256]	2018	74	EC vs. RP	chronic	Affinity immuno assay	92	Sweden	CCL4 and CCL7 have significantly higher levels in EC compared to VP. CXCL9, CXCL10, and CXCL11 significantly decreased levels in EC compared to VP
Sperk[257]	2021	42	EC vs. healthy	chronic	MS	81	Sweden	Increased in CCL4, CCL7, CCL20, and NOS3 in EC compared with HIV control
Leema[258]	2020	83	HIV-1 vs. healthy	chronic	Affinity immuno assay	92	Ethiopia	CCL4 chemokine was the most critical protein for successful classification between children who initiated ART at different time points.
Babu[259]	2019	44	HIV-1 vs. healthy	chronic	Affinity immuno assay	184	India	TRANCE, NT-3,CD8A siginificantly different
Kramer[260]	2010	19	Plasma donors	acute	MS	34	USA	Acute phase proteins were induced systemically prior to the first detection of viremia and before detectable increase in plasma cytokine levels.
Rodríguez-Gallego [261]	2019	18	VC vs. NC	chronic	MS		Spain	Eighteen proteins were found to exhibit statistically significant differences in plasma levels between TCs before the loss of HIV-1 control and PCs

Abbreviations: N, the number of samples used; P, the number of proteins; EC, elite controller; RP, rapid progressor; VC, viral controller; NC, non-controller; MS, mass spectrometry; TC, transient controllers; PCs, persistent controllers; ART, antiretroviral therapy; Cart, combination ART; NOS3, Nitric Oxide Synthase 3; TRANCE, Tumor necrosis factor (TNF)-related activation-induced cytokine.

Comparison of depleted and neat plasma

This extraordinarily wide dynamic range poses a considerable difficulty to profile the entire plasma/serum proteome by a proteomics technology (often based on LC-MS/MS). To address this issue, highly abundant plasma proteins (serum albumin (HSA), albumin, IgG, IgA, IgM, IgD, IgE, kappa and lambda light chains, alpha-1-acidglycoprotein, alpha-1-antitrypsin, alpha-2-macroglobulin, apolipoprotein A1, fibrinogen, haptoglobin, and transferrin) are frequently depleted, typically through columns with immobilized antibodies directed against the top 1 – 20 abundant proteins. This typically extends the coverage to 500–800 proteins[262].

Unfortunately, there are disadvantages to employing depletion systems, including **1**) the possibility of removing some low-abundance proteins along with the abundant proteins. Albumin, the most abundant protein present in plasma, is a transport protein that binds a variety of molecules including hormones, lipids and amino acids. As a result, removing albumin from plasma may result in the specific loss of some low-abundance peptides or small proteins of interest, such as cytokines[263]; **2**) Because antibodies are never totally specific, some untargeted proteins may be removed concurrently[264]. **3**) The depletion step is carried out under non-denaturing conditions, resulting in co-immunoprecipitation and removal of antigen-bound proteins; and **4**) the depletion step results in significant experimental variations[262]. In conclusion, immune-depletion columns can broadly improve the detection of nontargeted plasma proteins by a factor of four.

Aims of this doctoral dissertation

The overall aim of the studies presented in this thesis was to assess how HIV-1 evolves within and between adults and infants and to understand how hosts with different progression rates respond to HIV-1 infection during acute HIV-1 infection.

Specific Aims

Paper I: To investigate the molecular epidemiology of HIV-1 in Nigeria by analysis of *pol* sequences from four geo-political zones.

Paper II: To provide an overview of the contribution of phylogenetic inference in dissecting HIV-1 mixing between geographic areas and key populations in sub-Saharan Africa.

Paper III: To describe the relationship between innate immune responses and acute retroviral syndrome during acute HIV-1 infection.

Paper IV: To characterize the blood plasma proteome of study participants enrolled in cohorts from four African countries before, during and after acute HIV-1 infection, with an overall aim of elucidating virus-host mechanisms involved in the regulation of HIV-1 disease progression.

Paper V: To determine the *gag* and *nef* intra-host HIV-1 diversity and evolution within the first 15 months of life in infants with different rates of disease progression.

Materials and methods

Cohorts

Paper I

The Institute for Human Virology-Nigeria (IHVN), a PEPFAR implementing partner that offers ART to over 60,000 public sector patients through the AIDS Care and Treatment in Nigeria (ACTION) programme, initiated the two cohorts as well as collected the data and samples used in this study. The first cohort was a retrospective cross-sectional examination of archival samples from 175 HIV-infected individuals who experienced virologic failure and were recruited into an HIV treatment and care programme between November 2006 and December 2007. Participants in the study went to IHVN-supported facilities, including the University of Abuja Teaching Hospital (UATH) and the National Hospital Abuja (NHA). They qualified to be included in the cohort if they had HIV-1 RNA testing between November 2006 and December 2007, were above the age of 18, were on NNRTI-based first-line regimens, and had not received any protease inhibitor (PI)-based medication before viral load testing. Six first-line therapy regimens were frequently provided during the study period: stavudine, lamivudine, nevirapine, or efavirenz (D4T/3TC/NVP or EFV); zidovudine, lamivudine, nevirapine, or efavirenz (AZT/3TC/NVP or EFV); and tenofovir, emtricitabine, nevirapine or efavirenz (TDF/FTC/NVP or EFV). Patients were monitored for $CD4^+$ counts and toxicity in the laboratory on a six-monthly basis[265]. The average age was 38.1 years old (standard deviation (SD), 8.5), with 49% of the participants being female. The median log viral load was 4.7 and the median $CD4^+$ T-cell count was 128 cells/L[266]. The second cohort was a cross-sectional study as part of the ACTION programme, in which participants had viral load testing to confirm HIV treatment failure. It included 191 samples from patients who lived in Abuja between 2011 and 2014 and were subjected to genotypic drug resistance testing. The median age was 36, and 58% of the patients were female. The median number of $CD4^+$ T-cells was 154[266].

Paper II

This was systematic literature review based on an exhaustive search of the PubMed database for peer-reviewed research articles on HIV-1 phylodynamics in sub-Saharan Africa published in English in 1995-2021. The MeSH terms (HIV-1) AND (Africa) were used to select HIV-1 articles from African countries. The keywords "phylogenetic analysis" OR "phylodynamics" OR "evolution" OR "phylogeny" OR "molecular epidemiology" OR "transmission" were used to widen the scope and to ensure that all relevant research articles were included. Articles were independently reviewed and selected by two independent investigators.

Paper III

The data and samples used in this study came from four cohorts established by the International AIDS Vaccine Initiative (IAVI) in Kenya, Rwanda, Uganda, and Zambia. IAVI has funded and built research capacity at nine clinical research centres (CRC) in Kenya, Uganda, Rwanda, Zambia, and South Africa since 1999[267]. These cohorts were established to better assess HIV-1 risk populations, their suitability for HIV vaccine clinical trial participation, and their unmet requirements for preventive services and products[268]. Participants (over the age of 18) were included in this prospective, longitudinal, multi-centre study between 2006 and 2011 to examine laboratory, clinical, immunologic, and viral predictors of disease progression in recently HIV-infected volunteers (IAVI Protocol C). These cohorts mostly comprised participants who were at high risk of HIV-1 infection, such as ART-naive HIV-1 negative partners in serodiscordant partnerships, female sex workers, fisherfolk, and men who had sex with males. Volunteers with a diagnosis of hyper acute HIV-1 infection (hAHI) were eligible (defined as HIV-1 antibody negative and RNA (Fiebig stage I) or p24 antigen-positive (Fiebig stage II)). The analysis also included available matched pre-infection (HIV-1 negative) samples from study participants.

This study also included data and samples from Swedish study participants who were diagnosed with AHI before 2011. They were de-identified from Skåne University Hospital's electronic patient register (InfCareHIV).

Paper IV

The same study participants as analysed in paper III, were analysed also in this paper, except those from Sweden. This was due to a shortage of matching pre-infection plasma samples among the Swedish individuals. Instead, another prospective cohort of 15 HIV-uninfected women between the ages of 18 and 23 years old that were at risk of infection was included. These women were enrolled in the Females Rising through Education, Support, and Health (FRESH) research in KwaZulu-Natal, South Africa, between 2011 and 2016, to diagnose acute HIV infection during the Fiebig stage I[269]. This was made possible by testing for HIV RNA twice a week in this high-risk cohort, as well as collecting biological specimens (blood and mucosal sample) before and immediately after infection in local venues frequented by young people, such as cafes, nightclubs, and shopping malls. Women who were HIV-uninfected, sexually active, not pregnant, non-anaemic (haemoglobin 10 g/L), and free of major chronic conditions were eligible. The period of the surveillance ranged from 48 to 96 weeks. Three-time points were employed in this study: pre-infection, two weeks, and one month after infection.

Paper V

Between 1999 and 2002, data and samples were collected as part of the Prevention of HIV-1 Transmission Study in Nairobi, Kenya. The parent cohort included 465 HIV-infected women recruited during pregnancy and followed for up to two years postpartum with their infants[270]. Seventy-two infants acquired HIV-1 infection by the age of one month. To lower the risk of HIV-1 transmission, mothers were given a brief course of zidovudine (ZDV) during the third trimester. There was no national ART programme for children during the study period, thus all infants were treatment-naive. Infant plasma samples were collected at birth, as well as at months 1, 3, 6, 12, 15, 18, 21, and 24, whereas mother plasma samples were collected at week 32 of their pregnancies, during delivery, and one month later. HIV-1 infection was determined by a positive HIV-1-gag DNA PCR from dried blood spots or by detection of HIV-1 RNA from plasma in an infant was characterised as HIV-1 infection[271]. Infection was divided into three categories: *in utero*, peripartum/early breast milk, and late breast milk transmission. Infants that were infected by one month of age, and had three or more longitudinal plasma specimens available for examination were included in this study.

Study design

Table 7. Summary of study design and study participants included in the PhD projects.

Paper	Study Design	ART	Site, Country	Study Participants	Sample Size	Year of sampling	Variables
I	Retrospective Cross-sectional		Abuja, Nigeria	Adults	336	2006-2011	*pol* sequences, location
	Cross-sectional		Abuja, Nigeria	Adults	1076	1999-2013	*pol* sequences, location
II	Systematic		Sub-Sahara Africa		64	2021	Published articles
III	Prospective (Longitudinal) Matched	No*	Kilifi, Kenya		29	2005-2011	AHI symptoms, Innate immune responses, HIV-1 subtypes, VL, CD4, Age, sex
		No*	Kigali, Rwanda		4	2005-2011	
		No*	Masaka, Uganda		5	2005-2011	
		No*	Lusaka, Zambia		4	2005-2011	
			Skåne, SE	Adults	13	Before 2011	
IV	Prospective	No*	Durban, SA	Females	15	2012-2016	Protein expression
V	Prospective – paired (Longitudinal)	Yes	Nairobi, KE	Mothers	14	1999-2013	*gag, nef* sequences, CD4 count,
		No		Infants	14	1999-2013	CTL responses

Definitions

Paper III & IV

Estimated date of infection (EDI): The date between the last negative and first positive test in the case of HIV-1 antibody detection, 14 days before the test date in the case of p24 antigen detection only, or 10 days before the test date for volunteers who had a PCR-positive result prior to antibody or p24 antigen detection. If a volunteer can identify a clear exposure event, the date of that event may be used as the EDI at the discretion of the research team.

Paper V

Infant HIV-1 infection was classified as *in utero* if a positive *gag* PCR was detected within the first 48 hours of life, or peripartum (intrapartum, or through early nursing) if the *gag* PCR HIV-1 was negative or undetectable >48 hours after birth and one month of life.

Laboratory procedures (methods)

Paper I

On-site IHVN personnel performed wet laboratory procedures. **Figure 18** is a schematic presentation summarising the processes performed in Lund.

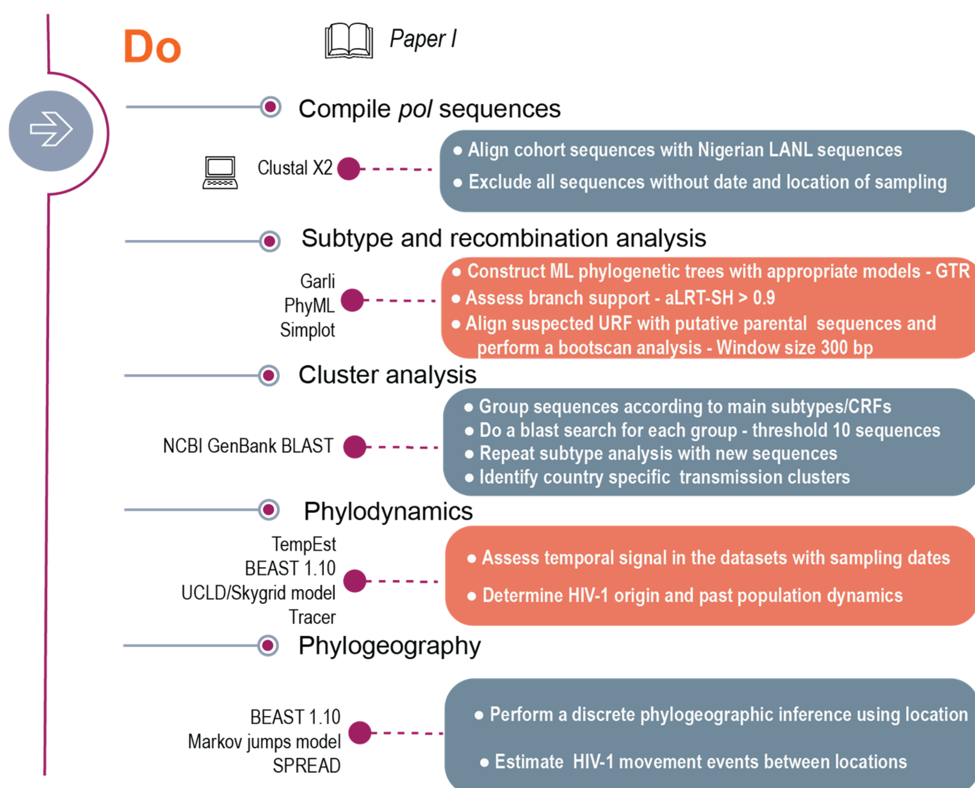


Figure 18. Paper I – Summary of analytical procedures.
Abbreviations: LANL, Los Alamos National Lab; GTR, general time-reversible model; aLRT-SH, approximate likelihood ratio test and Shimodaira–Hasegawa like branch support; CRFs, circulating recombinant forms, UCLD; uncorrelated relaxed clock with a lognormal distribution; BEAST, Bayesian Evolutionary Analysis by Sampling Trees; SPREAD; Spatial Phylogenetic Reconstruction of Evolutionary Dynamics; URF, unique recombinant forms

Paper II.
Approximately 2700 phylogenetics research articles published and indexed in the PubMed database were used to conduct this systematic review. About 360, 280, and 64 articles were identified based on the title, abstract, and full text respectively.

Paper III
HIV-1 RNA extraction, PCR, and DNA sequencing were performed on plasma from study participants to generate sequences needed for HIV-1 subtype characterization. Using a 96-well format working electrode covered with 40 capture antibodies on discrete and well-defined spots, 40 cytokines or chemokines were quantified using the Meso Scale discovery (MSD) immunoassay method. When a voltage is applied to antibodies bearing electrochemiluminescent labels, light is produced, and this is how the MSD device identifies them. The analyte concentration in the sample is proportional to the amount of light emitted, providing a quantitative measure of each analyte (pg/ml) in the sample. A standardised questionnaire administered during clinic visits was used to obtain the eleven symptoms of acute HIV-1 infection.

Paper IV
In collaboration with the Swedish National Infrastructure for Biological Mass Spectrometry (BioMS), all wet laboratory procedures were performed in Lund. **Figure 19** is a schematic presentation summarising the processes.

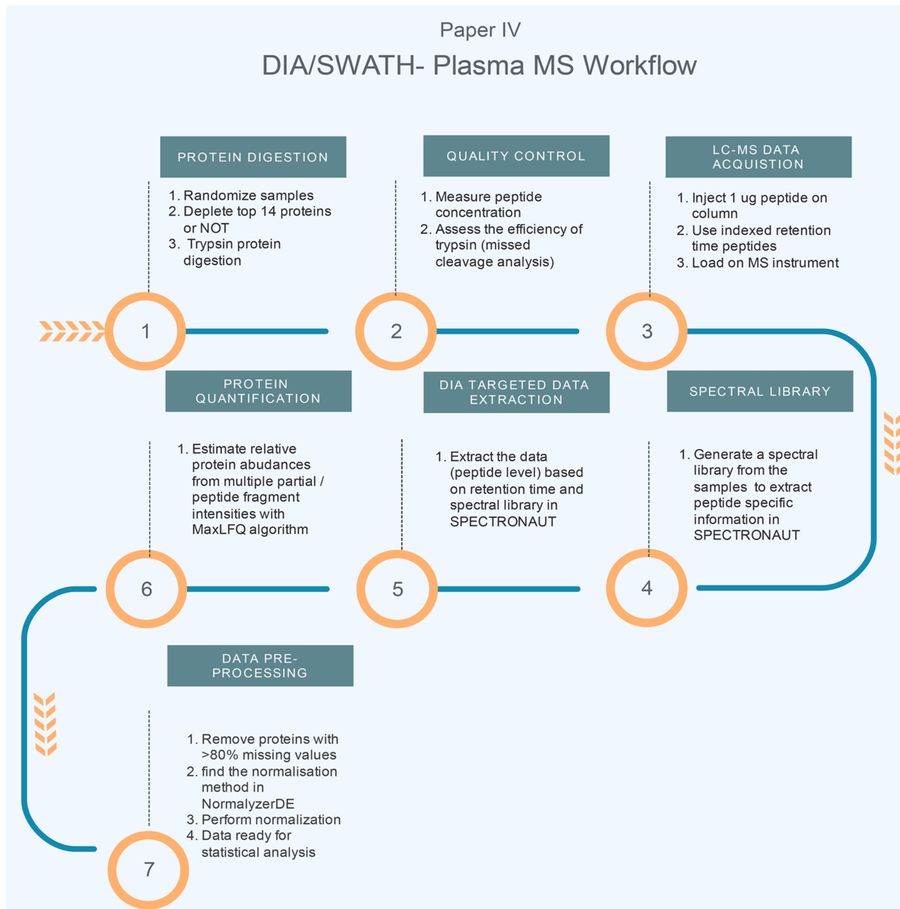


Figure 19. Paper IV – Summary of plasma proteomics laboratory procedures
Abbreviations: SWATH, Sequential Window Acquisition of all Theoretical Mass Spectra; DIA, data-independent acquisition

Paper V

Following RNA extraction, PCR, and clonal sequencing, approximately 15 – 25 clonal sequences for both the *gag* and *nef* regions of HIV-1 were obtained from each available mother and infant sample per time point. **Figure 20** summarizes the various dry laboratory methods used in this study.

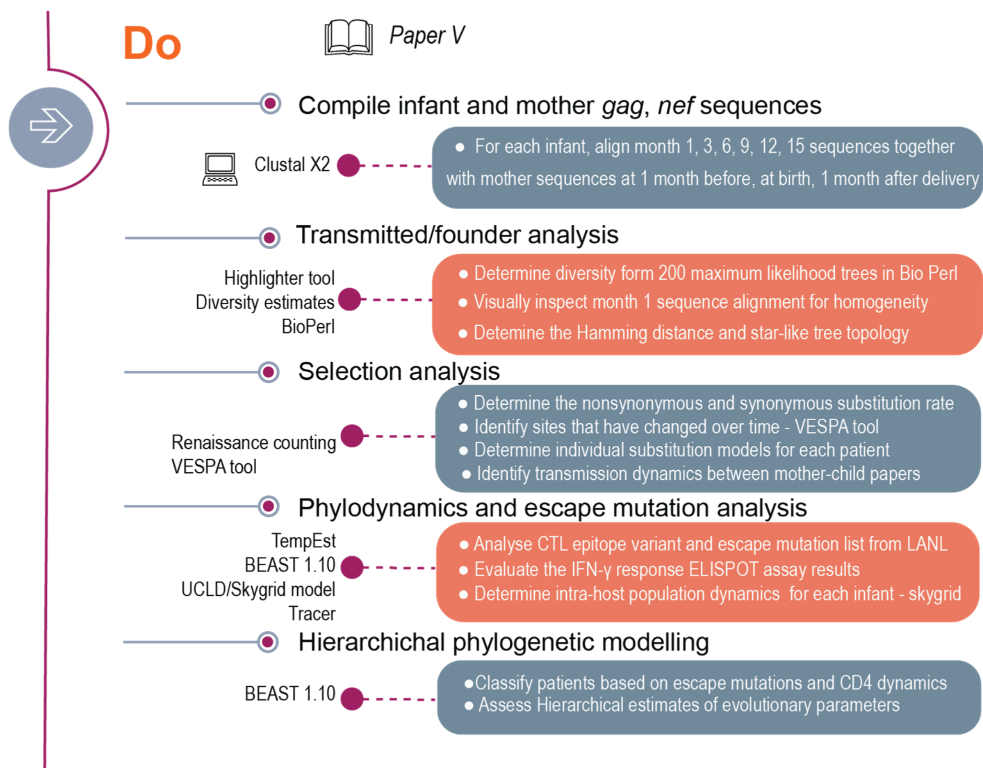


Figure 20. Paper V – Summary of analytical procedures.
Abbreviations: LANL, Los Alamos National Laboratory; GTR, general time-reversible model; aLRT-SH, approximate likelihood ratio test and Shimodaira-Hasegawa-like branch support; CRFs, circulating recombinant forms, UCLD; uncorrelated relaxed clock with a lognormal distribution; BEAST, Bayesian Evolutionary Analysis by Sampling Trees; SPREAD; Spatial Phylogenetic Reconstruction of Evolutionary Dynamics; CTL, cytotoxic CD8^{+} T-cells. ELISPOT; Enzyme-linked immunosorbent spot, IFN, interferon

Statistical analyses

Table 8: Summary of the statistical methods used in this thesis

Description	Dependent (outcome) variable	Independent (explanatory) variable	Parametric test	Non-parametric alternative test
Comparing means				
Average of two independent groups	Continous	Nominal /Binary	Independent t-test	Mann-Whitney
Average of >two independent groups	Continous	Nominal /Binary	One-way ANOVA	Kruskal Wallis
Average of two paired (matched) samples	Continous	Nominal /Time /condition	Paired t-test	Wilcoxon signed rank
> two measurements on the same subject	Continous	Nominal /Binary	Repeated measures ANOVA	Friedman
Investigating relationships				
Relationship between two continuous variables	Continous	Continous	Pearson's correlation coefficient	Spearman's correlation
Relationship between two nominal variables	Nominal	Nominal		Chi-squared test
Predicting a quantitative dependent variable by an independent variable	Continous	Any	Simple linear regression	Transform the data
Predict a nominal variable based on a single predictor variable	Nominal	Any	Logistic regression	
Predicting the dependent variable by the independent variable (inter-dependent data)	Continous	Any	Linear mixed model	Transform the data
A non–linear relationship between two continuous variables	Continous	Any	Spline regression	Transform the data
A non–linear relationship between two continuous variables	Continous	Any	Generalised additive model	Transform the data
Predict a nominal variable base ds on multiple predictor variables	Nominal	Any	Multiple Logistic regression	
Assessing the relation between two nominal variables large sample	Nominal	Nominal		Chi-square test
Examine the significance of association (contingency) between two kinds of classification)	nominal	nominal		Fisher's exact test
Survival /time to event analysis				
Estimate probability from observed survival time (Time it takes for an event to occur) – univariate	Continous – Time	Nominal		Kaplan-Meier
Comparing survival of one factor among two or more groups – univariate	Continous – Time	Nominal		Log-rank test

Assess the effect of several risk factors (one or more predictor variables) on survival time – multivariate	Continous – Time	Any	Cox proportional hazards regression model	
Unsupervised machine learning – Exploring and discovery				
Identify groups of similar objects in a multivariate data sets **without** sub diving the data				Hierarchical clustering
Identify groups of similar objects in a multivariate data sets by subdividing the data				K – means clustering
Summarise and visualise the most important information in a multivariate data set by reducing the dimensionality of the data without losing important information.	continuous			Principal component analysis
Supervised machine learning or classification techniques				
Predict the probability of belonging to a given class (or category) based on one or multiple predictor variables.	Nominal	Any		Discriminant analysis
Regression method based on dimension reduction technique that is supervised by the outcome				Partial least squares - PLS
Model validation and evaluation techniques				
Measure the performance of a predictive model by diving data into the training and testing set			Cross-validation set approach	
Graphical measure for assessing the performance or the accuracy of a classifier, which corresponds to the total proportion of correctly classified observations - plot the true positive rate (or sensitivity on the y-axis) against the false positive rate (or "1-specificity" on the x-axis) at all possible probability cutoffs.			Receiver operating characteristics (ROC curve)	
Summarises the overall performance of the classifier, and overall possible probability cutoffs. It represents the ability of a classification algorithm to distinguish 1s from 0s (i.e, events from non-events or positives from negatives)			The area under the curve (AUC)	
The proportion of true positives among all the individuals that have been predicted to be disease-positive by the model. This represents the accuracy of a predicted positive outcome. Precision = TruePositives/(true positives + false positives)			Precision	
True Positive Rate (TPR) or the proportion of identified positives among the disease-positive population (class = 1). Sensitivity = TruePositives/(TruePositives + FalseNegatives)			Sensitivity	
Measures the True Negative Rate (TNR), that is the proportion of identified negatives among the disease-negative population (class = 0). Specificity = TrueNegatives/(TrueNegatives + FalseNegatives)			Specificity	

Ethical considerations

Paper I

The National Hospital Abuja, the University of Abuja Teaching Hospital, the Nigerian National Health Research Ethics Committee, and the University of Maryland Baltimore Institutional Review Board all gave their approval to this study.

Paper II

No ethical aspects since it was a systematic review.

Paper III

All participants underwent an informed consent procedure before enrolment, and their signed informed consent was documented. The Kenya Medical Research Institute Ethical Review Committee, the University of Nairobi's Kenyatta National Hospital Ethical Review Committee, the Rwanda National Ethics Committee, the Uganda Virus Research Institute Science and Ethics Committee, the Uganda National Council of Science and Technology, the University of Cape Town Health Science Research and Ethics Committee, and the University of Zambia Research Ethics Committee all provided ethical approvals. The Swedish cohort received ethical approval from the Lund University Ethical Review Board (Diarienummer (Dnr) 2013/772).

Paper IV

The study was approved by the biomedical research ethics committee of the University of KwaZulu-Natal and the institutional review board of Massachusetts General Hospital.

Paper V

This study was conducted according to the principles expressed in the Declaration of Helsinki and the study protocol was reviewed and approved by the Kenyatta National Hospital Ethics Review Committee and the Institutional Review Board of the University of Washington. HIV-1 seropositive pregnant women were enrolled after the provision of written informed consent.

Results and discussion

This doctoral thesis consists of five papers focusing on different dynamics in HIV-1 disease and evolution, and can be divided into three categories: 1) Molecular phylogenetic description of the HIV-1 epidemic in Nigeria and Sub-Saharan Africa (with a focus on between-host HIV-1 evolution, Papers I and II); 2) Infant intra-host HIV-1 evolution and CTL responses during acute HIV-1 infection (Paper V); 3) Dynamics of early innate immune responses and plasma proteome before and during hyper-acute HIV-1 infection (Papers III and IV).

Molecular phylogenetic description of the HIV-1 epidemic in Nigeria and sub-Saharan Africa

Main findings

- In Nigeria, the most common circulating HIV-1 strains are CRF02_AG, CRF43_02G, and subtype G (Paper I).
- CRF43_02G primarily circulates in Nigeria compared to the rest of the world (Paper I).
- At least five large HIV-1 sub-epidemics were introduced in Nigeria in the 1960s and 1970s, close to the Nigerian Civil War (Paper I).
- In Nigeria, the HIV-1 epidemic began in large urban areas such as Abuja and Lagos and expanded to smaller rural communities (Paper I).
- The HIV-1 epidemic in sub-Saharan Africa is driven by various HIV-1 subtypes, which are typically geographically confined. The Southern African countries and Ethiopia is mainly driven by subtype C; in East Africa is mainly driven by sub-subtype A1 and subtype D; West Africa is mainly driven by CRF02_AG, sub-subtype A3, and subtype G; and Central Africa has a more complex and diversified pandemic (Paper II).
- HIV-1 infections in West and Central Africa appear to have started in cities and then expanded to rural areas, most likely due to human mobility (Paper

II). HIV-1 mixing has also been documented in rural and urban areas, as well as across national borders (Paper II).

- Some HIV-1 sub-epidemics appear to be localised, with no evidence of HIV-1 mixing with nearby groups. In addition, localized HIV-1 sub-epidemics, on the other hand, are important sources of HIV-1 transmission in nearby populations in different situations (Paper II).
- Human migration connected with economic activities such as mining and fishing may be associated with increased HIV-1 transmission (Paper II).
- We observed differences in the Subtype G nomenclature that has been previously used in Nigeria, which was probably caused by the misclassification of CRF43_02G. One of these categories includes subtype G'. Further dissection of this subtype G' cluster previously reported, showed that approximately half of the sequences belonged to subtype G, whereas the other half clustered with the CRF43_02G sequences in our analysis[272].

Dynamics of early innate immune responses and plasma proteome before and during hyper-acute HIV-1 infection

Main findings

- Symptoms of acute retroviral syndrome (ARS) are linked to a high innate immune response (cytokine storm) (Paper III).
- During AHI, plasma IP-10 was highly activated, and it was linked to ARS (Paper III).
- Plasma IP-10 was identified as a potential biomarker for detecting a greater innate immune response (Paper III).
- During AHI, the plasma proteome trajectories grouped into six distinct expression profiles (Paper IV).
- During AHI, 241 proteins were found to be significantly changed. These proteins have been shown to mediate pathological pathways including immune

system functions such as complement and coagulation, cell motility and adhesion involving actin polymerization, and platelet degranulation and metabolism (Paper IV).

- Twenty proteins were associated with ARS (**Paper IV**).
- Thirty-seven proteins that had prognostic value were associated with faster disease progression (**Paper IV**).

In **paper III**, we observed that 16 analytes were significantly elevated at Fiebig stage II (p24 antigen-positive, 15-20 days post-EDI) as compared to pre-infection levels. IP-10, IL-1α, IFN-γ, and IL-10 were more than 5-fold higher during hyperacute HIV-1 infection. This increase in plasma levels of multiple cytokines and chemokines (also known as cytokine storm) has previously been observed among 35 plasma donors with acute HIV-1 infection and 21 young females in South Africa[102, 158]. Large increases in IP-10 were observed in both these studies, with a peak in the hyperacute phase (4–11 days after the detection of viremia) but only partially resolved after peak viremia (24-32 days after the detection of viremia) and later time points. The intense early cytokine storm in acute HIV-1 infection, particularly IP-10, was found to be positively correlated with Gag-Protease replication capacity (higher IP-10, faster-replicating viruses), implying a role for viral virulence in driving the inflammatory response[158]. Furthermore, while this cytokine storm may contribute to viral replication control, it may also contribute to the early immunopathology of the infection and associated long-term consequences from prolonged immune activation.

Moreover, we observed that 18 innate immune analytes were higher in study participants who had AHI symptoms than in those who did not. And after controlling for age, the HIV-1 subtype, risk group, IP10 was associated with acute retroviral syndrome (ARS). Only one previous study in Thailand by Crowell *et al.* has reported on associations between innate immune responses during AHI and ARS (in 430 volunteers enrolled at Fiebig stage III or later), with TNF-α, IL-6, IL-7, IP-10, and MCP-1 being the only cross-cutting analytes[273]. In contrast, there was no association of IP-10 with ARS as observed in our study. This could be related to discrepancies in the sample collection time. Crowell and colleagues further demonstrated that ARS was associated with high viral burden, $CD4^+$ T-cell depletion, and immune activation.

While many of these disparities between participants with and without ARS were addressed following sustained ART, there was some evidence of increased chronic inflammation among participants who experienced ARS despite suppressive ART initiated during AHI. We also demonstrated that participants with stronger innate immune responses had increased odds of ARS and that IP10 can be used to distinguish individuals with stronger innate immune responses from those with weaker responses.

Paper IV is the first comprehensive and largest longitudinal MS-based plasma research of acute HIV-1 infection (AHI) that has been presented to date. We show that the plasma protein proteome profile is highly individual-specific, therefore using matched pre-infection sample as a reference to analyse changes in protein expression during AHI constitutes a major strength of this study, and allows interpretation of virus-induced alterations on the individual level. The longitudinal protein expression patterns were classified into six groups depending on how protein expressions changed from pre-infection levels: i.e. 1) decrease in protein expression following HIV-1 acquisition was either rapid and transient, more slowly initiated, or rapid and then resolved; 2) whereas the increase was either slowly initiated, persistent and large or, rapid and resolved; and the different proteins have been grouped in these categories including their respective functions. As part of the longitudinal trajectories, 241 of 1106 (21%) quantified plasma proteins changed significantly over the acute HIV-1 infection phase, revealing an extensive remodelling of the plasma proteome during early HIV-1 infection. It is crucial to establish the nature and timing of protein interactions during acute HIV-1 infection to inform prevention strategies, design prognostic markers for disease progression and severity, and further understand how the HIV-1 affects the host.

We also confirm that differences in host responses are associated with ARS. When we compared all proteins simultaneously, we identified a list of 20 proteins that were strongly associated with ARS. These findings point to a complex rearrangement of multiple factors involved in cell adhesion, binding and actin cytoskeleton, with many of these proteins decreasing beyond pre-infection levels after one month of infection. The top proteins associated with ARS were: zyxin (ZYX), secretoglobin family 1A member 1 (SCGB1A1). Moreover, Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 (SVEP1) and Serum amyloid A-4 were among the top proteins associated with viral control, where lower mean abundance levels of

these proteins were correlated with sustained viral control or lower viral load. SVEP1 protein structure suggested that SVEP1 potentially played a role in cellular adhesion and that the downregulation of its expression in vascular endothelial cells in patients with septic shock induces the release of chemokines such as IL-8, monocyte chemoattractant protein (MCP)-1, and MCP-3, which mediate the enhancement of cell chemotaxis[274, 275]. Therefore, having low levels of SVEP-1 could be correlated to lower or sustained VL. Finally, protein Kinase C beta type (PKCB) was among the top significant protein associated with the increased risk of clinical disease progression. PKCB stimulates Nuclear Factor-kappa-B (NF-κB) that regulates BCR-induced NF-κB B-cell activation, and also binds to the HIV-1 promoter to increase its transcription[276, 277]. In addition, PKCB plays a role in the rearrangement of the cytoskeleton that is required for viral entry[278]. The results point towards the importance of PKCB in HIV-1 replication, and it is possible that increased expression of PKCB might increase the risk of clinical progression.

Infant intra-host HIV-1 evolution and CTL responses during acute HIV-1

Main findings

- A single transmitted founder virus infected 78% of the infants (Paper V).
- Three months following infection, the immune-driven selection was observed in infants (Paper V).
- The majority of the identified CTL escape mutations were passed down from mothers, and did not reverse during the first year of life (Paper V).
- In infants, a higher HIV-1 evolutionary rate is linked to cytotoxic T lymphocyte escape mutations (Paper V).

It has previously been demonstrated that 67% of infants from the USA, and 54% of infants from Zambia were infected with a single transmitted founder virus[156, 279]. Furthermore, single transmitted founder viruses have been estimated to be 76-80% in heterosexual transmissions, as low as 62% in men who have sex with men, and 40% in

intravenous drug users[155, 157, 280, 281]. The percentage observed in our study is slightly higher than previously observed in other infant cohorts, but within the range of heterosexual transmission. There are two possible explanations for the observed discrepancies among the infants. The first is that we used a rigorous inclusion criterion, including only children with samples taken within one month of infection. The other studies used samples obtained more than three months after infection, implying that viruses that already had been under selection from infants' immunological pressure were included. The second is that the mothers in our study were given single-dose Nevirapine during pregnancy but not in the Women Infant Transmission Study (WITS) cohort in the USA.

HIV-1 specific $CD8^+$ T-cell responses are induced as early as the first week of life in 65% of *in utero* infected infants, with a median time of sample collection of nine days[136]. In the same cohort, these responses were detected in 52% of intrapartum infants with a median time of sample collection of 52 days (31-105 days) post-infection. Another South African cohort found $CD8^+$ T-cell responses as early as six weeks of age in infants[139]. In a previous sub-study of our cohort, $CD8^+$ T-cell responses were observed at 1 month in 58% of *in-utero* and 29% in peripartum infected Kenyan infants, and in approximately 50% of infants thereafter[282]. Over the first year of life, infants with IFN-γ release from $CD8^+$ T-cells at one month of age had peak plasma viral loads, rates of decline of viral load, and mortality risk similar to infants who lacked responses at one month of age. Therefore, the strength and breadth of IFN-γ responses at one month of age were not significantly associated with viral containment or mortality. However, the early presence of quantitatively higher Gag-specific $CD8^+$ T-cell responses has been associated with lower viral load and decreased mortality in the first year of life[140, 283]. Functional assessment of HIV-specific T-cells by multiparameter flow cytometry revealed that polyfunctional CD8+ T-cells were less prevalent in children before 24 months of age, and that HIV-specific CD4+ T-cell responses were of universally lower frequency among antiretroviral-naive children, and absent in young infants.

In addition, we found that viruses with CTL escape mutations evolved faster than those without, indicating that neonate $CD8^+$ T-cells exert selective pressure on the virus.

However, the magnitude of the cytotoxic specificity in killing HIV-1 remains to be quantified. No association was found between CTL escape mutations (observed from three months) and disease progression. The cross-sectional and longitudinal data suggested that, unlike HIV-1-infected adults who have strong cytotoxic T lymphocyte responses in the early stages of infection which are associated with reductions in viremia, HIV-1-infected infants generate HIV-1-specific $CD8^{+}$T-cell responses early in life that are not clearly associated with better clinical outcomes and that qualitative differences in the CD8 response, combined with a deficiency of HIV-specific CD4 cells, may contribute to the inability of young infants to limit replication of HIV[140, 282].

Conclusions, limitations and future perspectives

Our findings in **papers I and II** contribute to the growing body of evidence that HIV-1 transmission patterns vary by geographic location. This is primarily influenced by behavioural features, distinct migratory patterns, and the HIV-1 preventive measures available in each location. Particularly, in paper I, which was the first comprehensive HIV-1 phylodynamic research based on a countrywide set of HIV-1 sequences from Nigeria, we discovered a vast number of URFs and possibly novel CRFs. Further analysis indicated that HIV-1 first emerged and spread in major urban regions before relocating to smaller rural areas. In the early 2000s, the number of effective infections declined, corresponding with the introduction of free ART in Nigeria. However, the overall reliability of phylogenetic inference is restricted by sample size, patient demographics, and period of data collection, which results in missing links and smaller transmission clusters. Furthermore, despite the growing number of sequence data in public databases (e.g. the LANL HIV sequence database), these sequences still lack this information and are sometimes wrongly subtyped, therefore some caution should be used when employing sequences from these resources.

As a result, future HIV-1 phylogenetic research targeted at dissecting transmission dynamics must gather a bigger and proportional sample across all geographical areas, with an emphasis on time of collection and demographic information. Larger cohorts, such as the PANGEA-HIV consortium, established in five high HIV-1 prevalence countries, with dense homogeneous sampling, will improve and strengthen limitations from previous phylogenetic studies. In addition, they will provide a better picture or highlight where prevention efforts should be focused to most effectively reduce the HIV epidemic.[284].

In **papers III and IV**, we add to the existing data that i) early innate immune responses are elevated during acute HIV-1 infection (AHI) compared to pre-infection levels; ii) a second study that associates stronger early innate immune response, specifically IP-10 with the acute retroviral syndrome (ARS); iii) IP-10 is highly involved in AHI and

could be used to differentiate between stronger innate immune responses; iv) protein expression profiles and innate immune responses are highly individual specific highlighting the importance of longitudinal study designs that account for these individual differences; v) describe several proteins associated with HIV-1 disease progression, indicating that disease outcome could be driven by distinct virus- host interactions. However, there are a number of challenges associated with this type of research; for example, in plasma proteomics investigations, contamination of plasma by blood cellular components may induce systemic bias when quantifying the plasma proteome; ii) we used strict Fiebig I and II inclusion criteria to elucidate very early immune responses, but peak activation of analytes with delayed response may have been missed; iv) volunteers seeking hospital care are more likely to present with ARS than those from a routine longitudinal study; v) in paper III, the sample size was limited, and there was a e bias toward male volunteers (93%).

Taken together, it is clear that more research on the link between HIV-1 RNA levels, innate immune disturbances, and ARS is needed. Furthermore, bigger cohort validations of our findings in papers III and IV, as well as more sex balanced AHI investigations are required. On the other hand, our findings add to our understanding of how virus-host interactions during AHI interact with protein expression in the blood.

Paper V adds to the limited evidence on CTL-driven immune selection in HIV-1 vertically infected ART-naïve infants throughout the first 15 months of life. The immune-driven selection was found three months after infection, and the majority of reported CTL escape mutations were passed down from mothers, and did not reverse within the first year. Our work, like most intra-host HIV-1 evolution investigations, was limited by sample size, which could have contributed to the disparities in evolutionary rates when the samples were separated into groups. Nonetheless, this work contributed to our understanding of vertical HIV-1 transmission and how infant immune selection pressure shapes HIV-1 intra-host evolutionary dynamics. Our results support the notion that early immune-directed therapies may boost the efficacy of CTL responses in managing HIV-1 infections in infants.

Acknowledgements

Many people have contributed to this PhD dissertation in some way. I would, nevertheless, like to offer my heartfelt appreciation to the people named below.

Joakim Esbjörnsson, my main supervisor. Thank you for giving me this once-in-a-lifetime opportunity to work with you. You have opened several avenues for future collaborations and have assisted me in growing and becoming a better scientist. I haven't been a flawless PhD student, but you've helped me deal with a variety of situations, and I know that whatever success I achieve will never be attributed to my own efforts. I look forward to collaborating with you again in the future.

Amin S. Hassan, my co-supervisor. In 2018, we collaborated on synthesising over 200,000 PCRs, and it was around this time that I truly appreciated the craftsmanship you demonstrated in your work. Since then, you've helped me with everything from simple basic science thinking to hypothesis testing and complex proteomics interpretation.

Philippe Lemey, my co-supervisor. I first met you in 2009 as a fresh graduate on a phylogenetics course at the Uganda Virus Research Institute, and I believed you were the “king” of phylogeography at the time. After working with you on multiple projects, I believe you are the “god” of phylogeography. Thank you for always being accessible and responsive when I needed your guidance. I am eager to continue to collaborate on future projects.

Sarah Rowland Jones, my co-supervisor. I would not have worked with this unique infant HIV-1 cohort if you and Joakim had not collaborated on multiple projects. Thank you for allowing me to work not just with this content, but with your team as well. As a woman, seeing how respected and well-published you are in the HIV field has greatly inspired me.

Johan Malmström, my co-supervisor. It was a completely new experience for me when I started the proteomics project. We had a few introductory meetings during which you advised and introduced me to people who have continued to assist me to this day. This

project was a success thanks to you and your supportive team (Filip Årman, Melinda Rezeli, Christofer Karlson, Sven Kjellström, Tirthankar Mohanty, Hong Yan).

To the virology group PIs, it has been a privilege to work with colleagues I enjoy being around, who share similar research interests, and where everyone treats each other with respect, without judgement, and with a genuine want to help. Thank you, Patrik Medstrand and Marianne Jansson, for fostering an environment in which everyone can thrive as a scientist and, more importantly, as a person.

Eva Freyhult and Mun-Gwan Hong, the NBIS experts and my mentors. There hasn't been a single day since we started working together that I haven't been impressed with the degree of your work ethic and attention to detail. Mun-Gwan, who was my mentor under the NBIS programme and recommended various courses and resources for me to study, has substantially enhanced my R programming skills. I was inspired to be a better coder simply by seeing Eva Freyhult's R code, but I also learned a lot about statistics during our weekly sessions. Thank you for your motivation, effort, and readiness to help me learn so much more about proteomics data analysis than I did before meeting you.

Sara Karlson. So many wonderful things have transpired at the lab because you've been the foundation of a fully functional facility. You exemplify the qualities of a true friend: loyal, meticulous, caring, and loving. It's been a pleasure to sail with you on this PhD journey. I'm still looking forward to the wine celebration to commemorate our 4,000+ PCR plates.

Malin Neptin. I had a sense we'd get along great the first day you joined the team. We've always welcomed each other with a pleasant smile since that day. When I'm not feeling well, you inquire about me. Thank you for assisting me in the lab, for being more than just a co-worker, and for your contagious grin and lovely personality.

Emil Johansson. You are one of the kindest persons I have ever encountered in my life. You have been my immunology teacher, Swedish instructor, listener, moving help, discussion partner, and most importantly, a friend during the last four years.

George Makau. It was reassuring to know that there is someone in the same workplace who knew my hardships as a PhD student, who was willing to help and guide me

through these challenges and occasionally asked me to a BBQ to decompress. Thank you for including me in your life.

Zsófia Szojka, my personal photographer, a friend, a sister I never had. Thank you for all the support and time you invested to discuss several aspects of my PhD project.

Angelica Palm. You are the epitome of well balanced, and knowledgeable postdoc. You have been interested in knowing about the progress of my PhD studies and have always asked pertinent questions.

Sviataslau Sasinovich and Dawit Assefa: I'm still not over the "5-minute lunch" craze. Sviat, you've always been my go-to guy for any technical or software concerns. Dawit, your unparalleled knowledge of HIV-1 molecular epidemiology has made me want to always debate some difficult topics. Thank you very much for graciously assisting me through these difficulties.

To all current and previous PhD students, postdocs, master's students and colleagues in the virology team. Michael Boswell, Christian Bruhn, Malin Haglund, Fregenet Tesfaye, Oskar Olsson, Nordine Bakouche, Fariba Vaziri- Sani, thanks for the many thoughtful discussions,

Joseph Mukama, Mathew Lukenge, and Joseph Kazibwe, my family in Sweden. Every time we met and talked about life, I felt home. Thank you for allowing me to be a part of your lives and for your emotional support.

Hampus Martins, my best friend and partner. Life with you is just a blessing. I can never repay your kindness, support, and companionship. Weebale nyo, okumbelelawo!

Kassim Senjonyo, my brother. Thank you for always being there when I need you, for all the airport pickups and drop-offs, for running several errands for me and Isaac, and for taking care of our parents in my absence.

Eli Maama. Simanyi kyendikuuwa okulaga okusima kwange eli gyooli. Webaale nyo okulabilabilila mutabaani wange mukiseera wesomedde PhD. Webaale nyo o'kubeererawo mu business zange, webaale nyo o'kunzaala no kunsabila o'kutuuka ku kuula liino. Nsaba mukama akuwee o'buwanganzi.

Eli Taata. Oli muntu wabigambo bitini, naye buuli lwo yogeela nkuwuliliza ela nsimmidde daala buli kiimu kyonkolodde.Okusingila ddaala oku nyamba mu byo kuzimba, okuuza ”brother” woo, o’kulabilila emootoka noo kunzizangaamu amanyi buuli lwenkuuba esiimu. Nga bwobadde o’ngamba bulijjo, PhD bwegwa nga tunonya sente. Nsaba Allah akumpeele o’buwanganzi.

Isaac Anguria, the anchor of my life. You were the only person who kept me going when I was going through a hard patch. Being 10,000 kilometres apart from you has been the most difficult aspect of my PhD life, but thanks to Jaajas, we were able to make it work. You came into my life at a time when I needed you the most, and I aim to spend the rest of it by your side, nurturing you into a responsible and caring individual. Nkwagaala nyo mutabaani.

References

1. MS Gottlieb M, HM Schanker, MD, PT Fan, MD, A Saxon, MD, JD Weisman. **Pneumocystis Pneumonia --- Los Angeles**. *Morbidity and Mortality Weekly Report (MMWR)* 1981; 30(250):1-3.
2. Gallo RC. **The discovery of the first human retrovirus: HTLV-1 and HTLV-2**. *Retrovirology* 2005; 2:17.
3. Poiesz BJ, Ruscetti FW, Reitz MS, Kalyanaraman VS, Gallo RC. **Isolation of a new type C retrovirus (HTLV) in primary uncultured cells of a patient with Sézary T-cell leukaemia**. *Nature* 1981; 294(5838):268-271.
4. Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)**. 1983; 220(4599):868-871.
5. Klatzmann D, Barré-Sinoussi F, Nugeyre MT, Danquet C, Vilmer E, Griscelli C, et al. **Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes**. *Science (New York, NY)* 1984; 225(4657):59-63.
6. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T, et al. **T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV**. *Nature* 1984; 312(5996):767-768.
7. Rozenbaum W, Dormont D, Spire B, Vilmer E, Gentilini M, Griscelli C, et al. **Antimoniotungstate (HPA 23) treatment of three patients with AIDS and one with prodrome**. *Lancet (London, England)* 1985; 1(8426):450-451.
8. UNAIDS. **Start Free, Stay Free, AIDS Free**
In: *Final report on 2020 targets*; 2021.
9. Lupia R, Chien S-C. **HIV and AIDS Epidemic in Kenya: An Overview**. *Journal of Experimental & Clinical Medicine* 2012; 4(4):231-234.
10. Awofala AA, Ogundele OE. **HIV epidemiology in Nigeria**. *Saudi journal of biological sciences* 2018; 25(4):697-703.
11. Van De Perre P, Lepage P, Kestelyn P, Hekker A, Rouvroy D, Bogaerts J, et al. **Acquired Immunodeficiency Syndrome In Rwanda**. *The Lancet* 1984; 324(8394):62-65.
12. T. Ras GJ, Simson, I.W., Anderson, R., Prozesky, O.W., Hamersma. **Acquired immunodeficiency syndrome - a report of 2 South African cases**. 1983; 64(4):140-142.
13. Serwadda D, Mugerwa RD, Sewankambo NK, Lwegaba A, Carswell JW, Kirya GB, et al. **Slim disease: a new disease in Uganda and its association with HTLV-III infection**. *Lancet (London, England)* 1985; 2(8460):849-852.
14. Ryu W-S. **Chapter 17 - Retroviruses**. In: *Molecular Virology of Human Pathogenic Viruses*. Ryu W-S (editor). Boston: Academic Press; 2017. pp. 227-246.
15. Frankel AD, Young JA. **HIV-1: fifteen proteins and an RNA**. *Annual review of biochemistry* 1998; 67:1-25.
16. Hemelaar J. **The origin and diversity of the HIV-1 pandemic**. *Trends in molecular medicine* 2012; 18(3):182-192.
17. Motomura K, Chen J, Hu W-S. **Genetic Recombination between Human Immunodeficiency Virus Type 1 (HIV-1) and HIV-2, Two Distinct Human Lentiviruses**. *Journal of Virology* 2008; 82(4):1923-1933.
18. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. **Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes**. *Nature* 1999; 397(6718):436-441.
19. Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, et al. **Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (Cercocebus atys atys) from the Taï Forest, Côte d'Ivoire: Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2**. *Journal of Virology* 2005; 79(19):12515-12527.
20. Daniel MD, Letvin NL, King NW, Kannagi M, Sehgal PK, Hunt RD, et al. **Isolation of T-cell tropic HTLV-III-like retrovirus from macaques**. *Science (New York, NY)* 1985; 228(4704):1201-1204.
21. Apetrei C, Kaur A, Lerche NW, Metzger M, Pandrea I, Hardcastle J, et al. **Molecular Epidemiology of Simian Immunodeficiency Virus SIVsm in U.S. Primate Centers Unravels the Origin of SIVmac and SIVstm**. *Journal of Virology* 2005; 79(14):8991-9005.
22. Sharp PM, Hahn BH. **Origins of HIV and the AIDS Pandemic**. *Cold Spring Harbor Perspectives in Medicine* 2011; 1(1):a006841-a006841.
23. Apetrei C, Lerche NW, Pandrea I, Gormus B, Silvestri G, Kaur A, et al. **Kuru experiments triggered the emergence of pathogenic SIVmac**. *AIDS (London, England)* 2006; 20(3):317-321.
24. Sauter D, Kirchhoff F. **Key Viral Adaptations Preceding the AIDS Pandemic**. *Cell host & microbe* 2019; 25(1):27-38.
25. Gottlieb GS, Raugi DN, Smith RA. **90-90-90 for HIV-2? Ending the HIV-2 epidemic by enhancing care and clinical management of patients infected with HIV-2**. *The lancet HIV* 2018; 5(7):e390-e399.
26. Hemelaar J, Gouws E, Ghys PD, Osmanov S. **Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004**. 2006; 20(16):W13-W23.
27. Patiño-Galindo J, González-Candelas F. **The substitution rate of HIV-1 subtypes: a genomic approach**. *Virus evolution* 2017; 3(2):vex029.
28. Pennings PS, Kryazhimskiy S, Wakeley J. **Loss and recovery of genetic diversity in adapting populations of HIV**. *PLoS Genet* 2014; 10(1):e1004000.
29. Ayouba A, Akoua-Koffi C, Calvignac-Spencer S, Esteban A, Locatelli S, Li H, et al. **Evidence for continuing cross-species transmission of SIVsmm to humans**. *AIDS (London, England)* 2013; 27(15):2488-2491.

30. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. **HIV-1 nomenclature proposal**. *Science (New York, NY)* 2000; 288(5463):55-56.
31. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. **HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations**. *Science (New York, NY)* 2014; 346(6205):56-61.
32. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960**. *Nature* 2008; 455(7213):661-664.
33. Lemey P, Pybus OG, Rambaut A, Drummond AJ, Robertson DL, Roques P, et al. **The molecular population genetics of HIV-1 group O**. *Genetics* 2004; 167(3):1059-1068.
34. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. **Timing the ancestor of the HIV-1 pandemic strains**. *Science (New York, NY)* 2000; 288(5472):1789-1796.
35. Salemi M, Strimmer K, Hall WW, Duffy M, Delaporte E, Mboup S, et al. **Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2001; 15(2):276-278.
36. UNAIDS. **UNAIDS DATA 2021**. In; 2021.
37. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, et al. **Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis**. *The Lancet Infectious Diseases* 2019; 19(2):143-155.
38. Nazziwa J, Faria NR, Chaplin B, Rawizza H, Kanki P, Dakum P, et al. **Characterisation of HIV-1 Molecular Epidemiology in Nigeria: Origin, Diversity, Demography and Geographic Spread**. *Sci Rep* 2020; 10(1):3468.
39. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. **HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time**. *Science (New York, NY)* 1996; 271(5255):1582-1586.
40. Claeys E, Vermeire K. **The CD4 Receptor: An Indispensable Protein in T Cell Activation and A Promising Target for Immunosuppression.** . *Archives of Microbiology & Immunology* 2019; 3:133-150.
41. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. **Effects of CCR5 and CD4 Cell Surface Concentrations on Infections by Macrophagetropic Isolates of Human Immunodeficiency Virus Type 1**. 1998; 72(4):2855-2864.
42. Wilen CB, Tilton JC, Doms RW. **HIV: cell binding and entry**. *Cold Spring Harb Perspect Med* 2012; 2(8).
43. Geijtenbeek TBH, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GCF, Middel J, et al. **DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells**. *Cell* 2000; 100(5):587-597.
44. Arthos J, Cicala C, Martinelli E, Macleod K, Van Ryk D, Wei D, et al. **HIV-1 envelope protein binds to and signals through integrin α4β7, the gut mucosal homing receptor for peripheral T cells**. *Nature Immunology* 2008; 9(3):301-309.
45. Saphire ACS, Bobardt MD, Zhang Z, David G, Gallay PA. **Syndecans Serve as Attachment Receptors for Human Immunodeficiency Virus Type 1 on Macrophages**. 2001; 75(19):9187-9200.
46. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody**. *Nature* 1998; 393(6686):648-659.
47. Berger EA, Murphy PM, Farber JM. **CHEMOKINE RECEPTORS AS HIV-1 CORECEPTORS: Roles in Viral Entry, Tropism, and Disease**. *Annual Review of Immunology* 1999; 17(1):657-700.
48. Gao G, Wieczorek L, Peachman KK, Polonis VR, Alving CR, Rao M, et al. **Designing a Soluble Near Full-length HIV-1 gp41 Trimer**. *J Biol Chem* 2013; 288(1):234-246.
49. Berger EA, Doms RW, Fenyö E-M, Korber BTM, Littman DR, Moore JP, et al. **A new classification for HIV-1**. *Nature* 1998; 391(6664):240-240.
50. Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, et al. **Genotypic and phenotypic characterization of HIV-1 patients with primary infection**. *Science (New York, NY)* 1993; 261(5125):1179-1181.
51. Gilliam BL, Riedel DJ, Redfield RR. **Clinical use of CCR5 inhibitors in HIV and beyond**. *Journal of Translational Medicine* 2011; 9(S1):S9.
52. Jamjian MC, Mcnicholl IR. **Enfuvirtide: first fusion inhibitor for treatment of HIV infection**. *American Journal of Health-System Pharmacy* 2004; 61(12):1242-1247.
53. Kozal M, Aberg J, Pialoux G, Cahn P, Thompson M, Molina J-M, et al. **Fostemsavir in Adults with Multidrug-Resistant HIV-1 Infection**. 2020; 382(13):1232-1243.
54. Emu B, Fessel J, Schrader S, Kumar P, Richmond G, Win S, et al. **Phase 3 Study of Ibalizumab for Multidrug-Resistant HIV-1**. 2018; 379(7):645-654.
55. London R. **Structural Maturation of HIV-1 Reverse Transcriptase—A Metamorphic Solution to Genomic Instability**. *Viruses* 2016; 8(10):260.
56. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, et al. **Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition**. *J Mol Biol* 2009; 385(3):693-713.
57. Hu W-S, Hughes SH. **HIV-1 Reverse Transcription**. *Cold Spring Harbor Perspectives in Medicine* 2012; 2(10):a006882-a006882.
58. Jacques DA, Mcewan WA, Hilditch L, Price AJ, Towers GJ, James LC. **HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis**. *Nature* 2016; 536(7616):349-353.
59. Campbell EM, Hope TJ. **HIV-1 capsid: the multifaceted key player in HIV-1 infection**. *Nat Rev Microbiol* 2015; 13(8):471-483.
60. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, et al. **Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid**. *Cell* 1996; 87(7):1285-1294.

61. De Iaco A, Santoni F, Vannier A, Guipponi M, Antonarakis S, Luban J. **TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm**. *Retrovirology* 2013; 10(1):20.
62. Ambrose Z, Aiken C. **HIV-1 uncoating: connection to nuclear entry and regulation by host proteins**. *Virology* 2014; 454-455:371-379.
63. Rawle DJ, Harrich D. **Toward the "unravelling" of HIV: Host cell factors involved in HIV-1 core uncoating**. *PLoS pathogens* 2018; 14(10):e1007270.
64. Dick A, Cocklin S. **Recent Advances in HIV-1 Gag Inhibitor Design and Development**. *Molecules* 2020; 25(7):1687.
65. Doms RW, Trono D. **The plasma membrane as a combat zone in the HIV battlefield**. *Genes & development* 2000; 14(21):2677-2688.
66. Miller JT, Le Grice SFJ. **Reverse Transcriptase-Catalyzed HIV-1 DNA Synthesis**. In: *Encyclopedia of AIDS*. Hope TJ, Stevenson M, Richman D (editors). New York, NY: Springer New York; 2021. pp. 1-9.
67. Anderson EM, Maldarelli F. **The role of integration and clonal expansion in HIV infection: live long and prosper**. *Retrovirology* 2018; 15(1):71.
68. Lusic M, Siliciano RF. **Nuclear landscape of HIV-1 infection and integration**. *Nat Rev Microbiol* 2017; 15(2):69-82.
69. Savarino A. **In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate**. *Retrovirology* 2007; 4:21.
70. Lesbats P, Engelman AN, Cherepanov P. **Retroviral DNA Integration**. *Chemical reviews* 2016; 116(20):12730-12757.
71. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, Böni J, et al. **HIV rebounds from latently infected cells, rather than from continuing low-level replication**. *Proc Natl Acad Sci U S A* 2008; 105(43):16725-16730.
72. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, et al. **A role for LEDGF/p75 in targeting HIV DNA integration**. *Nature medicine* 2005; 11(12):1287-1289.
73. Sundquist WI, Kräusslich HG. **HIV-1 assembly, budding, and maturation**. *Cold Spring Harb Perspect Med* 2012; 2(7):a006924.
74. Martin-Serrano J, Neil SJ. **Host factors involved in retroviral budding and release**. *Nature reviews Microbiology* 2011; 9(7):519-531.
75. An W, Telesnitsky A. **HIV-1 genetic recombination: experimental approaches and observations**. *AIDS reviews* 2002; 4(4):195-212.
76. Onafuwa-Nuga A, Telesnitsky A. **The Remarkable Frequency of Human Immunodeficiency Virus Type 1 Genetic Recombination**. 2009; 73(3):451-480.
77. Punt JOJASSAJPP. **Kuby immunology**. 2019.
78. Delves PJ, Roitt IM. **The Immune System**. 2000; 343(1):37-49.
79. Rustagi A, Gale M. **Innate Antiviral Immune Signaling, Viral Evasion and Modulation by HIV-1**. *J Mol Biol* 2014; 426(6):1161-1177.
80. Banchereau J, Briere F, Caux C, Davoust J, Lebecque S, Liu YJ, et al. **Immunobiology of dendritic cells**. *Annu Rev Immunol* 2000; 18:767-811.
81. Blanco P, Palucka AK, Pascual V, Banchereau J. **Dendritic cells and cytokines in human inflammatory and autoimmune diseases**. *Cytokine & growth factor reviews* 2008; 19(1):41-52.
82. Stetson DB, Medzhitov R. **Type I Interferons in Host Defense**. *Immunity* 2006; 25(3):373-381.
83. Randall RE, Goodbourn S. **Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures**. 2008; 89(1):1-47.
84. Schoggins JW. **Interferon-stimulated genes: roles in viral pathogenesis**. *Curr Opin Virol* 2014; 6:40-46.
85. Schoggins JW, Rice CM. **Interferon-stimulated genes and their antiviral effector functions**. *Current Opinion in Virology* 2011; 1(6):519-525.
86. Horowitz A, Stegmann K, Riley E. **Activation of Natural Killer cells during microbial infections**. 2012; 2.
87. Kärre K. **Natural killer cell recognition of missing self**. *Nature Immunology* 2008; 9(5):477-480.
88. Lanier LL. **Up on the tightrope: natural killer cell activation and inhibition**. *Nature Immunology* 2008; 9(5):495-502.
89. Wren LH, Chung AW, Isitman G, Kelleher AD, Parsons MS, Amin J, et al. **Specific antibody-dependent cellular cytotoxicity responses associated with slow progression of HIV infection**. *Immunology* 2013; 138(2):116-123.
90. Vilches C, Parham P. **KIR: Diverse, Rapidly Evolving Receptors of Innate and Adaptive Immunity**. 2002; 20(1):217-251.
91. Paul S, Lal G. **The Molecular Mechanism of Natural Killer Cells Function and Its Importance in Cancer Immunotherapy**. 2017; 8.
92. Flórez-Álvarez L, Hernandez JC, Zapata W. **NK Cells in HIV-1 Infection: From Basic Science to Vaccine Strategies**. 2018; 9.
93. Gordon S, Taylor PR. **Monocyte and macrophage heterogeneity**. *Nature Reviews Immunology* 2005; 5(12):953-964.
94. Murray PJ, Wynn TA. **Protective and pathogenic functions of macrophage subsets**. *Nature Reviews Immunology* 2011; 11(11):723-737.
95. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, et al. **HIV evolution: CTL escape mutation and reversion after transmission**. *Nature medicine* 2004; 10(3):282-289.
96. Arango Duque G, Descoteaux A. **Macrophage Cytokines: Involvement in Immunity and Infectious Diseases**. 2014; 5.

97. Yagi Y, Watanabe E, Watari E, Shinya E, Satomi M, Takeshita T, et al. **Inhibition of DC-SIGN-mediated transmission of human immunodeficiency virus type 1 by Toll-like receptor 3 signalling in breast milk macrophages**. *Immunology* 2010; 130(4):597-607.
98. Lombardo E, Alvarez-Barrientos A, Maroto B, Boscá L, Knaus UG. **TLR4-Mediated Survival of Macrophages Is MyD88 Dependent and Requires TNF-α Autocrine Signalling**. 2007; 178(6):3731-3739.
99. Griffith JW, Sokol CL, Luster AD. **Chemokines and Chemokine Receptors: Positioning Cells for Host Defense and Immunity**. 2014; 32(1):659-702.
100. Zlotnik A, Yoshie O. **Chemokines: A New Classification System and Their Role in Immunity**. *Immunity* 2000; 12(2):121-127.
101. Wang J, Knaut H. **Chemokine signaling in development and disease**. *Development* 2014; 141(22):4199-4205.
102. Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, Heitman J, et al. **Induction of a Striking Systemic Cytokine Cascade prior to Peak Viremia in Acute Human Immunodeficiency Virus Type 1 Infection, in Contrast to More Modest and Delayed Responses in Acute Hepatitis B and C Virus Infections**. 2009; 83(8):3719-3733.
103. Ndhlovu Zaza M, Kamya P, Mewalal N, Kløverpris Henrik N, Nkosi T, Pretorius K, et al. **Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point**. *Immunity* 2015; 43(3):591-604.
104. Wang Z, Shang H, Jiang Y. **Chemokines and Chemokine Receptors: Accomplices for Human Immunodeficiency Virus Infection and Latency**. 2017; 8.
105. Jiao Y, Zhang T, Wang R, Zhang H, Huang X, Yin J, et al. **Plasma IP-10 is associated with rapid disease progression in early HIV-1 infection**. *Viral Immunol* 2012; 25(4):333-337.
106. Hassan AS, Hare J, Gounder K, Nazziwa J, Karlson S, Olsson L, et al. **A Stronger Innate Immune Response During Hyperacute Human Immunodeficiency Virus Type 1 (HIV-1) Infection Is Associated With Acute Retroviral Syndrome**. *Clinical Infectious Diseases* 2021; 73(5):832-841.
107. Sun JC, Williams MA, Bevan MJ. **CD4+ T cells are required for the maintenance, not programming, of memory CD8+ T cells after acute infection**. *Nature Immunology* 2004; 5(9):927-933.
108. Oh S, Perera LP, Terabe M, Ni L, Waldmann TA, Berzofsky JA. **IL-15 as a mediator of CD4⁺ help for CD8⁺ T cell longevity and avoidance of TRAIL-mediated apoptosis**. 2008; 105(13):5201-5206.
109. Perdomo-Celis F, Taborda NA, Rugeles MT. **CD8+ T-Cell Response to HIV Infection in the Era of Antiretroviral Therapy**. 2019; 10.
110. Weninger W, Manjunath N, Von Andrian UH. **Migration and differentiation of CD8+ T cells**. 2002; 186(1):221-233.
111. Lieberman J. **The ABCs of granule-mediated cytotoxicity: new weapons in the arsenal**. *Nature Reviews Immunology* 2003; 3(5):361-370.
112. Gulzar N, Copeland FTK. **CD8+ T-Cells: Function and Response to HIV Infection**. *Current HIV Research* 2004; 2(1):23-37.
113. Gonzalez SM, Taborda NA, Rugeles MT. **Role of Different Subpopulations of CD8+ T Cells during HIV Exposure and Infection**. 2017; 8.
114. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection**. 1994; 68(9):6103-6110.
115. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, et al. **Control of Viremia in Simian Immunodeficiency Virus Infection by CD8⁺ Lymphocytes**. 1999; 283(5403):857-860.
116. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, et al. **Dramatic Rise in Plasma Viremia after CD8+ T Cell Depletion in Simian Immunodeficiency Virus–infected Macaques**. *Journal of Experimental Medicine* 1999; 189(6):991-998.
117. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, et al. **HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells**. *Blood* 2006; 107(12):4781-4789.
118. Sáez-Cirión A, Lacabaratz C, Lambotte O, Versmisse P, Urrutia A, Boufassa F, et al. **HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection <i>ex vivo</i> and peculiar cytotoxic T lymphocyte activation phenotype**. 2007; 104(16):6776-6781.
119. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, et al. **Antiviral pressure exerted by HIV-l-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus**. *Nature medicine* 1997; 3(2):205-211.
120. Carrington M, O'Brien SJ. **The Influence of HLA Genotype on AIDS**. 2003; 54(1):535-551.
121. Bernardin F, Kong D, Peddada L, Baxter-Lowe LA, Delwart E. **Human immunodeficiency virus mutations during the first month of infection are preferentially found in known cytotoxic T-lymphocyte epitopes**. *J Virol* 2005; 79(17):11523-11528.
122. Collins DR, Urbach JM, Racenet ZJ, Arshad U, Power KA, Newman RM, et al. **Functional impairment of HIV-specific CD8+ T cells precedes aborted spontaneous control of viremia**. *Immunity* 2021; 54(10):2372-2384.e2377.
123. Moir S, Fauci AS. **B-cell responses to HIV infection**. 2017; 275(1):33-48.
124. McCoy LE, McKnight Á. **Lessons learned from humoral responses of HIV patients**. *Curr Opin HIV AIDS* 2017; 12(3):195-202.
125. Burton DR, Poignard P, Stanfield RL, Wilson IA. **Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses**. *Science (New York, NY)* 2012; 337(6091):183-186.
126. Burton DR, Hangartner L. **Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design**. *Annu Rev Immunol* 2016; 34:635-659.

127. Kollmann TR, Crabtree J, Rein-Weston A, Blimkie D, Thommai F, Wang XY, et al. **Neonatal Innate TLR-Mediated Responses Are Distinct from Those of Adults**. 2009; 183(11):7150-7160.
128. Levy O. **Innate immunity of the newborn: basic mechanisms and clinical correlates**. *Nature Reviews Immunology* 2007; 7(5):379-390.
129. Prendergast AJ, Klenerman P, Goulder PJR. **The impact of differential antiviral immunity in children and adults**. *Nature Reviews Immunology* 2012; 12(9):636-648.
130. Nakazawa T, Agematsu K, Yabuhara A. **Later development of Fas ligand-mediated cytotoxicity as compared with granule-mediated cytotoxicity during the maturation of natural killer cells**. *Immunology* 1997; 92(2):180-187.
131. Sundström Y, Nilsson C, Lilja G, Kärre K, Troye-Blomberg M, Berg L. **The expression of human natural killer cell receptors in early life**. *Scandinavian journal of immunology* 2007; 66(2-3):335-344.
132. Guilmot A, Hermann E, Braud VM, Carlier Y, Truyens C. **Natural Killer Cell Responses to Infections in Early Life**. *Journal of Innate Immunity* 2011; 3(3):280-288.
133. Ziegner U, Campbell D, Weinhold K, Frank I, Rutstein R, Starr SE. **Deficient Antibody-Dependent Cellular Cytotoxicity against Human Immunodeficiency Virus (HIV)-Expressing Target Cells in Perinatal HIV Infection**. 1999; 6(5):718-724.
134. Bernstein HB, Kinter AL, Jackson R, Fauci AS. **Neonatal Natural Killer Cells Produce Chemokines and Suppress HIV Replication in Vitro**. *AIDS Res Hum Retroviruses* 2004; 20(11):1189-1195.
135. Luzuriaga K, Holmes D, Hereema A, Wong J, Panicali DL, Sullivan JL. **HIV-1-specific cytotoxic T lymphocyte responses in the first year of life**. 1995; 154(1):433-443.
136. Thobakgale CF, Ramduth D, Reddy S, Mkhwanazi N, de Pierres C, Moodley E, et al. **Human immunodeficiency virus-specific CD8+ T-cell activity is detectable from birth in the majority of in utero-infected infants**. *J Virol* 2007; 81(23):12775-12784.
137. Mphatswe W, Blanckenberg N, Tudor-Williams G, Prendergast A, Thobakgale C, Mkhwanazi N, et al. **High frequency of rapid immunological progression in African infants infected in the era of perinatal HIV prophylaxis**. 2007; 21(10):1253-1261.
138. Goulder PJR, Brander C, Tang Y, Tremblay C, Colbert RA, Addo MM, et al. **Evolution and transmission of stable CTL escape mutations in HIV infection**. *Nature* 2001; 412(6844):334-338.
139. Shalekoff S, Meddows-Taylor S, Gray GE, Sherman GG, Coovadia AH, Kuhn L, et al. **Identification of human immunodeficiency virus-1 specific CD8+ and CD4+ T cell responses in perinatally-infected infants and their mothers**. 2009; 23(7):789-798.
140. Huang S, Dunkley-Thompson J, Tang Y, Macklin EA, Steel-Duncan J, Singh-Minott I, et al. **Deficiency of HIV-Gag-specific T cells in early childhood correlates with poor viral containment**. *Journal of immunology (Baltimore, Md : 1950)* 2008; 181(11):8103-8111.
141. Akinsiku OT, Bansal A, Sabbaj S, Heath SL, Goepfert PA. **Interleukin-2 Production by Polyfunctional HIV-1–Specific CD8 T Cells Is Associated With Enhanced Viral Suppression**. 2011; 58(2):132-140.
142. Barrios C, Brandt C, Berney M, Lambert P-H, Siegrist C-A. **Partial correction of the TH2/TH1 imbalance in neonatal murine responses to vaccine antigens through selective adjuvant effects**. 1996; 26(11):2666-2670.
143. Pantaleo G, Demarest JF, Schacker T, Vaccarezza M, Cohen OJ, Daucher M, et al. **The qualitative nature of the primary immune response to HIV infection is a prognosticator of disease progression independent of the initial level of plasma viremia**. *Proc Natl Acad Sci U S A* 1997; 94(1):254-258.
144. Cohen MS. **Preventing sexual transmission of HIV**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2007; 45 Suppl 4:S287-292.
145. Shen R, Richter HE, Smith PD. **Interactions between HIV-1 and mucosal cells in the female reproductive tract**. *American journal of reproductive immunology (New York, NY : 1989)* 2014; 71(6):608-617.
146. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. **Estimating per-act HIV transmission risk: a systematic review**. *AIDS (London, England)* 2014; 28(10):1509-1519.
147. Hladik F, Mcelrath MJ. **Setting the stage: host invasion by HIV**. *Nature Reviews Immunology* 2008; 8(6):447-457.
148. Gupta P, Collins KB, Ratner D, Watkins S, Naus GJ, Landers DV, et al. **Memory CD4(+) T cells are the earliest detectable human immunodeficiency virus type 1 (HIV-1)-infected cells in the female genital mucosal tissue during HIV-1 transmission in an organ culture system**. *J Virol* 2002; 76(19):9868-9876.
149. Hladik F, Sakchalathorn P, Ballweber L, Lentz G, Fialkow M, Eschenbach D, et al. **Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1**. *Immunity* 2007; 26(2):257-270.
150. Joseph SB, Swanstrom R, Kashuba ADM, Cohen MS. **Bottlenecks in HIV-1 transmission: insights from the study of founder viruses**. *Nat Rev Microbiol* 2015; 13(7):414-425.
151. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, et al. **Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission**. *Science (New York, NY)* 2004; 303(5666):2019-2022.
152. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, et al. **Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1**. *PLoS pathogens* 2009; 5(1):e1000274.
153. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, et al. **Phenotypic properties of transmitted founder HIV-1**. *Proc Natl Acad Sci U S A* 2013; 110(17):6626-6633.
154. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, et al. **Relative resistance of HIV-1 founder viruses to control by interferon-alpha**. *Retrovirology* 2013; 10:146.

155. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. **Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection**. *Proc Natl Acad Sci U S A* 2008; 105(21):7552-7557.
156. Kumar A, Smith CEP, Giorgi EE, Eudailey J, Martinez DR, Yusim K, et al. **Infant transmitted/founder HIV-1 viruses from peripartum transmission are neutralization resistant to paired maternal plasma**. *PLoS pathogens* 2018; 14(4):e1006944.
157. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. **High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men**. *PLoS pathogens* 2010; 6(5):e1000890.
158. Muema DM, Akilimali NA, Ndumnego OC, Rasehlo SS, Durgiah R, Ojwach DBA, et al. **Association between the cytokine storm, immune cell dynamics, and viral replicative capacity in hyperacute HIV infection**. *BMC Medicine* 2020; 18(1).
159. McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. **The immune response during acute HIV-1 infection: clues for vaccine development**. *Nature Reviews Immunology* 2010; 10(1):11-23.
160. Brenchley JM, Schacker TW, Ruff LE, Price DA, Taylor JH, Beilman GJ, et al. **CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract**. *The Journal of experimental medicine* 2004; 200(6):749-759.
161. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection**. *AIDS (London, England)* 2003; 17(13):1871-1879.
162. Lyles RH, Muñoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR, et al. **Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study**. *The Journal of infectious diseases* 2000; 181(3):872-880.
163. Robb ML, Eller LA, Kibuuka H, Rono K, Maganga L, Nitayaphan S, et al. **Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand**. *The New England journal of medicine* 2016; 374(22):2120-2130.
164. Tindall B, Barker S, Donovan B, Barnes T, Roberts J, Kronenberg C, et al. **Characterization of the acute clinical illness associated with human immunodeficiency virus infection**. *Archives of internal medicine* 1988; 148(4):945-949.
165. Sanders EJ, Price MA, Karita E, Kamali A, Kilembe W, Bekker LG, et al. **Differences in acute retroviral syndrome by HIV-1 subtype in a multicentre cohort study in Africa**. *AIDS (London, England)* 2017; 31(18):2541-2546.
166. Haase AT. **POPULATION BIOLOGY OF HIV-1 INFECTION: Viral and CD4+ T Cell Demographics and Dynamics in Lymphatic Tissues**. 1999; 17(1):625-656.
167. Stevenson M. **HIV-1 pathogenesis**. *Nature medicine* 2003; 9(7):853-860.
168. Dai J, Agosto LM, Baytop C, Yu JJ, Pace MJ, Liszewski MK, et al. **Human Immunodeficiency Virus Integrates Directly into Naïve Resting CD4⁺ T Cells but Enters Naïve Cells Less Efficiently than Memory Cells**. 2009; 83(9):4528-4537.
169. Chun T-W, Engel D, Berrey MM, Shea T, Corey L, Fauci AS. **Early establishment of a pool of latently infected, resting CD4 T cells during primary HIV-1 infection**. 1998; 95(15):8869-8873.
170. Blankson JN, Finzi D, Pierson TC, Sabundayo BP, Chadwick K, Margolick JB, et al. **Biphasic Decay of Latently Infected CD4+ T Cells in Acute Human Immunodeficiency Virus Type 1 Infection**. *The Journal of infectious diseases* 2000; 182(6):1636-1642.
171. Alter G, Teigen N, Ahern R, Streeck H, Meier A, Rosenberg ES, et al. **Evolution of Innate and Adaptive Effector Cell Functions during Acute HIV-1 Infection**. *The Journal of infectious diseases* 2007; 195(10):1452-1460.
172. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection**. *J Virol* 1994; 68(9):6103-6110.
173. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection**. *The Journal of experimental medicine* 2009; 206(6):1273-1289.
174. World Health Organization. **Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations: 2016 update**. *Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations: 2016 update* 2016.
175. Esbjörnsson J, Månsson F, Kvist A, da Silva ZJ, Andersson S, Fenyö EM, et al. **Long-term follow-up of HIV-2-related AIDS and mortality in Guinea-Bissau: a prospective open cohort study**. *The lancet HIV* 2018.
176. **Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe**. *Lancet (London, England)* 2000; 355(9210):1131-1137.
177. John GC, Nduati RW, Mbori-Ngacha DA, Richardson BA, Panteleeff D, Mwatha A, et al. **Correlates of mother-to-child human immunodeficiency virus type 1 (HIV-1) transmission: association with maternal plasma HIV-1 RNA load, genital HIV-1 DNA shedding, and breast infections**. *The Journal of infectious diseases* 2001; 183(2):206-212.
178. Luzuriaga K, Mofenson LM. **Challenges in the Elimination of Pediatric HIV-1 Infection**. *The New England journal of medicine* 2016; 374(8):761-770.
179. Tobin NH, Aldrovandi GM. **Immunology of pediatric HIV infection**. *Immunol Rev* 2013; 254(1):143-169.
180. Shalekoff S, Gray GE, Tiemessen CT. **Age-Related Changes in Expression of CXCR4 and CCR5 on Peripheral Blood Leukocytes from Uninfected Infants Born to Human Immunodeficiency Virus Type 1-Infected Mothers**. 2004; 11(1):229-234.

181. Bunders MJ, Van Der Loos CM, Klarenbeek PL, Van Hamme JL, Boer K, Wilde JCH, et al. **Memory CD4+CCR5+ T cells are abundantly present in the gut of newborn infants to facilitate mother-to-child transmission of HIV-1**. *Blood* 2012; 120(22):4383-4390.
182. Rouzioux C, Costagliola D, Burgard M, Blanche S, Mayaux MJ, Griscelli C, et al. **Estimated timing of mother-to-child human immunodeficiency virus type 1 (HIV-1) transmission by use of a Markov model. The HIV Infection in Newborns French Collaborative Study Group**. *American journal of epidemiology* 1995; 142(12):1330-1337.
183. Amin O, Powers J, Bricker KM, Chahroudi A. **Understanding Viral and Immune Interplay During Vertical Transmission of HIV: Implications for Cure**. 2021; 12.
184. Lagaye S, Derrien M, Menu E, Coïto C, Tresoldi E, Mauclère P, et al. **Cell-to-cell contact results in a selective translocation of maternal human immunodeficiency virus type 1 quasispecies across a trophoblastic barrier by both transcytosis and infection**. *J Virol* 2001; 75(10):4780-4791.
185. Ross AL, Cannou C, Barré-Sinoussi F, Menu E. **Proteasome-independent degradation of HIV-1 in naturally non-permissive human placental trophoblast cells**. *Retrovirology* 2009; 6:46.
186. Scarlatti G, Leitner T, Halapi E, Wahlberg J, Marchisio P, Clerici-Schoeller MA, et al. **Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers**. *Proc Natl Acad Sci U S A* 1993; 90(5):1721-1725.
187. Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, et al. **Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants**. *Science (New York, NY)* 1992; 255(5048):1134-1137.
188. Shearer WT, Quinn TC, LaRussa P, Lew JF, Mofenson L, Almy S, et al. **Viral Load and Disease Progression in Infants Infected with Human Immunodeficiency Virus Type 1**. 1997; 336(19):1337-1342.
189. Obimbo EM, Wamalwa D, Richardson B, Mbori-Ngacha D, Overbaugh J, Emery S, et al. **Pediatric HIV-1 in Kenya: pattern and correlates of viral load and association with mortality**. *Journal of acquired immune deficiency syndromes (1999)* 2009; 51(2):209-215.
190. Goulder PJ, Lewin SR, Leitman EM. **Paediatric HIV infection: the potential for cure**. *Nature Reviews Immunology* 2016; 16(4):259-271.
191. Muenchhoff M, Prendergast AJ, Goulder PJR. **Immunity to HIV in Early Life**. 2014; 5.
192. Newell ML, Coovadia H, Cortina-Borja M, Rollins N, Gaillard P, Dabis F. **Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis**. *Lancet (London, England)* 2004; 364(9441):1236-1243.
193. **Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis**. *The Lancet* 2000; 355(9210):1131-1137.
194. Gurdasani D, Iles L, Dillon DG, Young EH, Olson AD, Naranbhai V, et al. **A systematic review of definitions of extreme phenotypes of HIV control and progression**. 2014; 28(2):149-162.
195. de Wolf F, Spijkerman I, Schellekens PT, Langendam M, Kuiken C, Bakker M, et al. **AIDS prognosis based on HIV-1 RNA, CD4+ T-cell count and function: markers with reciprocal predictive value over time after seroconversion**. *AIDS (London, England)* 1997; 11(15):1799-1806.
196. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. **Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis**. *Proc Natl Acad Sci U S A* 2007; 104(44):17441-17446.
197. Kaleebu P, Nankya IL, Yirrell DL, Shafer LA, Kyosiimire-Lugemwa J, Lule DB, et al. **Relation Between Chemokine Receptor Use, Disease Stage, and HIV-1 Subtypes A and D: Results From a Rural Ugandan Cohort**. 2007; 45(1):28-33.
198. Wang B. **Viral Factors in Non-Progression**. 2013; 4.
199. Salvi R, Garbuglia AR, Caro AD, Pulciani S, Montella F, Benedetto A. **Grossly Defective <i>nef</i> Gene Sequences in a Human Immunodeficiency Virus Type 1-Seropositive Long-Term Nonprogressor**. 1998; 72(5):3646-3657.
200. Wang B, Spira TJ, Owen S, Lal RB, Saksena NK. **HIV-1 strains from a cohort of American subjects reveal the presence of a V2 region extension unique to slow progressors and non-progressors**. 2000; 14(3):213-223.
201. Claiborne DT, Prince JL, Scully E, Macharia G, Micci L, Lawson B, et al. **Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4⁺ T cells, and disease progression**. 2015; 112(12):E1480-E1489.
202. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. **Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma**. *J Virol* 2002; 76(5):2298-2305.
203. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. **Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS**. *Nature genetics* 2002; 31(4):429-434.
204. Gaudieri S, DeSantis D, McKinnon E, Moore C, Nolan D, Witt CS, et al. **Killer immunoglobulin-like receptors and HLA act both independently and synergistically to modify HIV disease progression**. *Genes and immunity* 2005; 6(8):683-690.
205. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, al. E. **The major genetic determinants of HIV-1 control affect HLA class I peptide presentation**. *Science (New York, NY)* 2010; 330(6010):1551-1557.
206. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, et al. **AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis**. *Nature medicine* 2005; 11(12):1290-1292.
207. Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, et al. **Copy number variation of KIR genes influences HIV-1 control**. *PLoS biology* 2011; 9(11):e1001208.

208. Meyer L, Magierowska M, Hubert JB, Rouzioux C, Deveau C, Sanson F, et al. **Early protective effect of CCR-5 delta 32 heterozygosity on HIV-1 disease progression: relationship with viral load. The SEROCO Study Group**. *AIDS (London, England)* 1997; 11(11):F73-78.
209. Hassan A, Esbjornsson J, Wahome E, Thiong'o A, Makau G, Price M, et al. **HIV-1 subtype diversity, transmission networks and transmitted drug resistance amongst acute and early infected MSM populations from Coastal Kenya**. *PLoS ONE* 2018; 13(12).
210. Robertson DL, Hahn BH, Sharp PM. **Recombination in AIDS viruses**. *J Mol Evol* 1995; 40(3):249-259.
211. Andrews SM, Zhang Y, Dong T, Rowland-Jones SL, Gupta S, Esbjörnsson J. **Analysis of HIV-1 envelope evolution suggests antibody-mediated selection of common epitopes among Chinese former plasma donors from a narrow-source outbreak**. *Sci Rep* 2018; 8(1):5743.
212. Arimide DA, Esquivel-Gómez LR, Kebede Y, Sasinovich S, Balcha T, Björkman P, et al. **Molecular Epidemiology and Transmission Dynamics of the HIV-1 Epidemic in Ethiopia: Epidemic Decline Coincided With Behavioral Interventions Before ART Scale-Up**. 2022; 13.
213. Pérez-Losada M, Jobes DV, Sinangil F, Crandall KA, Arenas M, Posada D, et al. **Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand**. *PLoS ONE* 2011; 6(3):e16902.
214. Esbjörnsson J, Mild M, Månsson F, Norrgren H, Medstrand P. **HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: origin, demography and migrations**. *PLoS ONE* 2011; 6(2):e17025.
215. Bbosa N, Ssemwanga D, Nsubuga RN, Salazar-Gonzalez JF, Salazar MG, Nanyonjo M, et al. **Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations**. *Sci Rep* 2019; 9(1):1-8.
216. Nduva GM, Otieno F, Kimani J, McKinnon LR, Cholette F, Sandstrom P, et al. **Phylogeographic Assessment Reveals Geographic Sources of HIV-1 Dissemination Among Men Who Have Sex With Men in Kenya**. 2022; 13.
217. Preston BD, Poiesz BJ, Loeb LA. **Fidelity of HIV-1 reverse transcriptase**. *Science (New York, NY)* 1988; 242(4882):1168-1171.
218. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, et al. **Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots**. *J Virol* 2002; 76(22):11273-11282.
219. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP. **High rate of recombination throughout the human immunodeficiency virus type 1 genome**. *J Virol* 2000; 74(3):1234-1240.
220. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. **Measurably evolving populations**. *Trends in Ecology & Evolution* 2003; 18(9):481-488.
221. Salemi M. **The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective**. *Infectious disease reports* 2013; 5(1S):3.
222. Leitner T. **Phylogenetics in HIV transmission: taking within-host diversity into account**. 2019; 14(3):181-187.
223. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. **Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection**. *Proceedings of the National Academy of Sciences* 2008; 105(21):7552-7557.
224. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. **Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus**. *PLoS pathogens* 2016; 12(5):e1005619.
225. Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, et al. **Modeling sequence evolution in acute HIV-1 infection**. *Journal of theoretical biology* 2009; 261(2):341-360.
226. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. **Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection**. *Journal of Virology* 1999; 73(12):10489-10502.
227. Norström MM, Buggert M, Tauriainen J, Hartogensis W, Prosperi MC, Wallet MA, et al. **Combination of immune and viral factors distinguishes low-risk versus high-risk HIV-1 disease progression in HLA-B*5701 subjects**. *J Virol* 2012; 86(18):9802-9816.
228. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. **Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics**. *PLoS Comput Biol* 2007; 3(2):e29.
229. Mild M, Gray R, Kvist A, Lemey P, Goodenow M, Fenyo E, et al. **High intrapatient HIV-1 evolutionary rate is associated with CCR5-to-CXCR4 coreceptor switch**. *Infection Genetics and Evolution* 2013; 19:369-377.
230. Rose R, Hall M, Redd AD, Lamers S, Barbier AE, Porcella SF, et al. **Phylogenetic Methods Inconsistently Predict the Direction of HIV Transmission Among Heterosexual Pairs in the HPTN 052 Cohort**. *The Journal of infectious diseases* 2019; 220(9):1406-1413.
231. Leitner T, Romero-Severson E. **Phylogenetic patterns recover known HIV epidemiological relationships and reveal common transmission of multiple variants**. *Nat Microbiol* 2018; 3(9):983-988.
232. Giardina F, Romero-Severson EO, Albert J, Britton T, Leitner T. **Inference of Transmission Network Structure from HIV Phylogenetic Trees**. *PLoS Comput Biol* 2017; 13(1):e1005316.
233. Volz EM, Romero-Severson E, Leitner T. **Phylodynamic Inference across Epidemic Scales**. *Mol Biol Evol* 2017; 34(5):1276-1288.
234. Mild M, Esbjornsson J, Fenyo E, Medstrand P. **Frequent intrapatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: Implications for coreceptor switch**. *Journal of Virology* 2007; 81(7):3369-3376.
235. Edo-Matas D, Lemey P, Tom JA, Serna-Bolea C, van den Blink AE, van 't Wout AB, et al. **Impact of CCR5delta32 host genetic background and disease progression on HIV-1 intrahost evolutionary processes: efficient**

hypothesis testing through hierarchical phylogenetic models. *Molecular biology and evolution* 2011; 28(5):1605-1616.
236. Suchard MA, Kitchen CMR, Sinsheimer JS, Weiss RE. **Hierarchical phylogenetic models for analyzing multipartite sequence data**. *Systematic biology* 2003; 52(5):649-664.
237. Mader SSWM. **Human biology**. 2020.
238. Anderson NL, Anderson NG. **The Human Plasma Proteome**. *Molecular & Cellular Proteomics* 2002; 1(11):845-867.
239. Deutsch EW, Omenn GS, Sun Z, Maes M, Pernemalm M, Palaniappan KK, et al. **Advances and Utility of the Human Plasma Proteome**. *Journal of Proteome Research* 2021; 20(12):5241-5263.
240. Geyer PE, Holdt LM, Teupser D, Mann M. **Revisiting biomarker discovery by plasma proteomics**. 2017; 13(9):942.
241. McFarlane I, McFarlane I, Bomford A, Sherwood R. **Liver disease and laboratory medicine**. London: ACB Venture Publications; 2000.
242. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, al. e. **Tissue-based map of the human proteome**. 2015; 347(6220):1260419.
243. Hortin GL, Jortani SA, Ritchie JC, Jr, Valdes R, Jr, Chan DW. **Proteomics: A New Diagnostic Frontier**. *Clinical Chemistry* 2006; 52(7):1218-1222.
244. Bruderer R, Muntel J, Müller S, Bernhardt OM, Gandhi T, Cominetti O, et al. **Analysis of 1508 Plasma Samples by Capillary-Flow Data-Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance**. *Mol Cell Proteomics MCP* 2019; 18(6):1242.
245. Keshishian H, Burgess MW, Gillette MA, Mertins P, Clauser KR, Mani DR, et al. **Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury**. *Mol Cell Proteomics* 2015; 14(9):2375.
246. Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA, et al. **Quantitative, Multiplexed Workflow for Deep Analysis of Human Blood Plasma and Biomarker Discovery by Mass Spectrometry**. *Nat Protoc* 2017; 12(8):1683.
247. Altelaar AFM, Munoz J, Heck AJR. **Next-generation proteomics: towards an integrative view of proteome dynamics**. *Nature Reviews Genetics* 2013; 14(1):35-48.
248. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis**. *Mol Cell Proteomics* 2012; 11(6):O111.016717.
249. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. **Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial**. *Molecular systems biology* 2018; 14(8):e8126.
250. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. **Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial**. 2018; 14(8):e8126.
251. Bernhardt OM, Selevsek N, Gillet L, Rinner O, Picotti P, Aebersold R, et al. **Spectronaut: a fast and efficient algorithm for MRM-like processing of data independent acquisition (SWATH-MS) data**. In: *60th American Society for Mass Spectometry Conference 2012*: F1000Posters 2014.
252. Consortium TU. **UniProt: the universal protein knowledgebase in 2021**. *Nucleic Acids Res* 2020; 49(D1):D480-D489.
253. Uhlén M, Karlsson MJ, Hober A, Svensson A-S, Scheffel J, Kotol D, et al. **The human secretome**. 2019; 12(609):eaaz0274.
254. Tarancón-Diez L, Rull A, Herrero P, Vazquez-Alejo E, Peraire J, Guillén S, et al. **Early antiretroviral therapy initiation effect on metabolic profile in vertically HIV-1-infected children**. *The Journal of antimicrobial chemotherapy* 2021; 76(11):2993-3001.
255. Al-Mozaini M, Alzahrani A, Alsharif I, Shinwari Z, Halim M, Alhokail A, et al. **Quantitative proteomics analysis reveals unique but overlapping protein signatures in HIV infections**. *Journal of infection and public health* 2021; 14(6):795-802.
256. Zhang W, Ambikan AT, Sperk M, van Domselaar R, Nowak P, Noyan K, et al. **Transcriptomics and Targeted Proteomics Analysis to Gain Insights Into the Immune-control Mechanisms of HIV-1 Infected Elite Controllers**. *EBioMedicine* 2018; 27:40-50.
257. Sperk M, Mikaeloff F, Svensson-Akusjärvi S, Krishnan S, Ponnan SM, Ambikan AT, et al. **Distinct lipid profile, low-level inflammation, and increased antioxidant defense signature in HIV-1 elite control status**. *iScience* 2021; 24(2):102111.
258. Lemma M, Petkov S, Bekele Y, Petros B, Howe R, Chiodi F. **Profiling of Inflammatory Proteins in Plasma of HIV-1-Infected Children Receiving Antiretroviral Therapy**. *Proteomes* 2020; 8(3).
259. Babu H, Sperk M, Ambikan AT, Rachel G, Viswanathan VK, Tripathy SP, et al. **Plasma Metabolic Signature and Abnormalities in HIV-Infected Individuals on Long-Term Successful Antiretroviral Therapy**. *Metabolites* 2019; 9(10).
260. Kramer HB, Lavender KJ, Qin L, Stacey AR, Liu MK, di Gleria K, et al. **Elevation of intact and proteolytic fragments of acute phase proteins constitutes the earliest systemic antiviral response in HIV-1 infection**. *PLoS pathogens* 2010; 6(5):e1000893.
261. Rodríguez-Gallego E, Tarancón-Diez L, García F, Del Romero J, Benito JM, Alba V, et al. **Proteomic Profile Associated With Loss of Spontaneous Human Immunodeficiency Virus Type 1 Elite Control**. *The Journal of infectious diseases* 2019; 219(6):867-876.
262. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, et al. **Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics**. *Journal of Proteome Research* 2010; 9(10):4982-4991.

263. Granger J, Siddiqui J, Copeland S, Remick D. **Albumin depletion of human plasma also removes low abundance proteins including the cytokines**. *Proteomics* 2005; 5(18):4713-4718.
264. Bellei E, Bergamini S, Monari E, Fantoni LI, Cuoghi A, Ozben T, et al. **High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins**. *Amino Acids* 2011; 40(1):145-156.
265. Etiebet MA, Shepherd J, Nowak RG, Charurat M, Chang H, Ajayi S, et al. **Tenofovir-based regimens associated with less drug resistance in HIV-1-infected Nigerians failing first-line antiretroviral therapy**. *AIDS (London, England)* 2013; 27(4):553-561.
266. Volz EM, Ndembi N, Nowak R, Kijak GH, Idoko J, Dakum P, et al. **Phylodynamic analysis to inform prevention efforts in mixed HIV epidemics**. *Virus evolution* 2017; 3(2):vex014.
267. Kamali A, Price MA, Lakhi S, Karita E, Inambao M, Sanders EJ, et al. **Creating an African HIV clinical research and prevention trials network: HIV prevalence, incidence and transmission**. *PLoS ONE* 2015; 10(1):e0116100.
268. Price MA, Kilembe W, Ruzagira E, Karita E, Inambao M, Sanders EJ, et al. **Cohort Profile: IAVI's HIV epidemiology and early infection cohort studies in Africa to support vaccine discovery**. *Int J Epidemiol* 2021; 50(1):29-30.
269. Dong KL, Moodley A, Kwon DS, Ghebremichael MS, Dong M, Ismail N, et al. **Detection and treatment of Fiebig stage I HIV-1 infection in young at-risk women in South Africa: a prospective cohort study**. *The lancet HIV* 2018; 5(1):e35-e44.
270. John-Stewart C, Grace, Mbori-Ngacha D, Payne L, Barbara, Farquhar C, Richardson A, Barbra, Emery S, et al. **HIV-1–Specific Cytotoxic T Lymphocytes and Breast Milk HIV-1 Transmission**. *The Journal of infectious diseases* 2009; 199(6):889-898.
271. Emery S, Bodrug S, Richardson BA, Giachetti C, Bott MA, Panteleeff D, et al. **Evaluation of performance of the Gen-Probe human immunodeficiency virus type 1 viral load assay using primary subtype A, C, and D isolates from Kenya**. *Journal of clinical microbiology* 2000; 38(7):2688-2695.
272. Chaplin B, Eisen G, Idoko J, Onwujekwe D, Idigbe E, Adewole I, et al. **Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy**. *AIDS Res Hum Retroviruses* 2011; 27(1):71-80.
273. Crowell TA, Colby DJ, Pinyakorn S, Fletcher JLK, Kroon E, Schuetz A, et al. **Acute Retroviral Syndrome Is Associated With High Viral Burden, CD4 Depletion, and Immune Activation in Systemic and Tissue Compartments**. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2018; 66(10):1540-1549.
274. Nakada TA, Russell JA, Boyd JH, Thair SA, Walley KR. **Identification of a nonsynonymous polymorphism in the SVEP1 gene associated with altered clinical outcomes in septic shock**. *Critical care medicine* 2015; 43(1):101-108.
275. Gilgès D, Vinit MA, Callebaut I, Coulombel L, Cacheux V, Romeo PH, et al. **Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor A domains**. *The Biochemical journal* 2000; 352 Pt 1(Pt 1):49-59.
276. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. **A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA**. *Cell* 1998; 92(4):451-462.
277. Griffin GE, Leung K, Folks TM, Kunkel S, Nabel GJ. **Activation of HIV gene expression during monocyte differentiation by induction of NF-kB**. *Nature* 1989; 339(6219):70-73.
278. Harmon B, Ratner L. **Induction of the Galpha(q) Signaling Cascade by the Human Immunodeficiency Virus Envelope Is Required for Virus Entry**. 2008; 82(18):9191-9205.
279. Nakamura KJ, Heath L, Sobrera ER, Wilkinson TA, Semrau K, Kankasa C, et al. **Breast milk and in utero transmission of HIV-1 select for envelope variants with unique molecular signatures**. *Retrovirology* 2017; 14(1):6.
280. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. **Wide variation in the multiplicity of HIV-1 infection among injection drug users**. *J Virol* 2010; 84(12):6241-6247.
281. Sagar M, Kirkegaard E, Long EM, Celum C, Buchbinder S, Daar ES, et al. **Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes**. *J Virol* 2004; 78(13):7279-7283.
282. Lohman BL, Slyker JA, Richardson BA, Farquhar C, Mabuka JM, Crudder C, et al. **Longitudinal Assessment of Human Immunodeficiency Virus Type 1 (HIV-1)-Specific Gamma Interferon Responses during the First Year of Life in HIV-1-Infected Infants**. *Journal of Virology* 2005; 79(13):8121-8130.
283. Nqoko B, Day CL, Mansoor N, De Kock M, Hughes EJ, Hawkridge T, et al. **HIV-Specific Gag Responses in Early Infancy Correlate with Clinical Outcome and Inversely with Viral Load**. *AIDS Res Hum Retroviruses* 2011; 27(12):1311-1316.
284. Abeler-Dörner L, Grabowski MK, Rambaut A, Pillay D, Fraser C. **PANGEA-HIV 2**. *Current Opinion in HIV and AIDS* 2019; 14(3):173-180.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

Dynamics of HIV-1 infection within and between hosts

JAMIRAH NAZZIWA graduated as a bioinformatician from Skövde University in 2015 and currently works at the Systems virology group, Department of Translational Medicine, Faculty of Medicine, Lund University.

A deep understanding of HIV-1 transmission, evolution, and the virus-host interactions during disease progression is needed to fast-track HIV-1 vaccine research. This doctoral thesis explores the transmission dynamics of HIV-1 within and between hosts, innate immune responses and longitudinal protein expression profiles during acute HIV-1 infection.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Translational Medicine
Systems Virology group

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:93
ISBN 978-91-8021-254-0
ISSN 1652-8220