Dissertation

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

For the degree of

Doctor of Natural Sciences

Presented by

MPhil Med. Microbiology, Yvonne Affram

born in Accra, Ghana.

Oral examination: 2nd March, 2016

# Intracellular DNA Levels of Drug Resistant and Drug Sensitive Human Immunodeficiency Virus Infections and its Correlation with Disease Progression Markers in a Population of Africans

Referees:

Prof. Dr. Hans-Georg Kräusslich

Prof. Dr. Stephan Urban

# Summary

There has been significant reduction in the scale of the global HIV pandemic over the last decade due to increased awareness and the effectiveness of Highly Active antiretroviral therapy (HAART). However, new challenges have emerged which when not effectively addressed may hinder the progress made so far. One important area of difficulty in the management of HIV infection is the emergence and transmission of drug resistant mutations (DRM), which threatens the long-term use of current HAART. Also, there is a growing need for more effective and relatively cheaper markers for monitoring HIV especially in resource-limited areas. One of such potential markers is HIV-l intracellular (IC) DNA load.

We analyzed plasma and buffy coat samples as well as clinical data of 86 HIV-1 infected drug naïve patients and eight follow-up patients with persistently high viral loads after 24 weeks of HAART from Nouna in Burkina Faso. The participants were predominantly females. Among drug naïve patients, paired RNA and DNA templates were polymerase chain reaction (PCR) amplified and Sanger sequenced for the detection of DRMs. Templates encompassed 1461 base pairs sequenced from the protease and reverse transcriptase region of HIV-1 pol. A total of eight patients harboured transmitted drug resistance mutations (TDRMs). Six had TDRMs to non-nucleoside reverse transcriptase inhibitors (NNRTIs), one to nucleoside reverse transcriptase inhibitors (NRTIs), and one to protease inhibitors (PIs). Given that interest is growing in the use of viral DNA sequencing to complement or replace RNA for DRM monitoring, we compared sequences from RNA and DNA templates, for similarities and differences in regions with DRMs and for nucleotide differences. A high level of concordance 78 (94%) was observed in regions with DRMs. Also, nucleotide sequences of paired templates were highly similar (84%). Furthermore, observed nucleotide differences greater than 10 within viral sequence pairs were resolved with deep sequencing using the so-called “Nextera tagmentation” approach. Deep sequencing of thirteen RNA and DNA template pairs revealed predominantly major reverse transcriptase (RT) mutations M230I and M184I within the minority viral population. It also confirmed differences in DRMs observed in Sanger sequencing and showed that patterns of synonymous and non-synonymous nucleotide changes were similar to those seen in Sanger sequencing. Predominant HIV-1 subtypes found in patients were CRF02_AG and CRF06_cpx. With the eight follow-up patients, DRMs and HIV-1 subtyping were determined from RNA templates only. Drug resistance mutations contributing to virological failure after HAART were assessed. Major DRMs to RTIs mainly K103N, E138Q, G190A and M230L were detected.

Finally we assessed the utility of HIV-1 IC DNA load as a tool for disease monitoring and found that HIV-1 IC DNA levels did not correlate with traditional HIV-1 disease monitoring markers such as $CD4^+$ T-cell counts ($r^2$=0.017; p=0.23) and plasma viral loads ($r^2$=0.003;

p=0.60), as well as other markers of disease progression. Also, there was no association (p=0.26) found between HIV-1 IC DNA levels of drug resistant (median 2.96 $\log_{10}$ copies/$10^6$ cells, IQR 2.30-3.52) and drug susceptible (median 2.62 $\log_{10}$ copies/$10^6$ cells, IQR 2.23-3.06) strains of HIV-1, signifying that the presence of TDRMs does not affect HIV-1 IC DNA levels. However, a significantly higher baseline HIV-1 IC DNA level (p=0.045) was found in patients failing HAART after 24 weeks of therapy (median 3.16 $\log_{10}$ copies/$10^6$ cells, IQR 2.75-3.62) as opposed to those who did not (median 2.63 $\log_{10}$ copies/$10^6$ cells, IQR, 2.12-3.04).

Our findings show that the prevalence of TDRMs is high and new DRMs develop over time making it necessary to institute resistance testing for more effective clinical management. Also the high concordance in DRMs between viral RNA and DNA templates suggests that DNA could be used for resistance monitoring, as it is cheaper and relatively easy to handle. HIV-1 IC DNA load is an independent marker that could be used alone or together with plasma viral load, $CD4^+$ T-cell counts and other markers to monitor disease progression. A larger follow-up study is recommended to confirm these findings.

# Zusammenfassung

Innerhalb des letzten Jahrzehnts haben sich die Krankheitslast und die Auswirkungen der globalen *HIV* (humanes Immundefizienz-Virus) Pandemie aufgrund des gesteigerten Interesses (z.B. die Aufnahme in die Millenium Development Goals der Vereinten Nationen) an der Erkrankung und der erhältlichen Therapie, bekannt als hochaktive antiretovirale Therapie (*HAART*), deutlich reduziert. Allerdings haben sich in dieser Zeit neue Herausforderungen herauskristallisiert, die, wenn sie nicht wirksam angegangen werden, die bisherigen Fortschritte kompromittieren. Ein großes Problem in der Behandlung der HIV-Infektion stellt die Entstehung und Übertragung von Mutationen (engl. *drug resitant mutations, DRM*) dar, die HI-Viren resistent gegen antiretrovirale Wirkstoffe machen und so den langfristigen Nutzen der aktuellen *HAART* beeinträchtigen. Auch gibt es einen zunehmenden Bedarf an effektiveren und verhältnismäßig billigeren Biomarkern zur Überwachung des Krankheitsverlaufs, insbesondere in Ländern mit unzureichenden finanziellen Mitteln. Ein solcher potenzieller Biomarker könnte die intrazelluläre DNA-Kopienzahl von *HIV-1* darstellen.

Wir analysierten Buffy-Coat- (Leukozytenfilm) und Plasmaproben mit den dazugehörigen klinischen Daten von 86 *HIV-1*-infizierten Therapie-naiven Patienten sowie longitudinale Daten von acht Patienten mit einer weiterhin hohen Viruslast nach einer 24wöchigen hochaktiven antiretoviralen Therapie in Nouna, Burkina Faso. In der Gruppe der Therapie-naiven Patienten amplifizierten wir je RNA- und DNA-Abschnitte (engl.: *templates*) mittels Polymerase-Kettenreaktion (PCR) und sequenzierten diese mit der Sanger-Methode zum Nachweis von *DRMs,* die Abschnitte umfassten 1461 Basenpaare aus der *HIV-1* pol-Region. Bei insgesamt acht Patienten konnten wir übertragbare Resistenzmutationen (engl. *transmitted drug resistance mutations, TDRMs*) nachweisen. Sechs hatten *TDRMs* für nicht-nukleosidische Reverse-Transkriptase-Hemmer (*NNRTI*), je ein Patient für nukleosidischen Reverse-Transkriptase-Hemmer (*NRTI*) und für Proteaseinhibitoren (*PIs*). Da ein großes Interesse daran besteht, die DNA-Sequenzierung des HI-Virus als Ergänzung oder gar als Ersatz für die aktuell durchgeführte Therapiekontrolle mittels RNA-Amplifikation zu verwenden, verglichen wir zusätzlich die Sequenzen von RNA und DNA-*templates* in den Regionen, die für Resistenzen gegen antiretrovirale Wirkstoffe kodieren. Bei 78 Proben (94%) haben wir ein hohes Maß an Übereinstimmung zwischen den beiden *templates* in Regionen mit *DRMs* beobachtet. Auch wiesen die Nucleotidsequenzen von Proben, die als RNA- und DNA-*template* vorlagen, große Ähnlichkeiten (84%) auf. Bei mehr als zehn unterschiedlichen Nukleotiden zwischen DNA- und RNA-*template* haben wir noch weiterführende Untersuchungen mittels deep sequencing, dem sogenannten „Nextera tagmentation“ Verfahren, durchgeführt. Diese zusätzlichen Untersuchungen bei dreizehn Paaren von RNA und DNA-*templates* ergaben überwiegend die *„major“* Mutationen M230I und M184I der reversen Transkriptase (Mutationen, die mit sehr wahrscheinlichen Auswirkungen auf die Resistenz gegen antiretrovirale Wirkstoffe assoziiert

sind) und die *„minor“* Protease-Inhibitor (PI) Mutationen V11I und L89I (Mutationen, die nur wenig zur Resistenzentstehung beitragen) innerhalb einer kleinen Anzahl von Viren der Gesamtviruspopulation. Hiermit wurden auch die Unterschiede in den Sequenzen der Sangermethode bestätigt und zeigte, dass die Muster von synonymen und nicht-synonymen Nukleotidänderungen große Ähnlichkeiten mit der Sanger-Sequenzierung aufzeigten. Die überwiegend gefundenen HIV-1-Subtypen unserer Patienten waren CRF02_AG und CRF06_cpx. Für unsere acht longitudinalen Proben wurden DRMs und HIV-1 Subtypen von RNA-*templates* ermittelt. Extrahierte cDNA Amplikons von Patienten dienten zur Untersuchungen der Entstehung von DRMs, die möglicherweise zum virologischen Versagen nach HAART beitrugen. Wir detektierten hauptsächlich die *„major“* Mutationen für RTIs: K103N, E138Q, G190A und M230L. Änderungen des dominierenden HIV-1-Subtyps wurde nicht beobachtet.

Zum Abschluss haben wir noch den Nutzen der HIV-1-IC DNA Viruslast zur Therapieüberwachung untersucht. Hier zeigte sich, dass diese Methode keine Korrelation mit den herkömmlichen Methoden wie CD4+T-Zellzahl, RNA-Viruslast im Plasma oder anderen Biomarker aufweist. Des Weiteren haben wir keinen Zusammenhang zwischen HIV-1-IC-DNA-Kopienzahl von resistenten und nicht-resistenten Virusstämmen von HIV-1 beobachtet. Dies könnte damit erklärt werden, dass die Anwesenheit von TDRMs die intrazelluläre HIV-1-DNA-Kopienzahl nicht beeinflusst. Allerdings haben wir bei Patienten mit virologischem Versagen nach 24 Wochen HAART signifikant höhere HIV-1 IC DNA-Kopienzahlen beobachtet, als bei Patienten ohne virologischen Versagen.

Unsere Ergebnisse zeigen, dass die Prävalenz von *TDRMs* bereits hoch ist und sich neue DRMs im Laufe der Zeit entwickeln. Dies erfordert wohl die Einführung von Resistenztestungen für ein effektiveres klinisches Management von HIV-infizierten Patienten. Die hohe Übereinstimmung zwischen viraler RNA und DNA in Bezug auf die Erkennung von DRMs könnte bedeuten, dass diese Methode mit viraler DNA zur Verlaufskontrolle von Resistenzen eingesetzt werden kann, zumal diese kostengünstig und relativ einfach in der Handhabung ist. Die HIV-1 IC DNA-Kopienzahl stellt einen unabhängigen Marker dar, der das Potenzial haben könnte, alleine oder in Kombination mit Plasma-Viruslast, CD4+T-Zellzahl und anderen Biomarkern in der klinischen Überwachungen einer HIV-Infektion verwendet zu werden. Hierfür sind jedoch noch weitere Untersuchungen an einem größeren Patientenkollektiv notwendig.

# Table of Content

# List of Abbreviations

| Number | Abbreviation | Full name |
|---|---|---|
| 1 | 3TC | Lamivudine |
| 3 | DDI | Didanosine |
| 4 | E | Glutamic acid |
| 5 | I | Isoleucine |
| 6 | IDV | Indinavir |
| 7 | L | Leucine |
| 8 | T | Threonine |
| 9 | TDF | Tenofovir |
| 10 | A | Alanine |
| 11 | aa | Amino acid |
| 12 | ABC | Abacavir |
| 13 | AIDS | Acquired Immune Deficiency Syndrome |
| 14 | ARDs | Antiretroviral drugs |
| 15 | ART | Antiretroviral therapy |
| 16 | ARV | Antiretroviral |
| 17 | AS | Allele specific |
| 18 | ATV | Atazanavir |
| 19 | AZT | Zidovudine |
| 20 | B2M | β2 microglobulin |
| 21 | BMI | Body mass index |
| 22 | bp | Base pair |
| 23 | C | Cysteine |
| 24 | CCR | C-C chemokine receptor |
| 25 | CD | Cluster of differentiation |
| 26 | CDC | Center for disease control and prevention |
| 27 | cDNA | Complementary Deoxyribonucleic acid |
| 28 | CHUYO | Center Hospital Universitaire Yalgado Ouedraogo |
| 29 | CRF | Circulating recombinant forms |
| 30 | CRP | C-reactive protein |

# List of Abbreviations continued

| Number | Abbreviation | Full name |
|---|---|---|
| 32 | CXCR | C-X-C chemokine receptor |
| 33 | D | Aspartic acid |
| 34 | D4T | Stavudine |
| 35 | DNA | Deoxyribonucleic acid |
| 36 | dNTPs | Deoxy-nucleoside-triphosphates |
| 37 | DR | Democratic republic |
| 38 | DRM | Drug resistance mutation |
| 39 | DRV | Darunavir |
| 40 | dsDNA | Double stranded Deoxyribonucleic acid |
| 41 | EDTA | Ethylenediaminetetraacetic acid |
| 42 | EFV | Efavirenz |
| 43 | ETR | Etravirin |
| 44 | F | Phenylalanine |
| 45 | FACS | Fluorescence activated cell sorter |
| 46 | FAM | Fluorescein |
| 47 | FDA | Food and drug administration |
| 48 | FPV | Fosamprenavir |
| 49 | FTC | Emtricitabine |
| 50 | G | Glycine |
| 51 | H | Histidine |
| 52 | HAART | Highly active antiretroviral therapy |
| 53 | Hb | Haemoglobin |
| 54 | HBV | Hepatitis B virus |
| 55 | HCL | Hydrochloric acid |
| 56 | HCV | Hepatitis C virus |
| 57 | HIV-1 | Human immunodeficiency virus type 1 |
| 58 | hrs | Hours |
| 59 | HTLV | Human T -cell lymphotropic virus |
| 60 | IC | Intracellular |
| 61 | IL-2R | Interleukin-2 receptor |
| 62 | IQR | Interquartile range |

# List of Abbreviations continued

| Number | Abbreviation | Full name |
|---|---|---|
| 63 | K | Lysine |
| 64 | kb | Kilo-base pair |
| 65 | LB | Lysogeny Broth |
| 66 | LB | Luria-Bertani |
| 67 | LPV | Lopinavir |
| 68 | LR | Low-level |
| 69 | LTR | Long terminal repeats |
| 70 | M | Molar concentration |
| 71 | M | Methionine |
| 72 | MB | Molecular Beacon |
| 73 | min | Minute |
| 74 | ml | Milliliter |
| 75 | MSM | Men who have sex with men |
| 76 | N | Asparagine |
| 77 | NCBI | National Center for Biotechnology Information |
| 78 | NFV | Nelfinavir |
| 79 | NGS | Next generation sequencing |
| 80 | NNRTI | Non-nucleoside reverse transcriptase inhibitor |
| 81 | NRTI | Nucleoside reverse transcriptase inhibitor |
| 82 | nt | Nucleotide |
| 83 | NTC | Non-template control |
| 84 | NVP | Nevirapine |
| 85 | OH | Hydroxyl |
| 86 | OLA | Oligonucleotide ligation assay |
| 87 | P | Proline |
| 88 | PBMC | Peripheral blood mononuclear cells |
| 89 | PCR | Polymerase chain reaction |
| 90 | PI | Protease inhibitors |
| 91 | PIC | Preintegration complex |
| 92 | pol | Gene coding viral enzymes |
| 93 | PR | Protease |

# List of Abbreviations continued

| Number | Abbreviation | Full name |
|---|---|---|
| 94 | PR/RT | Protease-reverse transcriptase |
| 95 | Q | Glutamine |
| 96 | qPCR | Quantitative polymerase chain reaction |
| 97 | r | Pearsons correlation coefficient |
| 98 | R | Arginine |
| 99 | $r^2$ | Coefficient of determination |
| 100 | RNA | Ribonucleic acid |
| 101 | RPV | Rilpivirine |
| 102 | RT | Reverse transcriptase |
| 103 | RT-PCR | Reverse transcriptase polymerase chain reaction |
| 104 | RTI | Reverse transcriptase inhibitor |
| 105 | s | Seconds |
| 106 | S | Serine |
| 107 | SGR | Second generation recombinants |
| 108 | SQV | Saquinavir |
| 109 | STS | Second template switch |
| 110 | TAE | Tris-acetate-EDTA buffer |
| 111 | TAMs | Thymidine analogue mutations |
| 112 | TDRM | Transmitted drug resistance mutation |
| 113 | TET | Tetrachloro-69-carbofluorescein |
| 114 | TPV | Tipranavir |
| 115 | UNAIDS | Joint United Nations program on HIV/AIDS |
| 116 | URF | Unique recombinant forms |
| 117 | V | Valine |
| 118 | W | Typtophan |
| 119 | WHO | World health organization |
| 120 | Y | Tyrosine |
| 121 | yrs | Years |

# List of tables

# List of figures

# Acknowledgements

Financial support for this study was obtained from the German Research Foundation (DFG) within the SFB 544, under the A6 project.

To the almighty God I say thank you lord for keeping, protecting and ordering my steps and to you be all the glory, honour, dominion and power forever and ever.

My profound gratitude goes to my supervisor Prof. Dr. Hans-Georg Kräusslich for giving me the opportunity to start this exciting project and steering it through to its completion. I also thank Prof. Dr. Stephan Urban for his interest in the project and also accepting the responsibility of being my supervisor. To Prof. Dr. Paul Schnitzler, I also say a very big thank you for all the wonderful scientific discussions and the editing of my doctoral thesis. It was great working with you. I further express my sincere gratitude to Dr. Steffen Geis for all his contribution to my bench work, discussions on how best to present figures and ideas and also for editing my thesis. I express my profound gratitude to you, Prof. Dr. Stefan Wiemann for being a part of my thesis advisory board committee and for all your scientific contributions that helped in steering the work to its completion and most especially accepting the challenge of finding skilled people to provide the needed service for me to carry out the deep sequencing experiments and also accepting the invitation to be my examiner. I further extend my sincere gratitude to Dr. Dirk Grimm for accepting the invitation to be part of my examination committee. I thank all the members of the diagnostic unit of the Virology laboratory especially Angelika Schillak for their contribution to various aspects of my work and also for allowing me to work with them. I am deeply indebted to the Deutscher Akademischer Austausch Dienst (DAAD) for offering me the scholarship that enabled me to come over to Deutschland and start my PhD.

I further extend my sincere gratitude to past and present members of the AG Kräusslich and the AG Müller groups especially to Fabrice Tiba for handing over available samples and clinical information to me and to the following persons Dr. Kathleen Börner, Dr. Susann Kummer, Robin Burk, Bianca Nouvertne, David Bejarano, Frauke Mücksch and Martin Obr for the time spent in helping with the editing of my thesis, I am most grateful. I also thank Anke-Mareil Heuser for all her support and friendship. I also thank my friends at the department

of Public Health who helped me with some of the statistical analysis especially Jacob Mazalale. I further extend my appreciation to Dr. Konrad Bode for helping me with my quantitative PCR troubleshooting and Prof. Dr. Alexander Dalpke for permitting me to use his quantitative PCR machine to run some of my experiments. I also thank all the patients and staff of the Centre de Recherche en Sante de Nouna (CRSN) in Burkina Faso especially Thierry Ouedraogo and Dr. Coulibaly for the transport of clinical samples.

Thanks to all my African family and friends in Heidelberg and Frankfurt, I say "Ayikoo". You guys have been my backbone of strength. You gave me something to laugh about whenever we met and you supported me with your prayers. You proved to me that a family is not only made up of blood relations but also of people who would make you feel wanted, loved, sheltered and comfortable even when the times are challenging. I thank my sister Faustina Kusi and family for all their love and support. I also extend my appreciation to Dr. Slyvestor Obiri Asamoah and family, Mr. David Dadson, Mrs Beatrice Tandoh, Mr. Kofi Anane for all your support. To my dear friends Dr. Mojdeh Heidary Navid and Eunice Okyere, knowing you was worth every minute spent together and I thank you guys for being there every step of the way.

To my family, I say I am very grateful to every one of you. To my dear husband Dr. Kwame Ofori Affram, I would like to use this platform to thank you for your invaluable support throughout my programme. You stood by me so I could step tall. You have been an emblem of strength to me and I love you. To my dear brother and friend Dr. Ernest Kenu, I say thank you for supporting me throughout my education. You have indeed been there to wash every dust off my feet and I thank you for supporting my studies. I also thank my parents and my siblings for all their love, prayers and support.

`

# 1 Introduction

## 1.1 The human immunodeficiency virus type 1

The Human Immunodeficiency Virus was first discovered in the early 1980s at which time it was referred to as lymphadenopathy associated virus (Barre-Sinoussi et al., 1983) and as well as human T -cell lymphotropic virus III (HTLVIII). It was shown to be the etiological factor in Acquired Immune Deficiency Syndrome (AIDS) (Broder and Gallo, 1984; Essex et al., 1985), a condition in humans which compromises the immune system and makes patients liable to life-threatening opportunistic infections (Coffin et al., 1986). HIV belongs to the retrovirus family and lentivirus group of viruses (Petropoulos, 1997).

## 1.2 Global epidemiology of HIV

HIV/AIDS has been a global pandemic for the past 3 to 4 decades (Maartens et al., 2014; Piot and Quinn, 2013). In comparison to other global pandemics, HIV/AIDS has been the largest with an estimated 78 million infected since the initial case diagnosis and an estimated 39 million deaths (UNAIDS, 2013b, 2014b). Since the advent of antiretroviral drugs (ARDs), the dynamics of HIV epidemiology has significantly changed with reducing incidence and increasing prevalence as mortality has drastically reduced (Maartens et al., 2014; Piot and Quinn, 2013). The reason for this pattern of epidemiological change is that, while ARVs are effective in reducing viral load and preventing full blown AIDS, so far, they are unable to completely eradicate the virus from the system of a sufferer (Piot and Quinn, 2013). The incidence of HIV infections has fallen by 33% since 2001 (Maartens et al., 2014). According to the UNAIDS HIV epidemic update report in 2013 as indicated in figure 1, the worldwide estimation of people living with HIV at the end of 2012 were 35.3 million, an increase from 31 million within a decade. Sub-Saharan Africa has the largest population of infected people with an estimated 25.0 million people presently living with HIV infection (Maartens et al., 2014; UNAIDS, 2013b, 2014b). Also in 2013, a total of 1.6 million new HIV infections and 1.2 million deaths among infected adults and children in sub-Saharan Africa were reported (UNAIDS, 2013b).

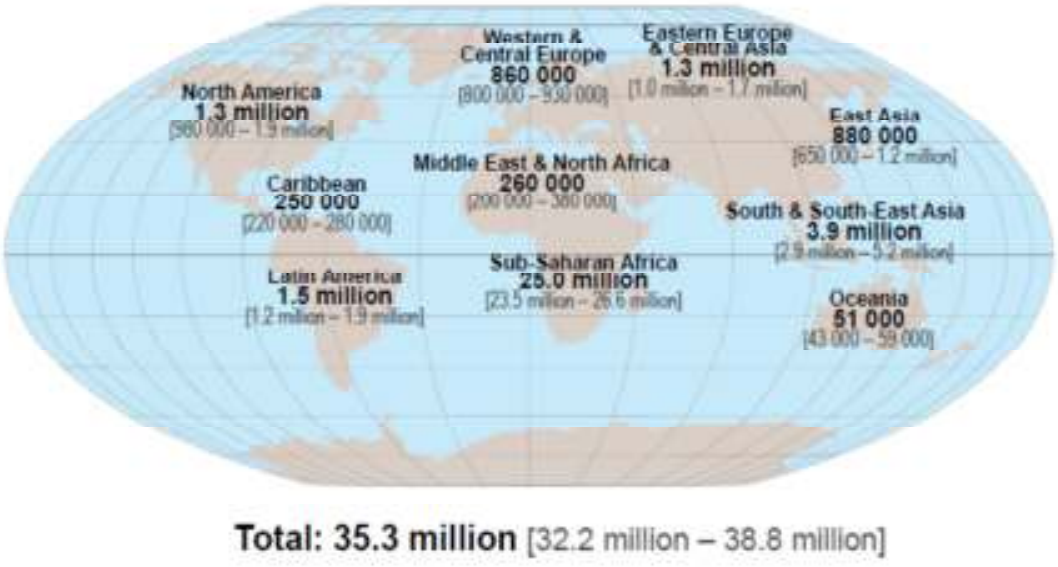



**Figure 1: Worldwide HIV epidemic update in 2013.**

**Reproduced from https://www.aids.gov/images/global-hivaids-overview.jpg.**

### 1.2.1 Worldwide distribution of HIV-1

Due to its heterogeneity, HIV-1 has been subdivided into three major groups, M, N and O. Of the groups, group M accounts for majority of global infections. Group M is further divided into 10 subtypes or clades named subtypes A-K (Buonaguro et al., 2007). The distribution of the major HIV-1 subtypes across the globe varies widely (Abecasis et al., 2013; Lau and Wong, 2013; Lihana et al., 2012). Global trends have shown changing patterns of recombination and spread of strains since the inception of the pandemic thus making it imperative that a keen interest should be taken to understand the dynamics of global subtype diversity (Lau and Wong, 2013; Lihana et al., 2012). HIV-1 subtype diversity can affect viral fitness and pathogenicity and pose difficulties for vaccine development and optimal therapy (Lau and Wong, 2013; Lihana et al., 2012). Subtypes differ from each other by an average nucleotide percentage ranging from 25–35% while intrasubtype variations are between 15-20% (Lau and Wong, 2013). The classification and distribution of HIV-1 is made more complex by the existence of many recombinant forms between the various clades of viruses. Circulating recombinant forms (CRFs) are defined as those viruses that have been fully sequenced and found in 3 or more

epidemiologically unlinked individuals (Hemelaar, 2012). Currently, there are 72 CRFs for HIV-1. The recombined viruses that do not fall in the criteria of CRFs are called unique recombinant forms (URFs). Also recombination may occur between CRFs and these are called second-generation recombinants (SGRs) (Hemelaar, 2012).

The West and Central African sub-regions harbour the widest diversity of HIV-1 strains however, only a few subtypes and recombinant forms account for most infections worldwide (Lihana et al., 2012). The important global subtypes include A, B, C, D, G and CRFs (CRF01_AE, CRF02_AG and CRF07_BC) (Lau and Wong, 2013; Lihana et al., 2012). In Europe, North America, Australia as well as parts of Asia and North Africa, Subtype B is the predominant HIV-1 subtype (Hemelaar, 2012). Subtype B is predominantly transmitted among homosexual men. It is therefore not surprising that a male to female ratio of 3:2 prevails in Europe (WHO, 2013b). The other major subtypes that circulate in Europe are A, C and G (Abecasis et al., 2013; Hemelaar et al., 2011) while the main circulating recombinant forms found in Europe includes CRF02_AG and CRFO2_AE (Abecasis et al., 2013; Hemelaar et al., 2011).

All the subtypes of HIV-1 have been reported in Africa (Buonaguro et al., 2007; Hemelaar et al., 2011) with significant regional variation in regional distribution (Lihana et al., 2012). There is also a high recombination rate with increasing circulating and unique recombinant forms especially in West and Central Africa, the epicenter of HIV-1 diversity (Lihana et al., 2012). The HIV-1 epidemic in Africa is most severe in Southern Africa with the predominant strain being subtype C (Buonaguro et al., 2007). The predominance of subtype C in southern Africa and in India has contributed to subtype C becoming the most predominant HIV-1 strain worldwide (Buonaguro et al., 2007; Neogi et al., 2012). Transmission is driven by heterosexual intercourse in Africa with more females infected than males (Muula, 2008). In West Africa, CRF02_AG (Buonaguro et al., 2007) is the dominant strain of HIV-1 involved in the epidemic with subtype G being the next most dominant (Hemelaar, 2012). Central Africa has the highest diversity of HIV-1 globally with the dominant strain being CRF01_AE. CRF01_AE originated in Central Africa and is becoming the most dominant recombinant form globally (Lau and Wong, 2013). Subtype A is the predominant HIV-1 subtype in East Africa followed by

subtype D (Buonaguro et al., 2007; Hemelaar, 2012). In south east Asia, CRF01_AE has been shown to be the predominant form of HIV-1 (figure 2) (Buonaguro et al., 2007).

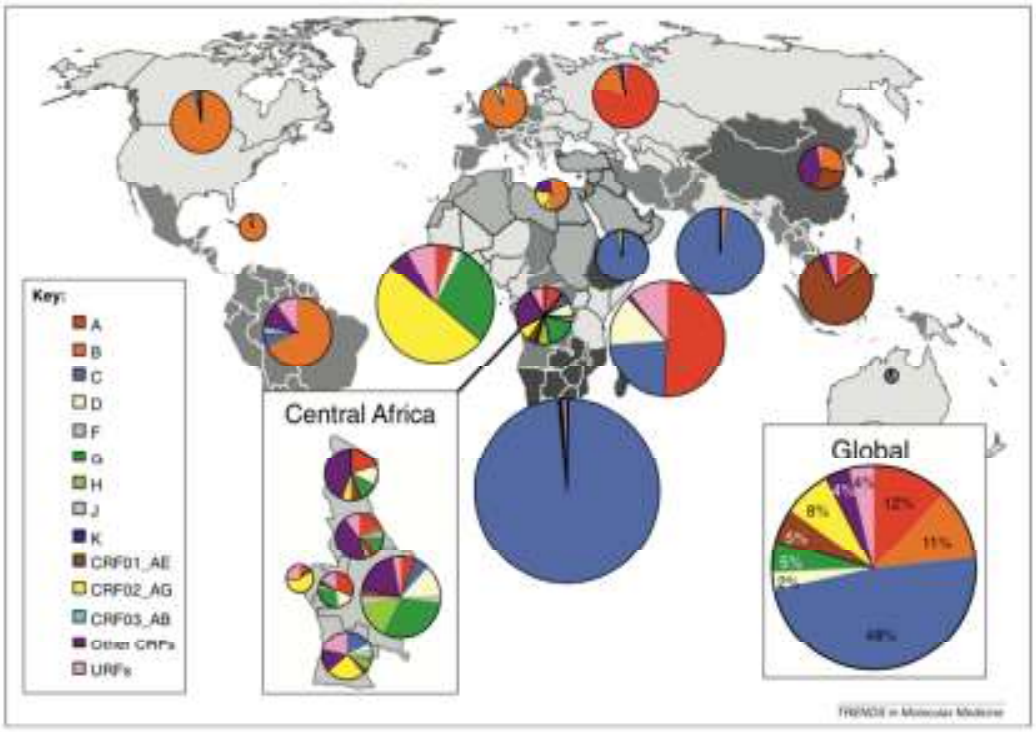



**Figure 2: Distribution of HIV-1 subtypes around the world.**

**Pie chart representing the distribution of major HIV subtypes in the various continents of the world Reproduced from (Hemelaar, 2012).**

## 1.3 HIV-1 genome structure and organization

HIV-1 is composed at its core of two copies of positive sense single-stranded RNA, which constitutes its genome, and associated proteins enclosed by a fullerene conical capsid (Dudek, 2007). The capsid is further enclosed by a matrix, which has on the outer surface the viral envelope (env) protein (Dudek, 2007). The genome of HIV-1 is quite complex with multiple reading frames (Cohen et al., 1990), splice sights (Mueller et al., 2014), internal ribosomal entry sites and pseudoknots (Watts et al., 2009). The RNA is about 9.5 kilobases (Li et al., 1992; Ratner et al., 1985) with a 5′ cap and a poly A tail (Ratner et al., 1985). The RNA genome comprises of major genes as well as minor genes, as is characteristic of retroviruses (Watts et al., 2009). The three major genes code for gag pol and env proteins, responsible for

structure, enzymes and envelope protein synthesis respectively. These major genes are unspliced except to for env (Lutzelberger et al., 2006; Watts et al., 2009). On the other hand, minor and accessory proteins including vif, vpu and nef are expressed through a complicated system of alternate splicing of the genes ((Lutzelberger et al., 2006; Schwartz et al., 1990). The 5' and 3' ends of HIV-1 proviral DNA have repeat sequences called long terminal repeats (Ferrer et al.). LTR sites are important for binding of primers during proviral transcription and integration of viral DNA into the host genome (Delviks-Frankenberry et al., 2011). They also serve as promoter and enhancer regions for viral RNA synthesis (Delviks-Frankenberry et al., 2011). LTRs of HIV-1 are segmented into U3, R and U5 regions at both the 5' and 3' regions of the proviral DNA. While the LTR regions are complete in the proviral DNA, the HIV-1 viral RNA only contains the R and U5 segments at the 5' end and U3 and R at the 3' ends (figure 5) (Delviks-Frankenberry et al., 2011).

Proteins that make up the viral particle are generally divided into structural, enzyme and accessory proteins. Structural proteins facilitate the integrity of the virus particle, support the internal organization and facilitate anchoring and infection of host cells (Frankel and Young, 1998). Structural proteins include matrix proteins, capsid proteins and membrane proteins. The matrix proteins form a layer on the inner side of the membrane and help to organize the transmembrane part of the envelope proteins as well as trafficking viral replication particles early in its life cycle (Hill et al., 1996). Capsid proteins deliver the viral RNA into cells during infection (Pornillos et al., 2009). Envelope proteins gp41 and gp120 ensure viral binding and penetration into host cells (Kwong et al., 2000). The protein p7 associates and supports the RNA genome (Dawson and Yu, 1998) while p6 is an important membrane interacting protein for viral assembly (figure 3) (Solbak et al., 2013).

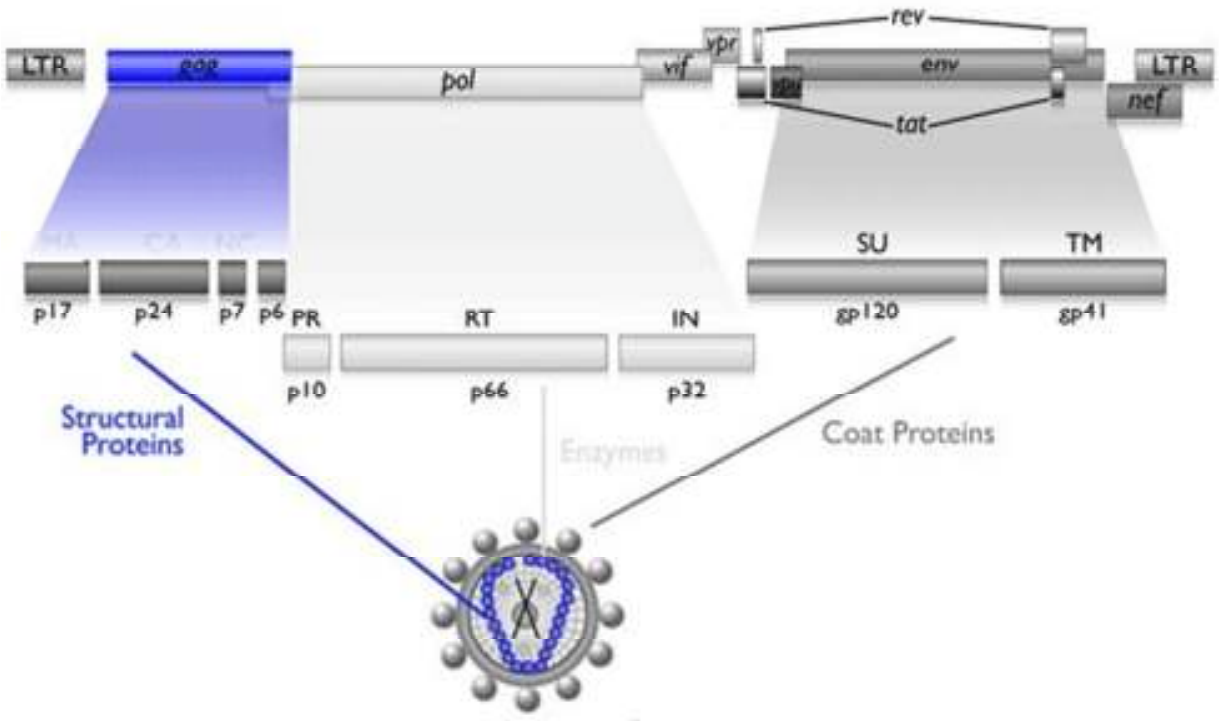



**Figure 3: Genome organization of HIV-1.**

**The figure illustrates the HIV genome above with the corresponding components of the viral particle they form below. (Reproduced from http://www.sciencedaily.com/releases/2011/12/111221140349.htm).**

## 1.4 Clinical presentation of HIV

The clinical presentation of HIV/AIDS varies widely from persistently asymptomatic to a severe immune compromise and wasting with many opportunistic infections and ultimately death (WHO, 2007). For clinical and research purposes HIV-1 infection has been classified into stages by the CDC and WHO. Staging of the disease is important for timing of management strategies and monitoring of disease progression during therapy as well as for research and epidemiological purposes (Schneider et al., 2008).

### 1.4.1 WHO clinical staging

The WHO staging of HIV/AIDS is based essentially on clinical manifestations of the disease (Kassa et al., 1999; WHO, 1993, 2007). There is also an immunological classification based on $CD4^+$ T-cell count (WHO, 2007). The clinical emphasis enables clinicians in resource

poor environments, where many laboratory parameters cannot be readily assessed, to effectively stage the disease and make good judgment on initiation of opportunistic disease prophylaxis and ART (WHO, 2007). WHO has divided HIV-1 infection into four clinical stages (1-4) based on the severity of clinically staged HIV-1 infection, according to the following categories: Asymptomatic, mild symptoms, advanced symptoms and severe symptoms which generally coincide to four stages (WHO, 2007).

### 1.4.2 CDC classification system of HIV-1 infection

While the WHO classification is essential for clinical management and patient care, the CDC staging is essentially based on measurement of $CD4^+$ T-cell counts with other associated clinical conditions (CDC, 1992; Schneider et al., 2008). Case definitions based on the CDC criteria are mainly oriented at public health surveillance, epidemiologic monitoring and control rather than for clinical management (Schneider et al., 2008). The CDC stages are also further classified based on age groups, namely; adults, adolescents and children (Schneider et al., 2008). Initial CDC classification for adults and adolescents is in three categories (1, 2 and 3) based on $CD4^+$ T-cell counts. Category 1 represents laboratory diagnoses of HIV-1 infection with $CD4^+$ T-cell counts greater than 500 cells/µl of blood. $CD4^+$ T-cell counts between 499 and 200 cells/µl of blood and counts less than 200 cells/µl of blood represent categories 2 and 3 respectively (CDC, 1992). The $CD4^+$ T-cell counts categories were further placed under 3 clinical groups categorized as A, B and C (CDC, 1992).

The CDC revised the classification using a simpler classification and retaining all the AIDS defining conditions of category C. The stages are as follows: stage 1, stage 2, stage 3 and stage unknown (Schneider et al., 2008). Since the $CD4^+$ T-cell count or percentage are critical for CDC staging, it is recommended that every effort should be made to acquire $CD4^+$ T-cell count information although the stage can still be classified as unknown. CDC stage 1 is defined as laboratory diagnosed HIV-1 infection with a $CD4^+$ T-cell count greater than 500 cells/µl of blood or CD4 percentage of greater than twenty-nine. Stage 2 refers to $CD4^+$ T-cell counts between 200 and 499 cells/µl of blood or a CD4 percentage between 14 and 28. In stage 3, the $CD4^+$ T-cell count is less than 200 cells/µl of blood or a CD4 percentage of less than 14 (Schneider et al., 2008).

## 1.5 Life cycle

HIV-1 gains access to the body mainly via parenteral routes. By far most of the cases in the global pandemic were transmitted through heterosexual intercourse (Maartens et al., 2014). On gaining entrance into the body, HIV-1 primarily infects mononuclear cells (Fu et al., 2011). The cells mainly targeted by HIV-1 are $CD4^+$T helper cells, monocytes, macrophages and dendritic cells (Coleman and Wu, 2009). These cells express CD4 on their cell surface, which serves as the major receptor for the viral membrane protein gp120 (Weiss, 2013). Besides CD4, co-receptors that facilitate viral attachment and tropism include the chemokine receptors CCR5 and CXCR4 (Weiss, 2013). CCR5 facilitates viral entry early during infection while CXCR4 is thought to be more important in the later stages of the infection (Weiss, 2013). However some HIV-1 strains can utilize both CCR5 and CXCR4 and are referred to as R5X4 viruses (Lamers et al., 2008).

Upon receptor recognition and attachment using gp120, the HIV-1 surface glycoprotein gp41 is important in membrane fusion and eventual entry onto the cells (Garg and Blumenthal, 2008). Within the cytoplasm of infected cells, the HIV-1 genome is reverse transcribed by reverse transcriptase to DNA (Hu and Hughes, 2012). Reverse transcription is a complicated process that involves two obligated template-switching steps involving both the minus and plus strands of the synthesized DNA (Delviks-Frankenberry et al., 2011).

The resulting HIV-1 DNA product of reverse transcription is referred to as a pre-integration complex (PIC), which is actively transported into the nucleus for integration into the genome as provirus (Engelman, 2009; Piller et al., 2003). Integration is effected by the virus encoded protein integrase (Albanese et al., 2008). Figure 4 illustrates the entire process.

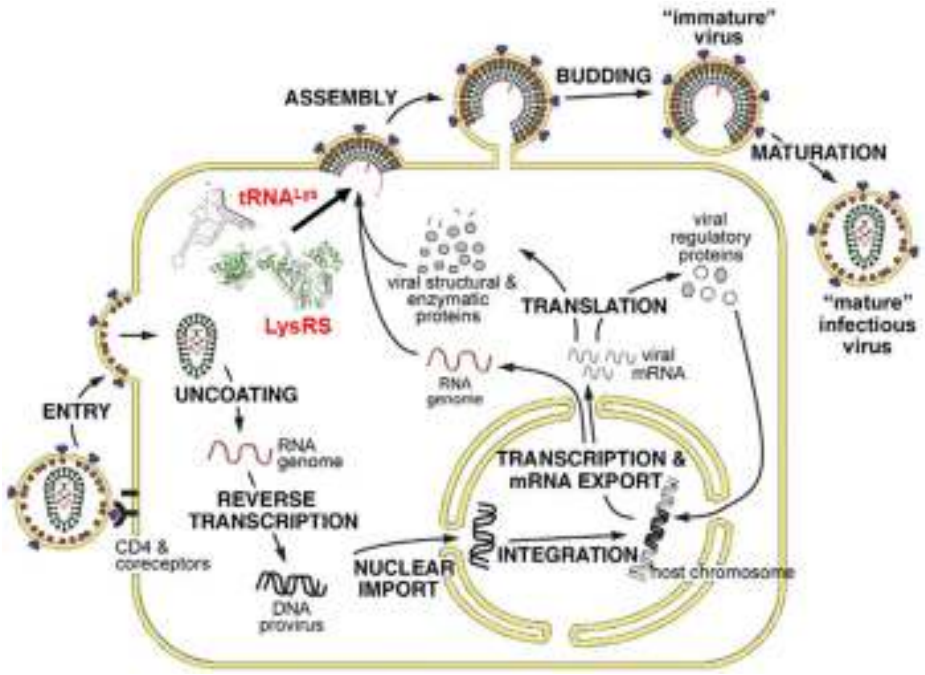



**Figure 4: The HIV life cycle.**

**The figure illustrates viral entry followed by reverse transcription and integration into the host genome. Viral RNA transcription follows with assembly and formation of a new infectious particle. (Reproduced from http://www.slideshare.net/sudhakorwar/presentation-6608205)**

### 1.5.1 The fate of HIV viral DNA

A small percentage of the linearized dsDNA gets integrated into the genome of the host to form the provirus. The provirus acts as a latent reservoir and ensures persistence of infection (Demetriou et al., 2010). The remaining linearized dsDNA is circularized to form 2-LTR circles (Butler et al., 2001; Vandegraaff et al., 2001). Homologous recombination in 2-LTR circles results in 1-LTR circles. Auto integration into LTR circles can also result in various rearranged circular products. Also, most of the HIV-1 DNA remain un-circularized and un-integrated and add up to the episomal population of HIV-1 viral DNA. The entire nonintegrated DNA is a dead end product, which may be degraded or persists in the nucleus of the cell (Butler et al., 2001; Vandegraaff et al., 2001). The HIV-1 DNA forms are illustrated in figure 5 (Kostrikis et al., 2002a)

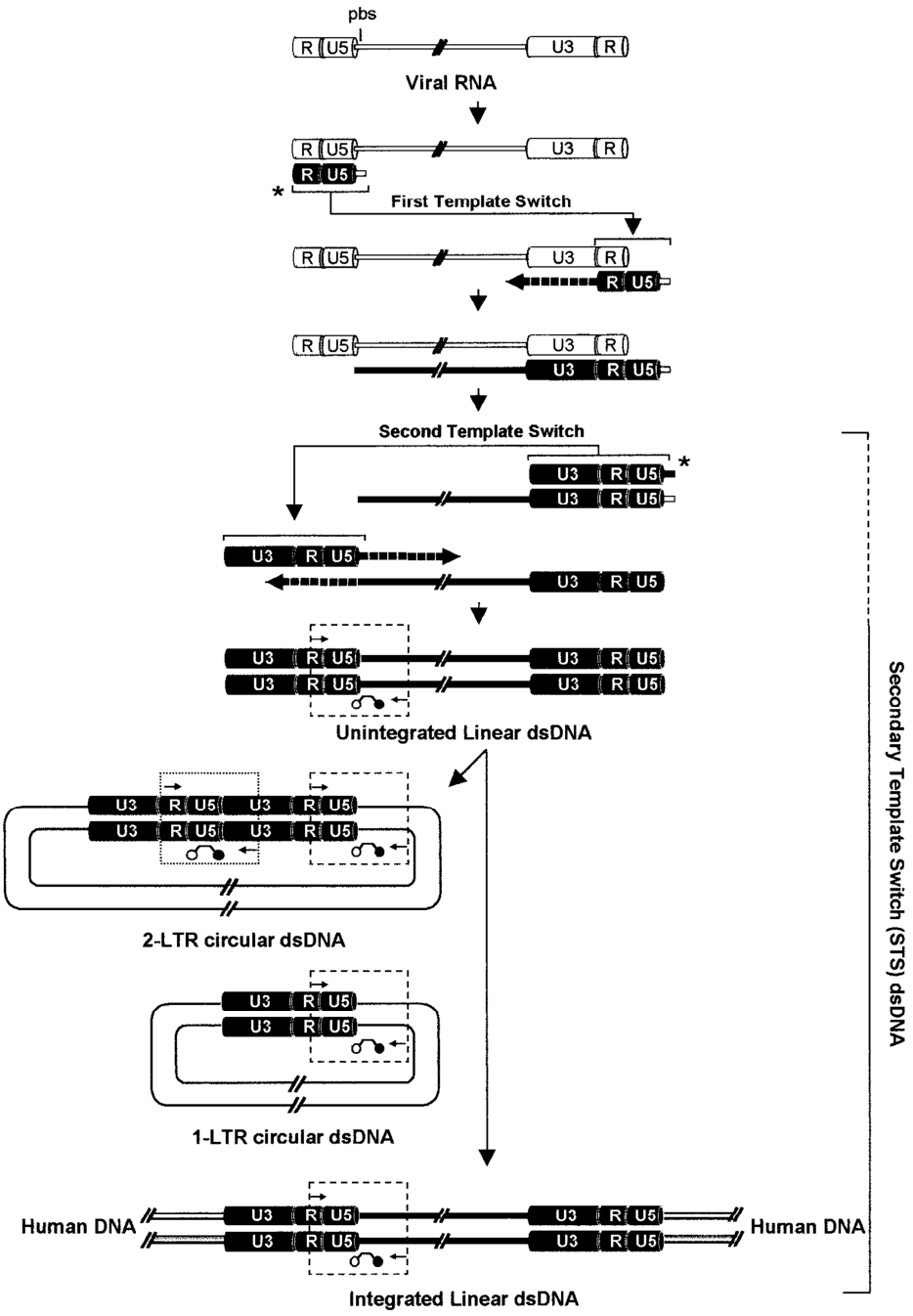



**Figure 5: Schematic representation of the intermediate products formed during the process of HIV-1 reverse transcription of the viral RNA to form the HIV-1 second template switch DNA.**

**The figure illustrates the formation of the linear dsDNA and its possible outcomes; the various LTR circles and the integrated provirus. Reproduced from (Kostrikis et al., 2002a).**

## 1.6 Management of HIV/AIDS

Management of HIV/AIDS has evolved significantly since the initial diagnosis in the early 1980s. While clinical management was ineffective in the early stages of the pandemic, the development of effective antiretroviral drugs has significantly improved clinical outcomes over the last couple of decades (Maartens et al., 2014). Till date, over 24 licensed drugs have been approved for treatment of HIV/AIDS (Maartens et al., 2014).

Antiretroviral drugs are generally classified based on the stage of the viral life cycle targeted. Currently there are 5 main classes of ARVs approved by the FDA for clinical use, which are: Viral Entry inhibitors, nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), and viral integration inhibitors (figure 6) (Arts and Hazuda, 2012). HIV-1 drug resistance mutations are quite common and with over 70% resistant rates for some individual ARVs (Nii-Trebi et al., 2013; Richman et al., 2004). Drug resistant mutations arise from key changes in base pairs that confer an escape for a sub population of the virus from the ARV effects. While resistance may develop rapidly to some antiretroviral drugs, the process can be slow and occur over time with accumulation of mutations that will ultimately confer resistance (Das and Arnold, 2013). Mechanisms of drug resistance development vary between drug families (Das and Arnold, 2013). Below are descriptions of the actions of the various groups of ARTs and their mechanisms of action and resistance.

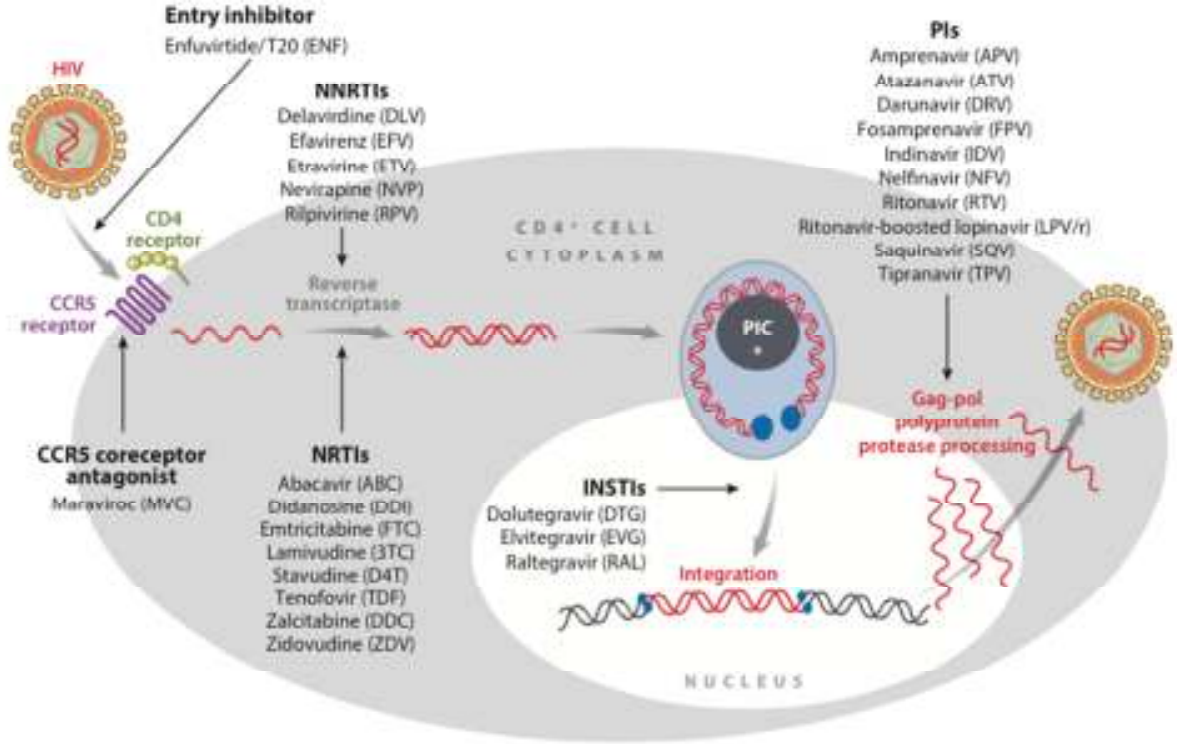



**Figure 6: The HIV-1 life cycle and drug targets at different stages of viral processing.**

**The figure illustrates the various points in the HIV-1 life cycle targeted by the various approved classes of drugs with specific examples of each group. Reproduced from (Chabria et al., 2014).**

## 1.6.1 Nucleoside/Nucleotide reverse transcriptase inhibitors

The critical importance of RT to the life cycle of HIV-1 has led to the development of many drugs targeting it. Nucleoside reverse transcriptase inhibitors (NRTIs) are nucleoside analogues that incorporate into the DNA chain synthesized by RT and hence interrupt elongation. NRTIs lack a 3' OH group and thus prevent the addition of additional nucleotides (figure 7A and 7B). Examples of NRTIs include zidovudine, lamivudine and abacavir (Das and Arnold, 2013).

Resistance to nucleoside reverse transcriptase inhibitors (NRTIs) occurs through two basic mechanisms. The primary mechanism involves mutations (M184V, Q151M) in the drug binding site that increases discrimination against drug binding and hence resistance (Johnson et

al., 2010; Larder and Kemp, 1989). The other mechanism involves increased ability of the reverse transcriptase enzyme to remove incorporated NRTIs from the chain by a process called pyrophosphorolysis and hence reduces their ability to inhibit reverse transcription of viral RNA (Arion and Parniak, 1999; Sluis-Cremer et al., 2000) (figure 7A and 7B). Some studies have demonstrated that pyrophosphorolysis is augmented by the presence of thymidine analogue mutations (TAMs) in reverse transcriptase (Ray et al., 2003; von Kleist et al., 2012). Common TAMs include M41L and L210W among several others (Hu et al., 2006).

Development of resistance to NRTIs has been shown to vary based on virus subtype and region (Wainberg et al., 2011). For example patients with subtype C treated in Botswana with zidovudine and didanosine were shown to have higher TAMs as compared to patients infected with subtype C in India, South Africa and Malawi (Wainberg et al., 2011). Also, it was shown that patients with CRF06_cpx viruses are predisposed to having higher TAMs than those carrying CRF02_AG recombinant viruses in Burkina Faso (Wainberg et al., 2011).

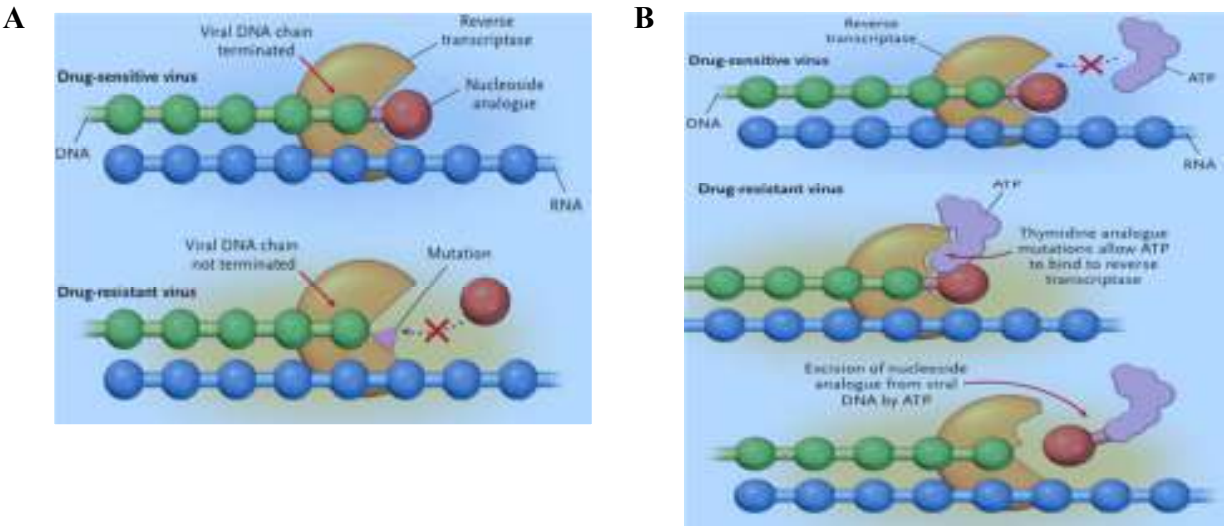



**Figure 7: Mechanisms of action of nucleoside reverse transcriptase inhibitors and development of resistance.**

**NRTIs are incorporated into the growing chain and prevent elongation. Mutations in the binding pocket of the enzyme prevent incorporation of the NRTI thereby leading to resistance development (A). NNRTIs can also prevent ATP binding to prevent reverse transcription. Mutations that allow ATP to bind can result in removal of the NRTI for generation of resistance. Reproduced from (Clavel and Hance, 2004).**

### 1.6.2 Non-nucleoside reverse transcriptase inhibitors

Non-nucleoside reverse transcriptase inhibitors (NNRTI) bind to a hydrophobic pocket of reverse transcriptase and allosterically reduce the catalytic ability of the enzyme (Das and Arnold, 2013) (Figure 8A, 8B and 8C). Examples of NNRTIs include nevirapine and efavirenz (Das and Arnold, 2013). Resistance to non-nucleoside reverse transcriptase inhibitors (NNRTIs) occurs due to mutations that lower drug binding and prevents allosteric inhibition of reverse transcriptase (Akinsete et al., 2004; Tambuyzer et al., 2009). The mutation K103N is one of several mutations that confer resistance against almost all NNRTIs (Akinsete et al., 2004; Tambuyzer et al., 2009).

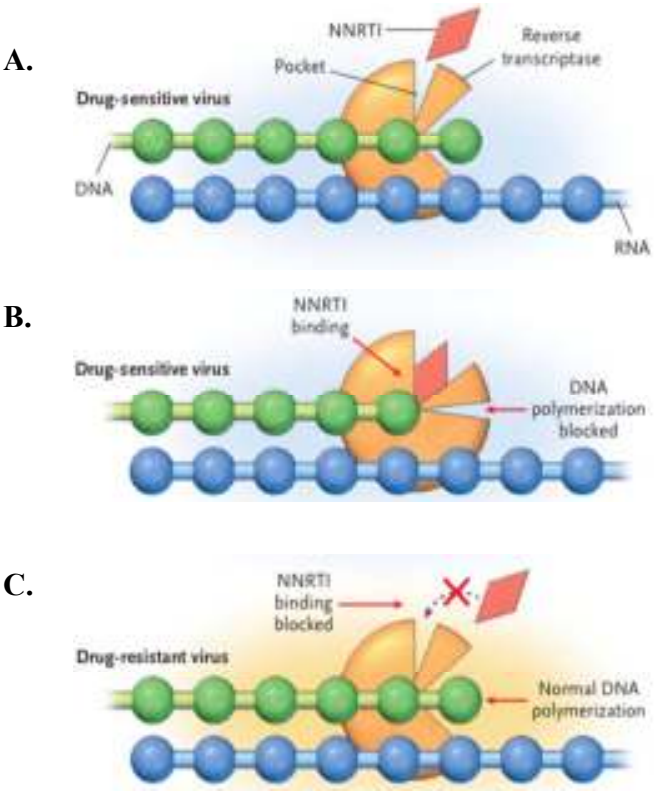



**Figure 8: Mechanism of action and resistance of HIV to Non-nucleoside Reverse-Transcriptase Inhibitors.**

**NNRTIs bind to an allosteric pocket of the RT and reduce activity at the active site of the enzyme. Mutations that preventing drug binding of the enzyme in the binding pocket confers resistance. Reproduced from (Clavel and Hance, 2004).**

### 1.6.3 Protease inhibitors

Protease inhibitors inhibit the HIV-1 protease enzyme, a heterodimer that cleaves HIV-1 polyproteins into functional forms without which the infectious virus cannot be formed (Fun et al., 2012). Protease cleaves the HIV-1 large precursor polyproteins gag and gag-pol into functional proteins, which include p7, p24, p17 and p6 as well as viral enzymes reverse transcriptase, protease and integrase (Baca and Kent, 1993; Fun et al., 2012). Protease inhibitors bind the substrate-binding site of the enzyme heterodimer and by that they inhibit cleavage of the polyproteins (Fun et al., 2012). Mutations in the substrate-binding pocket of the enzymes reduce affinity for the drugs but not the natural substrates and thus confer resistance to the protease inhibitors (Fun et al., 2012).

## 1.7 HIV Resistance Development and Transmission of Resistance

The high rate of mutation in HIV-1 generates numerous HIV-1 strains within a patient referred to as viral quasispecies (Domingo et al., 2012). In a patient, one of HIV quasispecies is predominant at the initial stages of infection (Metzner, 2006). In ARV treated individuals, the balance of the quasispecies is altered and minority species that can overcome the bottleneck due to the ARVs become the dominant species, as they are fitter under the unique circumstance of therapy (Metzner, 2006). The ability of a strain of HIV to overcome adverse conditions in the host, based on its unique DNA alterations, is referred to as viral fitness (Buckheit, 2004). The concept of viral fitness is described by the Red Queen hypothesis where viral quasispecies in competition tend to increase in replication ability (Fitness) with successive generations (Domingo et al., 2012). Understanding of HIV resistance mutation development and viral fitness is critical to understanding the long term virulence and transmission as well as drug susceptibility and therapeutic strategies (Arnott et al., 2010; Domingo et al., 2012).

### 1.7.1 Transmitted drug resistance mutations

Highly active antiretroviral drugs (HAART) have in recent times become increasingly available to patients infected with the human immunodeficiency virus (HIV-1) in Africa and large scale programs have been implemented in many African countries for the administration of

these drugs (UNAIDS, 2013a). The implementation of HAART has evidently led to a significant decrease in morbidity and mortality among infected individuals. However, the increasing use of HAART across the world has led to the emergence of drug resistant virus strains, known to reduce susceptibility to the ART and which can also be transmitted to treatment naive patients (Avila-Rios et al., 2014) (Johnson et al., 2010). This implies that an increasing number of drug naive patients are being infected with viruses that are already resistant to ARTs termed transmitted drug resistant viruses (Pillay et al., 2006). These viruses pose a major clinical and public health issue. On the other hand, resistant mutations developed by an individual on ART are termed acquired drug resistance mutations (DRMs). These mutations usually occur as a result of a drug selection pressure (Metzner et al., 2009).

### 1.7.2 Prevalence of transmitted drug resistant mutations in HIV-1 newly diagnosed patients

Transmitted drug resistant mutations surveillance is becoming an important factor in the global management of HIV-1. In 2012, The WHO HIV-1 drug resistance report estimates indicate increasing trends in TDRM globally due to increasing use of ARVs (WHO, 2013b). In developed territories like Europe, USA and Japan, 10 to 17 percent of newly diagnosed HIV-1 patients have at least one TDRM (WHO, 2012). In developing countries as well, the WHO reports an increasing trend of TDRM between 2003 and 2010 (WHO, 2012). The implications of TDRMs to success of ARV therapy was demonstrated in a large European study which showed significant increase in treatment failure in patients who had TDRMs as compared to those who did not (Wittkop et al., 2011).

With increasing ARV therapy availability in Africa, it follows that the prevalence of TDRMs will also increase, highlighting the importance of TDRM monitoring (WHO, 2012). Recent studies conducted among a cohort of sub-Saharan Africans in eight clinical research centers (Kigali, Rwanda; Lusaka, and Copperbelt, Zambia; Masaka and Entebbe, Uganda; Nairobi and Kilifi, Kenya; and Cape Town, South Africa) have confirmed low but increasing trends of TDRM (Price et al., 2011b). The prevalence of transmitted drug resistant mutations

(TDRM) in newly diagnosed patients in most African countries ranges from 3% and 17% (Derache et al., 2008; Ndembi et al., 2011; Ojesina et al., 2006).

## 1.8 Assays for HIV drug resistance

The development of assays that can detect resistant mutations has enabled testing for drug resistance and enable decision-making on the choice of ARVs and alterations in therapy (Hirsch et al., 2008). Phenotypic and genotypic resistance assays are the two types of assays that are currently in use for the assessment of DRMs. Genotypic resistance assays use data from sequenced HIV genes from patients to detect mutations that may confer resistance to a given ARV. Phenotypic resistance assays are cell culture based assays that assess viral replication capabilities in the presence or absence of a given ARV (Hirsch et al., 2008).

### 1.8.1 Genotypic resistance assays

Genotypic resistance assays can be done using standard kits or in-house assays based on resources of a given lab. In-house assays use regular sequencing techniques like the Sanger sequencing to determine nucleotide sequences of patient derived viruses. To perform Sanger sequencing, the viral RNA is reverse transcribed to DNA prior to sequencing usually targeting the RT and PR regions (Geretti et al., 2014; Hirsch et al., 2008). The FDA approved the Truegene HIV-1 kit for use in genotypic monitoring in 2001 (Grant et al., 2003). In genotypic assays, rule-based algorithms on HIV databases (Hirsch et al., 2008) are used for predictions on the relevance of mutations for clinical decision-making.

Genotypic resistance testing is recommended for routine resistant mutation testing among drug naïve patients (WHO, 2006). However several limitations are recognized with genotypic resistance assays. First, there is the general difficulty in predicting clinical outcomes entirely based on algorithms. Secondly, there is the difficulty in assessing viral strains due to the heterogeneity of the viral genome of an infected patient (Hirsch et al., 2008). This normally leads to the production of patient derived viral chromatographs that are difficult to interpret. Since most tests have been based on B-subtypes, non-B and C subtypes pose some difficulty in interpreting mutations that affect drug susceptibility (Hirsch et al., 2008). There is also the

difficulty of detection when HIV viral load is low. Recently however, some groups have demonstrated the ability to do resistance testing with viral loads less than 1000 copies /ml of plasma with a significant success rate below 50 copies/ml of plasma (Gonzalez-Serna et al., 2014; Santoro et al., 2014).

## 1.8.2 New trends in drug resistance assays

Besides standard phenotypic and genotypic assays, other assays are used for viral fitness tests. They include allele-specific PCR assays, ultra-deep and single genome sequencing. These assays are used to investigate minority variants with mutations that are hard to detect by standard approaches (Hirsch et al., 2008).

### 1.8.2.1 Deep sequencing

Population based Sanger sequencing has been the standard method that has been used in DNA sequencing over the past decades but in recent times next generation or deep sequencing is revolutionizing the field of genomics including clinical virology (Gibson et al., 2014; Quinones-Mateu et al., 2014).

During Sanger sequencing, the DNA template is only read once, thus increasing the chances of errors within the growing nucleotide sequence and also making it difficult to differentiate single nucleotide polymorphisms from resistant mutations (Chabria et al., 2014). Other limitations of Sanger sequencing includes inability to read minority strains less than 20% of viral quasi species (Chabria et al., 2014). It is known that minority variants gain fitness in the presence of drugs and can become dominant and therefore, knowledge of such variants is important for clinical practice (Chabria et al., 2014).

Deep sequencing employs next generation sequencing (NGS) strategies that allow multiple reading of a single nucleotide in the DNA sequence and thus increase the precision (Fox et al., 2014). In the clinical virology settings, when it comes to dealing with minority variants in DRMs, deep sequencing techniques, allow for reading of sequences of minority variants as low as 0.5-1% in the viral quasi species (Gibson et al., 2014; Quinones-Mateu et al., 2014). In deep sequencing the average number of times a nucleotide is read is referred to as coverage (Sims et al., 2014). Various platforms have been developed for deep sequencing using specific methods

however; they all employ the use of reverse transcription of viral RNA and PCR in order to generate amplicons for sequencing. Multiple primers ad adapters are used in order to capture all the sequences. Examples of NGS platforms currently in use include Roche, Illumina ad Ion Torrent (Chabria et al., 2014). Deep sequencing however has some drawbacks including the use of different primer pairs that can result in a distortion in the composition of the viral quasi species as primer efficiency could vary (Chabria et al., 2014; Jabara et al., 2011) and the occurrence of PCR recombinants that result in false strain identification (Chabria et al., 2014; Jabara et al., 2011; Kozal et al., 2011).

## 1.9 Comparing HIV proviral DNA and RNA for mutations

During HIV treatment, the selective pressure on the virus quasispecies results in selection of resistant mutations (Nikolenko et al., 2011). The mutations are generally measured in the viral RNA, which is mainly extracellular. However while mutations occurring in intracellular DNA may be similar to the RNA forms, discordance has been shown to exist between them as well (Banks et al., 2012). In resource poor countries however, DNA analysis for mutation may be cheaper to undertake in comparison to RNA (Banks et al., 2012). Some studies have shown that proviral DNA may differ in mutations compared to RNA. One reason for possible disparities is the difference in turnover rates (Banks et al., 2012; Delaugerre et al., 2012). In one study done in Zimbabwe, discordance between viral RNA and DNA drug resistance mutations were reported for PIs, NRTIs and NNRTIs (Banks et al., 2012). On the other hand, a high concordance between viral RNA and DNA has been reported in several studies (Ferrer et al., 2013; Mayers et al., 1998).

## 1.10 Therapeutic goals and treatment outcomes during HIV management

The primary goal of current HIV treatment is to reduce viral load of patients to below detectable levels (Okulicz et al., 2014) and restore immune function (Okulicz et al., 2014) although efforts are being made to completely eradicate the virus (de Mendoza et al., 2014) or attain cure based on prevention of reservoir development or viral endogenization (Ananworanich

et al., 2015; Colson et al., 2014). Also treatment and prevention of opportunistic infections and improvement of overall quality of life are important aims of the management of HIV/AIDS patients (Dybul et al., 2002).

Although the mainstay of clinical management of HIV is antiretroviral therapy, a critical issue has been the optimal time to initiate therapy (Ananworanich et al., 2015; Okulicz et al., 2014). Some studies have demonstrated that a short course antiretroviral therapy soon after infection may significantly slow disease progression (Ananworanich et al., 2015; Saez-Cirion et al., 2013). However there are questions about the benefits of early treatment in the face of increasing drug resistance (Maartens et al., 2014). Generally, patients are initiated on ARVs when their $CD4^+$ T-cell counts are below 350 cells/µl, however patients with $CD4^+$ T-cell counts over 350 cells/µl who show signs of complications of AIDS are started on ARVs (Maartens et al., 2014). In spite of the limited scope of trials in early initiation of therapy, the consensus on initiation of ARVs is inclined to early initiation of therapy where possible as this has been shown to improve and sustain viral load reduction, as well as optimize immune recovery (Ananworanich et al., 2015; Okulicz et al., 2014).

### 1.10.1 Disease monitoring in HIV

Monitoring of HIV-1 infection and disease progression is an important part of management especially in the current era of ARV. Monitoring disease progression determines decision-making, especially the timing of initiation of ARVs or opportunistic infection prophylaxis and also when to alter therapy (Laurent et al., 2011). There are several markers that have been used over the years to monitor HIV and disease progression but the most widely used are plasma viral load and $CD4^+$ T-cell counts (Mellors et al., 1995; Mellors et al., 1997a). Besides viral load and $CD4^+$ T-cell counts, there are other surrogate markers for HIV disease progression and these include hemoglobin (Ioannidis et al., 2001a), serum albumin, total lymphocyte count, liver enzymes (Venkataraman, 2013) and in recent times intracellular viral DNA levels (Williams et al., 2014).

### 1.10.2 Response to antiviral drug therapy

Although drug therapy has significantly revolutionized clinical management of HIV/AIDS, the response to therapy can vary in practice (Maartens et al., 2014). A good understanding of response to drug therapy and its implications is important for optimal clinical management. Given that the goal of therapy is to restore immune response and reduce viral load to below detection levels, response to therapy is divided into immune response and virological response as well as clinical response (Phillips et al., 2001). While response to drug therapy has been associated with time of initiation and baseline $CD4^+$ T-cell counts, no significant association has been made between virus subtype and response to therapy (Phillips et al., 2001).

After initiation of ART, viral load usually reduces to below detection limits for most assays. This usually occurs within 3 months of treatment initiation (Maartens et al., 2014). In contrast to virological response, immunological response lags behind and is usually achieved within 6 months of therapy (Maartens et al., 2014). While virological response is usually followed by immunological response during therapy, some proportions of treated patients have a discordant response to treatment. In one study, 56% of treated patients had a good virological and immunological response, 19% of patients had good virological response with poor immunological response and 14.8% had a good immunological response with poor virological response (Maartens et al., 2014; Tuboi et al., 2007).

## 1.11 Determinants of disease progression

Determinants of disease progression in HIV-1 infection have been an area of intense research due to its importance in clinical practice (Singh and Spector, 2009). Being able to predict disease progression can help clinicians tailor management regimes to suit individual cases in the hope for successful outcomes of therapy. Factors that determine disease progression include those due to viral characteristics, host genetic polymorphisms (Ioannidis et al., 2001b; Singh and Spector, 2009) and host immune characteristics (Hogan and Hammer, 2001). The viral load prior to therapy has been identified as a key determinant in disease progression (Mellors et al., 1997b). The viral load usually assessed is the viral load in plasma. However, it has recently

been determined that intracellular viral load may be an important and independent determinant of disease progression (Demetriou et al., 2010; Kostrikis et al., 2002b).

## 1.12 Transmitted drug resistant mutations and disease progression

Transmitted drug resistant mutations and disease progression in HIV-1 infection remains an area not completely clarified. Some studies have indicated that ART resistance is not a major determinant of disease progression and that continued adherence to HAART could still be protective in the face of drug resistant mutations (Lucas, 2005; Lucas et al., 2004; Recsky et al., 2004). However others studies have indicated a clear correlation between number of resistant mutations and virological outcomes (Ledergerber et al., 2004), which presupposes that, transmitted drug resistance may increase the risk of virologic failure (Daar and Richman, 2005). In 2005, a multi-resistant strain was identified in a New Yorker who progressed rapidly to AIDs, however, the prevalence of this strain was found to be low (Blick et al., 2007). Thus far, the role of drug resistance on disease progression requires a lot more research in order to clarify its relevance clinically.

In Africa, TDRM have been reported (Price et al., 2011b; WHO, 2012) however, not much has been studied on its effect on disease progression. Since Africa bears the brunt of the world's HIV-1 burden, it would be worth determining the effect of resistant mutations on disease progression, as it would duly inform clinical practice in the wake of the increasing development of mutants.

## 1.13 HIV-1 intracellular DNA load and disease progression

Plasma HIV-1 viral load has been shown to be one of the most important and independent determinants of HIV-1 disease progression (Mellors et al., 1995; Mellors et al., 1996). While the clinically determined viral load is RNA, the form of the virus in the nucleus of the cell is DNA (Craigie and Bushman, 2012; Demetriou et al., 2010). HIV-1 IC DNA load is a marker associated with the viral reservoir and with the spread of the virus (Parisi et al., 2012). Studies in patients with primary HIV-1 infection and advanced HIV-1 disease have demonstrated that early levels of HIV-1 IC DNA load prior to initiation of highly active anti-retroviral therapy (HAART) in peripheral blood mononuclear cells (PBMC) and in $CD4^{+}$T -cells have a predictive

value for long-term virological outcome and for disease progression, independently of $CD4^+$ T-cell counts and plasma viral RNA load (Russell et al., 2001; Saitoh et al., 2002; Verhofstede et al., 1994). Recently, Williams et al showed that HIV-1 intracellular viral DNA may even be a better predictor of disease progression than plasma RNA viral load (Williams et al., 2014).

Thus far, the factors that determine HIV-1 IC DNA levels are not well understood. In a recent study conducted in a cohort of newly diagnosed Europeans with HIV infection, it was found that the resistant viral strains did not have any correlation with HIV-1 intracellular viral DNA load prior to initiation of treatment as compared to non-resistant viral strains (Demetriou et al., 2010).

## 1.14 Therapy failure with HAART

The WHO defines treatment failure by plasma viral load and $CD4^+$ T-cell count decline (Hirnschall et al., 2013). According to the WHO, HIV viral load monitoring is the preferred approach for assessing response to treatment and treatment failure (Hirnschall et al., 2013). Two parameters are defined with respect to changes in viral load following initiation of therapy. Viral suppression refers to a reduction in viral load below detection for the assay, usually below 50 copies cells/ml of plasma however, virologic failure is said to have occurred when viral load persistently remains above 1000 copies cells/ml of plasma after 6 months of ART (WHO, 2013a). While immunological response and clinical outcomes may be used for complementing viral load and monitoring ART success, they have a poor sensitivity and positive predictive value for identifying virologic failure in adults and children (Hirnschall et al., 2013). Where it is impossible to measure viral load during ART, it is also recommended that immunological and clinical responses should be used to assess success of therapy (Laurent et al., 2011; WHO, 2013a). WHO defines immunological failure as $CD4^+$ T-cell counts falling to baseline or below or persistent $CD4^+$ T-cell counts below 100 cells/µl of blood (WHO, 2013a) and clinical failure represent a new or recurrent event as a result of severs immunological dysfunction after 6 months of therapy (WHO, 2013a).

During treatment with HAART, some patients responding to treatment with undetectable viral loads experience a period of low-level detectable viremia, followed by suppression of viral replication (Miller et al., 2004). These transient increases in viral load, are

referred to as blips (Fidler et al., 2014; Miller et al., 2004). Between 10 to 20 percent of patients on HAART experience blips (Fidler et al., 2014; Ibrahim et al., 2012) although up to 90 percent has been reported in one study (Nettles et al., 2005). The mechanism of virological blips is not clearly understood, however, it has been suggested that changes in immune activation and reservoir by stimuli like infections may account for it (Jones and Perelson, 2007). Virological blips have been attributed to random biological and statistical variations around mean viral levels (Nettles et al., 2005) and not poor adherence to medication (Miller et al., 2004). Virological blips are however not predictive for virologic failure (Fidler et al., 2014; Ibrahim et al., 2012; Nettles et al., 2005).

## 1.15 Overview of HIV in Burkina-Faso

HIV was first diagnosed in Burkina Faso in 1986 (WHO, 2005). The HIV prevalence in Burkina Faso varies widely with high rates in urban centers. A study conducted in urban Burkina Faso (Bobo-Dioulasso) reported a prevalence of 5.2% in 2004 (Lagarde et al., 2004). However the general prevalence of HIV in adults between the ages of 15 and 49 years is generally low, estimated at 1.1% (UNAIDS, 2013b). The predominant strains of HIV in Burkina Faso are CRF02_AG and CRF06_cpx although several strains are present (Tebit et al., 2006; Tebit et al., 2009).

Up-scale of HAART was initiated in Burkina Faso in 2003 and this has resulted in an increase of individuals on treatment alongside with the development of drug resistant mutations in treated and drug naïve patients (Somda et al., 2012; Tebit et al., 2008; Tebit et al., 2009). One important contribution to resistance mutation development is the use of nevirapine in the prevention of mother to child transmission (Tebit et al., 2006; Tebit et al., 2008). According to Tebit et al, the prevalence of resistance mutations in patients failing therapy was 40% to PIs, 76% NNRTIs and 85% to NRTIs (Tebit et al., 2008). Several studies have assessed the prevalence of DRMs among urban populations in Burkina Faso. Among pregnant women, a moderate (5-15%) rate of drug resistance mutations was reported (Somda et al., 2012). Also, Tebit et al reported a TDRM rate of 12.5% in a study in Ouagadougou (Tebit et al., 2009). In rural Burkina Faso, where the present study was conducted, there has been no study reporting TDRM mutations and how it correlates with markers of disease progression.

### 1.15.1 Rationale of the study

The improvement in survival of HIV infected individuals due to the introduction of HAART has been a major milestone in the global HIV pandemic. However, new challenges have arisen by virtue of the lifelong nature of therapy and the potential for treatment failure due to development of drug resistant strains of the virus. Also, there is the need for development of more sensitive markers for predicting disease progression in the clinics.

Plasma HIV RNA load and $CD4^+$ T-cell counts are important determinants of disease progression clinically. However in recent times, HIV-1 IC DNA has been shown to be an important independent predictor of disease progression (Kostrikis et al., 2002a). The use of HIV-1 IC DNA could provide an alternative for viral RNA measurements and also can be assessed in the case of undetectable RNA levels.

Although most of the global burden of HIV is in sub-Saharan Africa, studies on TDRMs and its implications on HIV-1 IC DNA have only been carried out in developed countries. In this study, we assessed the effect of TDRMs on HIV-1 IC DNA levels and the utility of HIV-1 IC DNA as a predictor of disease progression in a group HIV infected patients from Nouna Burkina Faso in sub-Saharan Africa.

### 1.15.2 General objectives

Using samples from HIV-1 infected patients from Nouna Burkina-Faso, West Africa, this study assessed the following:

1. The predominant DRMs in the protease (PR) and reverse transcriptase (RT) region of HIV-1 pol from RNA and DNA templates of treated and drug naïve HIV-1 infected patients starting ART.
2. How TDRMs affect HIV-1 IC DNA levels in HIV-1 infected drug naïve patients.

3. The correlation between baseline HIV-1 IC DNA levels of infected patients and established primary markers of disease progression such as plasma RNA viral load, and $CD4^+$ T-cell count as well as other indicators associated with disease progression.

# 2 Materials and Methods

## 2.1 Materials

### 2.1.1 Media and buffers

**Table 1: List of prepared media and buffers used in this study**

| Bacterial culture | Constinuents |
|---|---|
| LB medium | 1 % peptone, 0.5 % yeast extract, 171 mM NaCl |
| LB agar | 13 % agar in LB medium |
| LB-ampicillin or kanamycin ag | 100 µg/ml ampicillin or kanamycin |
| 50x TAE buffer | 2 M Tris, 1 M NaAc, 0.1 M EDTA |
| 10x DNA loading buffer | 100% glycerol, 2.5% bromphenol blue and 2.5% xylene cyanol |
| **PCRs and Extractions** | |
| Tris-EDTA buffer | with HCl |
| TaqMan qPCR buffer A | 500 mM KCI, 100 mM Tris-HCI, 0.1 mM EDTA, 600 nM Passive Reference 1 (Rox dye), pH 8.3 at room temperature (solution was autoclaved) |

### 2.1.2 Commercial reagents kits and instruments

**Table 2: List of commercial kits and consumables used in this study**

| Commercial kit or consumable | Supplying company |
|---|---|
| Viral DNA extraction kit | QIAmp viral DNA mini kit, Qiagen, Hilden, Germany |
| Viral RNA extraction kit | QIAmp viral RNA mini kit, Qiagen, Hilden, Germany |
| SS III One Step HIFI RT-PCR amplification kit | Life Technologies, Darmstadt, Germany |
| Platinum PCR SuperMix | Life Technologies, Darmstadt, Germany |
| Taq DNA Polymerase with Standard Taq Buffer PCR Kit | Life Technologies, Darmstadt, Germany |
| Phusion high fidelity DNA polymerase | New England BioLabs, Frankfurt, Germany |
| Verbatim high fidelity DNA polymerase enzyme | Life Technologies, Darmstadt, Germany |
| Qiaquick PCR purification kit | Qiagen, Hilden, Germany<br>Nucleospin Extract II kit, Macherey-Nagel, Düren, Germany |
| Nucleospin Gel and PCR Clean-up kit | Macherey-Nagel, Düren, Germany |
| NucleoBond MaxiPrep Kit | Macherey-Nagel, Düren, Germany |
| TaqMan PCR Core Reagent with AmpliTaq Gold DNA polymerase | Life Technologies, Darmstadt, Germany |
| TOPO® TA Cloning® Kit | Life Technologies, Darmstadt, Germany |
| DNA LoBind eppendorf tubes 1.5 ml | Eppendorf AG, Hambug, Germany |
| Agencourt AMPure XP PCR purification system | Beckman Coulter, Inc., Beverly, Massachusetts, USA |
| Quant-iT Picogreen dsDNA reagent and kits | Invitrogen, Life Technologies |
| Nextera XT DNA Library Preparation Kit for the MiSeq® System | Illumina, Inc., San Diego, CA, USA |

### 2.1.3 Reagents and chemicals

**Table 3: List of reagents and chemicals used in this study**

| Reagents and Chemicals | Supplying company |
|---|---|
| Agarose | Carl Roth, Karlsruhe, Germany |
| Ethanol (99%) | Sigma-Aldrich, Steinheim, Germany |
| Pure glycerol (100%) | AppliChem GmbH, Darmstadt, Germany |
| Deoxyribonucleoside triphosphate mix (dNTP mix) | Fermentas, Thermo Scientific, Schwerte, Germany |
| Ampicillin | Carl Roth, Karlsruhe, Germany |
| Midori Green Advanced DNA stain | NIPPON Genetics Europe GmbH, Düren, Germany |
| Kanamycin | Carl Roth, Karlsruhe, Germany |
| RNASE Away | Life Technologies, Darmstadt, Germany |
| DNAZAP | Life Technologies, Darmstadt, Germany |
| GeneRuler 1kb Plus DNA ladder mix | Fermentas, Thermo Scientific, Schwerte, Germany |
| 100 bp DNA Ladder | New England BioLabs, Frankfurt, Germany |
| Nuclease free water | Life Technologies, Darmstadt, Germany |
| MiSeq Reagent Kits v2 | Illumina, Inc., San Diego, CA, USA |

### 2.1.4 Primers and molecular-beacons

**Table 4: List of oligonucleotide primers and molecular-beacons used in this study**

| Name | Target region | Sequence 5´-3´ | Position | length of amplified region | Reference |
|---|---|---|---|---|---|
| **PCR primers** | | | | | |
| 1832 ( F1 ) | pol | CAGCATGYCAGGGAGTRGGRGGACC | 1832–1856 | 1751 | Kousiappa et al., 2009 |
| 3583 (R1) | pol | GGYTCTTGRTAAATTTGATATGTCCATTG | 3555–3583 | | Kousiappa et al., 2009 |
| 2078 (F2) | pol | AGGCTAATTTTTTAGGGAARATYTGGCCTTCC | 2078–2109 | 1461 | Kousiappa et al., 2009 |
| 3539 (R2) | pol | CTGTATTTCTGCTAYTAAGTCTTTTGATGG | 3510-3539 | | Kousiappa et al., 2009 |
| BFp1 (F1) | pol | AGACACAGGAGCAGATGATACAGT | 2324–2347 | 491 | Tebit D.M et al., 2006 |
| BFp4b (R1) | pol | TCCCAGAAGTCTTGAGTTCTCTTATT | 2790–2815 | | Tebit D.M et al., 2006 |
| BFp3 (F2) | pol | ATGATAGGGGGAATTGGAGG | 2388–2407 | 406 | Tebit D.M et al., 2006 |
| BFp4c (R2) | pol | TTATTGAGTTCTCTGAAATCTACTAATTTTCTCC | 2761–2794 | | Tebit D.M et al., 2006 |
| RT1 (F) | pol | AGTAGGACCTACACCTGTCAACATAATTGG | 2480-2509 | 1054 | Kanokporn 2009 |
| RT-Rev-Stu1 (R) | pol | TTTCTGCTACTAGGCCTTTTGCTGGGTCATAA-TAGACTCCATGTACAGGTTCTTTT | 3479-3534 | 1423 | Kanokporn 2009 |
| Pol1 (F1) | pol | TCAAGGGAAGGCCAGGGAATTT | 2111-2132 | | Kanokporn 2009 |
| Prot-for-EcoRI (F) | pol | GAGGGACAAGGAATTCTACCCTCCTTTAGCTTCC CTCAAA | 2223-2259 | 1269 | Kanokporn 2009 |
| RT4- Rev | pol | CTTTTAGAATTTCCCTGTTCTCTGCCAATTC | 3453-3483 | | Kanokporn 2009 |
| RT3- For | pol | AATATGTTGACTCAGATTGGTTGTACTTTAAAT | 2514-2547 | 964 | Kanokporn 2009 |
| NewF1 | pol | TCCCCCAGAGCAGACCGGAAC | 2132-2152 | 1429 | This work |
| NewR1 | pol | TCCATTGGTCTTGCCCTTGTTTCTG | 3537-3561 | | This work |
| NewF2 | pol | CCGGAACCATCAGCCCCACC | 2146-2165 | 1304 | This work |
| NewR2 | pol | ACTCTGCTTCCTCAGTCAGTGGT | 3428-3450 | | This work |

**Table 4 continued.**

| Name | Target region | Sequence 5´-3´ | Position | Reference |
|---|---|---|---|---|
| **Sequencing primers** | | | | |
| 2136 (F) | pol | YCAGARCAGACCAGAGCCAACAGCCCC | 2136–2162 | Kousiappa et al. 2009 |
| 2216 (F) | pol | AGGAGCMGAWAGACARG | 2216–2232 | Kousiappa et al. 2009 |
| 2454 (F) | pol | GGAMAWAARGCTATAGGTACAG | 2454–2475 | Kousiappa et al. 2009 |
| 2610 (R) | pol | CYTTTGGGCCATCCATTC | 2593–2610 | Kousiappa et al. 2009 |
| 2650 (R) | pol | AATGCTTTTATTTTYTCTTCTGTCAATGGC | 2621–2650 | Kousiappa et al. 2009 |
| 2734 (R) | pol | GCAAATAYTGGAGTATTRTATGGATTTTCAG | 2703–2734 | Kousiappa et al. 2009 |
| 3003 (F) | pol | GGATGGAAAGGATCACC | 3003–3019 | Kousiappa et al. 2009 |
| 3019 (R) | pol | GGTGATCCTTTCCATCC | 3003–3019 | Kousiappa et al. 2009 |
| 3462 (R) | pol | CTGCCARTTCTARYTCTGCTTC | 3441–3462 | Kousiappa et al. 2009 |

F1 and R1 represent forward and reverse primers respectively of the first round PCR reaction and F2 and R2 represent forward and reverse primers respectively of the second round PCR reaction.

**Table 4 continued.**

| Name | Target region | Sequence 5´-3´ | Position | Reference |
|---|---|---|---|---|
| **HIV-1 quantification molecular-beacons and primers** | | | | |
| MB684 | LTR | FAM-CCGCTGCAAGCCGAGTCCTGCGTCGAGACAGCG G-Dabcyl | 684–705 | Demetriou et al. 2010 |
| 623 For | LTR | AAATCTCTAGCAGTGGCGCCCGAA | 623–646 | Demetriou et al. 2010 |
| 788 Rev | LTR | TCTCTCCTTCTAGCCTCCGCTAGT | 765–788 | Demetriou et al. 2010 |
| **CCR5 quantification molecular-beacons and primers** | | | | |
| LK155 | CCR5 | TET-GCGCCTATGACAAGCAGCGGCAGGAGGCGC-Dabcyl | 623–641 | Kostrikis et al., unpublished data |
| LK46 For | CCR5 | GCTGTGTTTGCGTCTCTCCCAGGA | 478–501 | Kostrikis et al., unpublished data |
| LK47.new Rev | CCR5 | CACAGCCCTGTGCCTCTTCTTCTCA | 690–713 | Kostrikis et al., unpublished data |
| **TOPO TA sequencing primers** | | | | |
| M13 (F) | | GTAAAACGACGGCCAG | | http://tools.lifetechnologies.com/content/sfs/manuals/topotaseq_man.pdf |
| M13 (R) | | CAGGAAACAGCTATGAC | | |
| T3 | | ATTAACCCTCACTAAAGGGA | | |
| T7 | | TAATACGACTCACTATAGGG | | |

The reporters; FAM and TET stand for fluorescein and Tetrachloro-69-carbofluorescein respectively. The quencher Dabcyl stands for 4-(49-dimethylamino phenylazo) benzoic acid. Underlined sequences within the molecular-beacons represent the target recognition sequences within HIV and CCR5. The underlined sequences at the 5´ and 3 ends represent complementary sequences forming the hairpin.´

### 2.1.5 Machines and instruments

**Table 5: List of machines and instruments used in this study**

| Machines and Instrumets | Supplying company |
|---|---|
| Incubators | Infors, Einsbach GmbH, Germany |
| NanoPhotometer P300 | Implen, München, Germany |
| Laminar flow Labotect | GmbH, Göttingen, Germany |
| Gel Running Chambers | Minigel-Twin Biometra, Göttingen, Germany |
| Tabletop ultracentrifuge TL-100 | Beckman Coulter, Fullerton, CA, USA |
| Biofuge fresco microcentrifuge | Heraeus, Kleinostheim, Germany |
| Centrifuges J2HC or J2HS with rotors JA-10, JA-17 and JA-20 | Eppendorf Deutschland, Hamburg, Germany |
| PTC-200 PCR Thermo cycler | GMI, Ramsey, MN, USA |
| Spectrophotometer DU 640 | Beckman Coulter, Fullerton, CA, USA |
| Computer Monitor and CPU | Manufactures |
| ABI 7500 real-time PCR system | Thermo Scientific, Darmstadt, Germany |
| Tabletop ultracentrifuge TL-100 | Beckman Coulter, Fullerton, CA, USA |
| Abbott RealTime HIV-1 m2000rt viral load assay equipment with automated nucleic acid extraction station | Abbott Molecular Inc., Des Plaines, IL, USA |
| FLUOstar OPTIMA SP002 BMG | BMG LABTECH, Ortenberg, Germany |
| Illumina MiSeq™ | Illumina, Inc., San Diego, CA, USA |

## 2.2 Study population

The study population consisted of 86 HIV-1 infected drug naïve patients originating from Nouna Burkina Faso who were mostly followed up either for a period of 48 weeks or 72 weeks. Eligible patients were identified when HIV infected patients came to the Nouna district hospital in Nouna Burkina Faso, for HIV counseling and care. This study was made possible through an established collaboration between the Virology Unit of the Department of Infectious Diseases of the University Hospital, University of Heidelberg and the Centre de Recherche en

Sante de Nouna (CRSN) in Burkina Faso. Blood samples as well as all other clinical and socio-demographic data were collected within a period of six years. The blood samples were separated into plasma and buffy coats. The patients were then followed up and blood samples were consecutively collected from January 2009 to August 2010 in Nouna. The longitudinal patient samples were collected at 2, 4, 12, 24, 36 and 48 weeks after patients had been put on HAART. In a few instances patients were followed up at 68 and 72 weeks post HAART and had their blood samples collected. Patients followed over time were on a first line HAART regimen of two NRTI (3TC, d4T or AZT, d4T) and 1NNRTI (NVP or EFV).

## 2.3 Study design and sample size

This consisted of two parts, a cross-sectional part, consisting of 86 HIV-1 infected drug naïve patients attending the Nouna district hospital in Nouna Burkina Faso whose paired templates of DNA and RNA were genotyped and sequenced for the presence of DRMs in the protease-reverse transcriptase ((PR/RT) region. This was followed by resolving differences in resistance mutations and nucleotides changes within patient cDNA and DNA sequences of the same patient with deep sequencing (details are summarized in figure 9). The HIV-1 IC DNA levels of all the drug naïve patients were also quantified using a molecular-beacon-real-time qPCR assay. The second part was a longitudinal study consisting of 8 patients whose plasma samples were genotyped and again sequenced for the presence of DRMs in the PR/RT region of HIV-1 over time. The follow-up for DRMs over time was done because these patients did not attain virologic suppression at 24 weeks of therapy (see section 2.3.1 for details on sample groupings).

The $CD4^+$ T-cell counts and plasma viral load of patients prior to initiation of HAART and at all-time points during the follow-up period were determined in Burkina Faso, West Africa and undetermined plasma viral loads were determined in Heidelberg Germany.

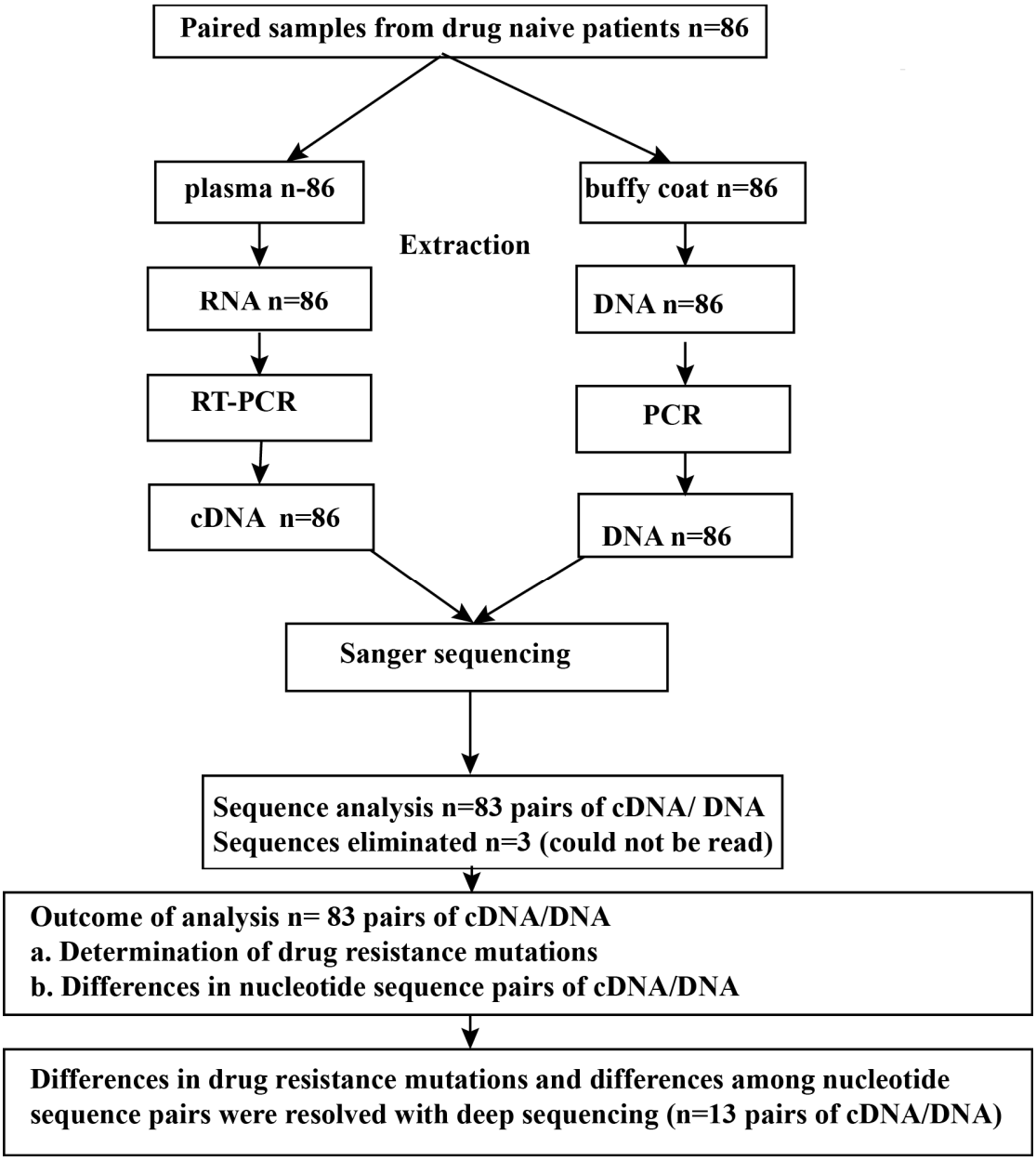



**Figure 9: Flow chart on genotyping and sequencing of HIV-1 infected samples from Nouna Burkina Faso.**

### 2.3.1 Sample groupings and characteristics

Various groups of HIV-1 infected samples were genotyped for the presence of DRMs. The first group of samples was paired samples of viral RNA and DNA extracted from plasma and buffy coats of 86 drug naïve patient samples and used as starting templates for drug resistance genotyping and HIV subtyping. HIV-1 genotyping and sequencing were used to

determine the presence of transmitted drug resistance mutations (TDRMs) and to determine similarities existing between RNA and DNA template pairs of the same patient. Three patient samples were eliminated from the analysis because the sequences were too heterogeneous and could not be read. The remaining 83 pairs were used in the analysis. The workflow for the entire genotyping and sequencing process is shown in figure 9.

The second group consisted of 8 selected patient plasma samples followed longitudinally over time from the drug naïve stage. These patients had high plasma viral loads (≥ 1000 copies per ml of plasma) at 24 and/or 48 weeks of antiretroviral therapy (ART). Patients with plasma viral loads depicting viral blips due to for example non-adherence to therapy were excluded from this group. The additional criteria for the selection of these 8 samples was that the patients had plasma samples available for at least two of the following time points 24, 48, 64 and 72 weeks post ART and a plasma viral load of ≥1000 copies per ml of plasma. The later criterion was to enable amplification of a PCR product that could be sequenced. The genotyping and sequencing process was the same as indicated in figure 9 except that samples were longitudinal samples collected over time. Drug resistance patterns were assessed.

The third group was made up of retrospective genotyping and sequencing of 5 patient plasma samples over all other remaining time points (4, 12 and 36 weeks post ART). These patients were selected from the second group above. The criterion for the selection was based on the evidence that the patient had major DRMs in RT. This mutation(s) should have either persisted at all time points or an additional mutation(s) developing over the course of infection from one time point to the other or a mixture of both of the aforementioned scenarios occurring simultaneously.

The fourth and final group comprised of selected drug naïve paired patient samples whose RNA and DNA were extracted from plasma and buffy coats respectively and used as starting templates in deep sequencing. These deep sequenced samples had two main characteristics. Firstly mutations found in the PR/RT region of paired RNA and DNA templates from the same patient using traditional Sanger sequencing were differing and secondly aligned nucleotide sequence pairs derived from Sanger sequenced paired RNA and DNA as stating templates had more than 10 changes.

## 2.4 Ethical clearance

Ethical clearance was obtained from the Centre de Recherche en Santé de Nouna (CRSN), Nouna Burkina Faso and the Virology Unit of the Department of Infectious Diseases of the University Hospital, Heidelberg (refer to appendix I and II for samples of the ethical clearances). All ethical considerations were strictly adhered to.

## 2.5 Informed consent

Patients who agreed to be part of the study filled informed consent forms before being recruited into the study.

## 2.6 Questionnaires and clinical data on patients

Each patient had their clinical and socio-demographic data recorded via laboratory diagnosis and interactive sessions respectively. Clinical data obtained included the patient's HIV status, co-infection with Hepatitis B and C (HBV and HCV), presence or absence of secondary bacterial or parasitic infections such as candidiasis, tuberculosis, salmonellosis and toxoplasmosis, Hb levels, CDC clinical staging of the HIV-1 infection, the type of anti-retoviral treatment, body mass index as indicated by the patient's weight in kilograms and their height squared (Beaudrot et al., 2015). The socio-demographic data consisted of age, gender, marital status, level of education, number of children, knowledge of the HIV status and possible exposure of sexual partner to HAART.

## 2.7 Inclusion and exclusion criteria

All HIV-1 infected drug naïve patients attending the Nouna district hospital in Nouna Burkina Faso and were not on any ARTs were included in the study. All patients registered at the Nouna district hospital in Nouna Burkina who were already on ARTs were excluded from the study.

## 2.8 Viral load and $CD4^+$ T-cell count determination

Patient plasma viral load determination was carried out at the Virology Unit of the Center Hospital Universitaire Yalgado Ouedraogo (CHUYO) in Ouagadougou with the Abbott RealTime HIV-1 assay with the automated m2000 system (Abbott Molecular Inc., Des Plaines, IL, USA) and for those samples where viral load was not determined before being shipped to Heidelberg, Germany, viral load was determined by the Abbott RealTime HIV-1 assay with the automated m2000 system (Abbott Molecular Inc., Des Plaines, IL, USA) at Virology Unit of the Department of Infectious Diseases of the University Hospital, Heidelberg. The $CD4^+$ T-cell counts of infected patients were determined from whole blood within 6 hours of sample collection with a fluorescence activated cell sorter (FACS) count flow cytometer (Becton Dickinson and Company, San Jose, USA).

## 2.9 Genomic DNA extraction from buffy coats

Genomic DNA of drug naïve and selected longitudinal samples were extracted from 140 µl of each patient buffy coat using the QIAmp DNA Blood Mini kit (Qiagen, Hilden, Germany) and eluted twice with 50ul elution buffer following the manufacturer's instructions. DNA quality and quantity was checked for all samples by UV spectrophotometry using NanoPhotometer P300 (Implen, München, Germany). DNA concentrations of patient samples ranged from 40 ng/µl to 195 ng/µl and were enough for PCR reactions.

## 2.10 Viral RNA extraction from plasma

Viral RNA was extracted from 140ul or 500ul (depending on the plasma viral load) of each patient plasma sample using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) and eluted twice with 40 µl elution buffer following the manufacturer's instructions. The quality and quantity of viral RNA was checked for all samples with UV spectrophotometry using NanoPhotometer P300 (Implen, München, Germany). RNA concentrations of patient samples ranged from 33ng/ul to 108ng/ul and were also sufficient for PCR reactions.

## 2.11 Reverse transcriptase polymerase chain reaction and nested PCR of viral RNA and genomic DNA samples

HIV-1 sequences encoding approximately 1461bp of gag (p6) and the pol gene consisting of the entire PR region of 99 amino acids (aa) and 2/3 of the RT region (between 312 to 335 aa) as indicated in figure 10 were amplified from each patient sample by both nested polymerase chain reaction (PCR) and RT-PCR, using extracted genomic DNA and viral RNA respectively. Predominant genotypic DRMs to antiretroviral drugs (Sage et al. 2010) were determined by sequencing of purified PCR amplicons within the pol gene of HIV-1 group M strains. The PCRs were performed with primers spanning the p6 of HIV-1 gag, PR and p51 of the RT region of the HIV-1 pol gene thus from nucleotide position 2078 to 3539 (see table 4 for PCR and sequencing primers and their respective HBX2 positions as well as alternative primers used for sequencing and figure 10, for a schematic diagram of the region amplified). The outer primers were1832 (F1) and 3583 (R1) corresponding to the HBX2 positions 1832-3583 within the HV-1 pol region. The inner primers were 2078 (F2) and 3539 (R2) also corresponding to the HBX2 positions 2078- 3539 within the HV-1 pol region (Kousiappa et al., 2009).

Viral RNA samples were reverse transcribed with the SuperScript III One-Step RT-PCR System (Life Technologies, Darmstadt, Germany) by following the manufacturer's protocol. The first round RT-PCR amplification mix for HIV-1 gag-pol region was prepared in 0.5ml PCR Eppendorf tubes (Eppendorf, North America) as follows; 15 µl of 33- 108 ng/µl l of viral RNA template, 25 µl of 2x RT-PCR reaction mix, 2 µl of 10 µM/µl outer primer (F1), 2 µl of 10 µM/µl outer primer (R1), 1 µl of 5 units SuperScript III RT/ Platinum Taq high fidelity enzyme mix and nuclease free water (Life technologies, Darmstadt, Germany) added to make a total unit volume of 50 µl. The thermocycling conditions was as follows; one cycle at 52°C for 1 h, 1 cycle at 94°C for 2 min, 40 cycles at 94°C for 15 s, 52°C for 30 s, 68°C for 2 min, one cycle at 68°C for 5 min and 4 °C at infinity (Kousiappa et al., 2009).

The first round PCR reaction mix for the genomic DNA included approximately 60–100 ng of patient DNA with 2 µl of 10 µM/µl each forward and reverse outer primers F1 and R1 (table 4) and 43µl of 1.1X Platinum PCR SuperMix (Life Technologies, Darmstadt, Germany) in a 50 µl unit volume. The thermocycling conditions were as follows, one cycle at 94°C for 2 min,

45 cycles at 94°C for 20 s, 52°C for 30 s, 72°C for 2 min, and one cycle at 72°C for 10 min (Kousiappa et al., 2009).

The second round PCR reactions of the cDNA and the genomic DNA included 60–100 ng or 3 µl of amplified patient DNA with 2 µl of 10 µM/µl each forward and reverse inner primers F2 and R2 (table 4) and 43µl of 1.1X Platinum PCR SuperMix (Life Technologies, Darmstadt, Germany) in a 50 µl unit volume. The thermocycling conditions were as follows, one cycle at 94°C for 2 min, 45 cycles at 94°C for 20 s, 52°C for 30 s, 72°C for 2 min, and one cycle at 72°C for 10 min (Kousiappa et al., 2009). The Eppendorf Master Cycler (Eppendorf, Hamburg, Germany) or the PTC-200 PCR Thermo cycler (GMI, Ramsey, MN, USA) was used for all the PCR amplifications.

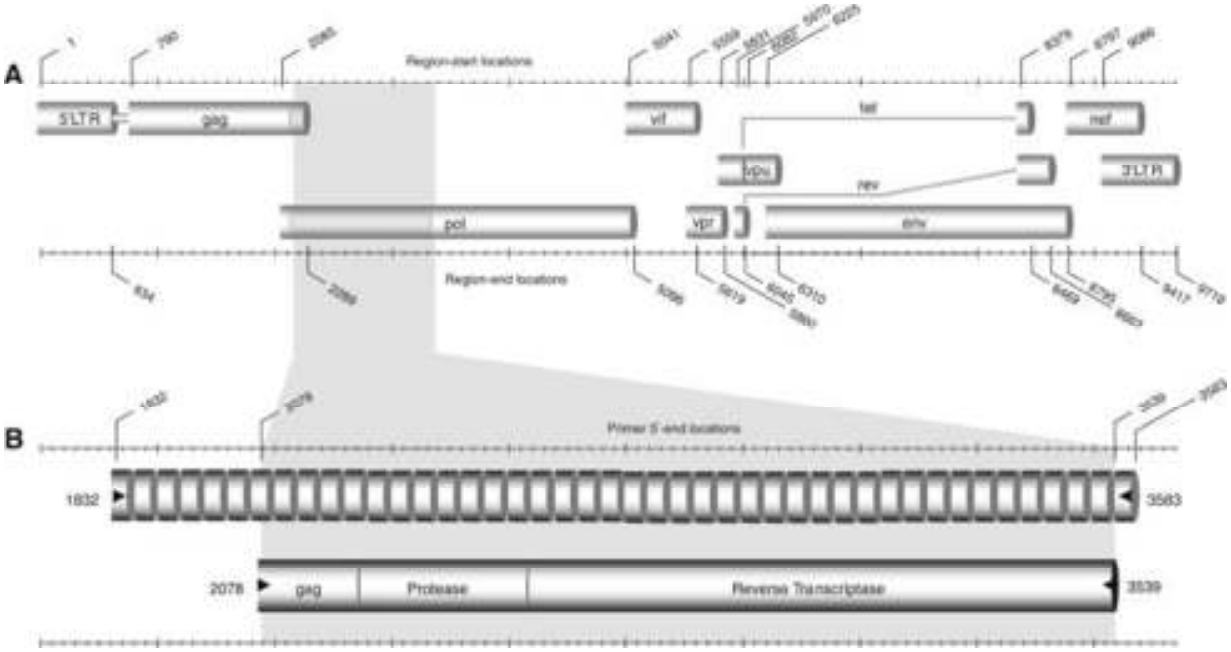



**Figure 10: A schematic diagram highlighting the PR/RT of HIV-1 pol from the M group.**

**This drawing shows the region genotyped among HIV-1 infected patients and assessed for the presence of HIV-1 genotypic drug resistance.(A) Shows the entire HIV genome architecture (as can be seen from the GenBank accession number K03455) and the corresponding HXB2 regions shown on the rulers above and beneath represent demarcations for each genetic region of the viral genome. The rulers above and beneath the viral genome also have segments of 100 base pairs from which the length of each of the regions in the viral genome can be estimated. The gag (p6) and pol (PR and p51 RT) regions amplified in all patient derived RNA and DNA samples are shaded in grey. (B) This shows the region genotyped in the nested PCR and RT-PCR reactions for all patients. The first round PCR primers corresponded to the HXB2 positions1832 and 3583**

**and these were also used as primer names for both the forward (FI: 1832) and reverse (R1: 3583) primers in the PCR mix. The second round PCR primers also corresponded to the HXB2 positions 2078 and 3539 and these were also the primer names for both the forward (F2: 2078) and reverse (R2: 3539) primers in the PCR mix (Demetriou et al., 2010).**

### 2.11.1 PCR optimizations

Ribonucleic acid and DNA PCR reactions for the HIV-1 pol region were set-up with in-house protocols as shown in appendix III at 50°C annealing temperatures for both RNA and DNA reactions. PCR conditions for the HIV-1 pol region were optimized for cDNA synthesis and first round PCR initially with SuperScript™ III One-Step RT-PCR System with Platinum® taq polymerase High Fidelity enzyme (Life Technologies, Invitrogen) and components were pipetted as indicated in appendix III. Template RNA concentrations were increased by 5 µl for each trial as shown in appendix III. Primer sequences were as indicated in the table 4. A nested PCR reaction with the verbatim high fidelity DNA polymerase enzyme (Thermo Scientific, Darmstadt, Germany) was set-up as indicated in appendix III. The PCR primer sequences for the second round reaction are given in table 4 above. The PCR consisted of 40 cycles with 1°C increments in annealing temperature of up to 60°C. Upon gel electrophoresis, no PCR band was realized although viral load of samples used was ≥ 500,000 copies/ml of plasma. Nested PCR for DNA samples were carried out with the same primers pairs and first and second round PCRs were done as shown in appendix III, but there were also no PCR bands found on agarose gels. Note that only one condition was changed at a time for each round of optimization.

Upon recommendations from a more experienced lab member who also worked on a similar project, primers were changed to the following combinations: cDNA and first round synthesis; 2.5 µl of 10 µM/µl each of Pol1 forward and RT-rev-stu1 reverse primers and the second round nested PCR reaction contained the same amounts of primer concentrations for Prot-for-EcoR1 forward primer and RT4 reverse primer as given in table 4. Nested PCR conditions were the same as shown in appendix III. No bands were seen on agarose gel photographs. PCR reagents were changed to the phusion high fidelity DNA polymerase (New England BioLabs, Frankfurt, Germany) and the same PCR conditions were repeated with the same primers and primer concentrations as mentioned above with 10 µl of 5x Phusion high fidelity or GC rich buffer but again no bands were seen after taking gel photographs.

Polymerase chain reactions were again tried with various primer combinations as shown in appendix III using the phusion high fidelity DNA polymerase (New England BioLabs, Frankfurt, Germany), the same PCR cycling conditions as described in the paragraph above. This procedure also yielded no results.

Three plasma samples were selected and tried on the ViroSeq HIV-1 Genotyping Kit (Abbott GmbH & Company, Wiesbaden, Germany) at the Diagnostic section of the University hospital, Heidelberg Germany following the manufacturer's protocol and amplification was sucessful. The experiment therefore proved that patient RNA samples were still viable for PCR reactions. In order to preserve patient samples for the actual experiments, 10 plasma samples from HIV infected patients from the diagnostic unit of the Heidelberg University Hospital were obtained and RNA extractions were done with the ViroSeq HIV-1 RNA extraction Kit (Abbott GmbH & Company, Wiesbaden, Germany) and the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany). PCR was again repeated with the ViroSeq HIV-1 Genotyping Kit which had its own primer sets (Abbott GmbH & Company, Wiesbaden, Germany) and the SuperScript™ III One-Step RT-PCR System with Platinum® Taq polymerase High Fidelity enzyme (Life Technologies, Invitrogen) and the Phusion High Fidelity DNA Polymerase PCR kit (New England BioLabs, Frankfurt, Germany). The ViroSeq protocol worked but the SuperScript RT-PCR System with the Phusion PCR protocols did not work. The annealing temperature was at 55°C for 45 minutes during the cDNA synthesis and at 55°C for 30s for both first and second round DNA synthesis. The ViroSeq primers could not be used with my PCR protocol because the primer sequences and concentrations were unknown.

As a proof of concept, further proceeded by using previously designed primers sets that specifically amplify a 400bp region spanning the 5' end of HIV-1 RT to amplify the Burkina Faso samples (Tebit et al., 2006). These primers were tried again on 5 pairs of RNA samples extracted with the ViroSeq HIV-1 RNA extraction Kit (Abbott GmbH & Company, Wiesbaden, Germany) and the QIAamp Viral RNA Mini Kit. PCRs were performed with the combination of the SuperScript RT-PCR System with the Phusion PCR protocols using the same conditions as described in the previous paragraph and all 5 pairs of samples were amplified. This suggested that the PCR primers used previously were not working and had to be changed. The primers were BFp1 for and BFp4b rev were first round primers and BFp3 for and BFp4c rev were second

round primers (table 4). Although these set of primers worked well, they could not be used to amplify the expected region for the purpose of this study.

New sets of HIV-1 pol primers were designed with the help of Stefan Seitz from the Molecular Virology Department of the UniversitätKlinikum, at the Otto Meyerhof Centre in the Heidelberg University Hospital. We designed HIV-1 sequences encoding approximately 1300bp of gag (p6), PR and about 255 aa in RT. These primers were designed based on consensus sequences of an HIV-1 CRF02_AG subtype isolated in 2006 with the GenBank accession number AB231895. These primer pairs were namely New F1 and New R1 used in the first round PCR reaction and New F2 and New R2 used in the second round PCR reaction. These primers were used together with the SuperScript RT-PCR System and Phusion PCR protocols, which gave non-specific bands including the band of interest on agarose gel. Gradual 1°C increments of annealing temperatures and 30s increments on annealing times did not resolve the problem. Activation of PCR reactions at 95°C for 5 min to denature pre-annealed primers, before addition of the polymerase also did not resolve the problem.

Since the protocol described above did not work efficiently, new PCR protocols were adapted from Kousiappa et al 2009. The protocol did work along with its published primers (table 4) and with minor modifications as described in section 2.11. The amplified region was about 1460bp of the HIV-1 pol region.

## 2.12 Agarose gel electrophoresis and DNA quantification

Ten microliters of amplified products were loaded onto ethidium bromide or 0.05% midori green (NIPPON Genetics Europe GmbH, Düren, Germany) stained 1.0 % agarose gels in 1 × TAE buffer (Brody et al, 2004; Sambrook et al, 2001). A 1kb plus molecular weight marker (Fermentas, Thermo Scientific, Schwerte, Germany) was always loaded on gels along with PCR products, to confirm expected molecular weight of the amplification products which were run at 90 volts for 45minutes. Gel products were then visualized under a spectrophotometer or a transilluminator (Beckman Coulter, Fullerton, CA, USA). If a single distinct band of the expected molecular weight was seen in lanes containing patient amplicons, then PCR purifications were carried out on the remaining 40 µl of amplified PCR products. On the other hand, for instances where non-specific bands were seen along with the expected bands, the

remaining 40 µl of amplified PCR products were run again on 1% agarose gels and the expected bands were excised from the gel and used in a gel purification assay. DNA concentrations of purified products were afterwards quantified by UV absorbance spectrophotometry using the NanoPhotometer® P 300 (Implen, München, Germany).

## 2.13 Purification and sequencing of amplified PCR products

Amplified products generated from the second round PCRs were purified using the NucleoSpin Extract II PCR clean-up and gel extraction kit (Macherey- Nagel GmbH and Co.KG, Dueren, Germany) following the manufacturers protocol. Minor alterations were made to the protocol to increase DNA yield. These included centrifuging PCR amplicons for 1 minute at 11,000rpm and eluting DNA only after heating DNA filled columns having elution buffer between 5-8 minutes at $70^0$C. Elution was done twice with 20 µl elution buffer to concentrate DNA. DNA concentrations were quantified by UV absorbance spectrophotometry using the NanoPhotometer® P 300 (Implen, München, Germany).

## 2.14 Sanger sequencing of PCR amplified DNA and cDNA samples

All purified PCR amplicons of DNA and cDNA were sequenced commercially from GATC Biotech (Konstanz, Germany) with the ABI 3130 genetic analyzer (Applied Biosystems, Foster City, CA). Briefly, purified second round PCR products containing 1461bps of HIV-1 PR and RT were bi-directionally sequenced using the sequencing HIV-1 pol primers annealing at HBX2 positions 2454 and 3019 respectively. In PCR amplicons where sequencing could not be achieved with the aforementioned primers, pairs of the following alternate gag-pol primers were used for sequencing 2136, 2216, 2610, 2650, 2734, 3003, 3462, and 3539 (table 4 and appendix VIII). The primer names also represent their annealing positions on the HIV-1 HBX2 template DNA sequencing reactions were analyzed with the BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA) according to the manufacturer's recommendations. Samples exhibiting partial or extensive viral diversity by direct sequencing were re-sequenced twice and if they still remained heterogeneous, the other sequence pair which in all cases was successfully sequenced was checked to make sure that they did not contain DRMs or unique HIV-1 subtypes before being eliminated from the sequencing analysis.

## 2.15 Determination of drug resistance mutations among HIV infected samples over time

Preceding sequence editing, patient derived HIV-1 PR/RT paired sequences of DNA and cDNA, encoding approximately 1461bp of gag (p6) and the pol gene consisting of the entire PR and 2/3 of the RT region obtained, were aligned to HIV reference sequences from the HIV sequence database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html) using Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html).

In order to identify mutations known to confer resistance to PR/RT inhibitors, the PR/RT sequence of each patient sample was analyzed with the Stanford HIV drug-resistance algorithm (Rhee et al., 2003). The Stanford HIV drug-resistance algorithm compares the query sequence to consensus subtype B reference sequences of HIV that have shown resistance to ARTs. The result differences determined between the query sequence and the consensus subtype B reference sequences are then used as an input for the assessment DRMs from the database. Genotypic drug resistance was defined as the presence of at least one DRM affecting the drug of any class of ARD as given by the Stanford HIV Drug Resistance database (Shafer, 2006) and the International AIDS Society (Avila-Rios et al.)-USA database(Johnson.VA et al., 2013).

Analyses of the possible impact of DRMs on therapeutic response among drug naive patients and selected patients followed over time were assessed using the drug resistance penalty score within the Stanford HIV drug-resistance database. This score assigns a mutation penalty to DRMs conferring resistance to any particular class of ARDs. The penalty scores for mutations within PR or RT conferring resistance to a particular PR or RT inhibitor were then summed up for this class of ARV and placed under one of five categories of drug resistance estimates defined by the Stanford HIV drug-resistance database. The estimates and their interpretations are as follows;

a. A range of 0 to 9 meant the virus was susceptible to the anti-HIV drug and presents no evidence of mutation.
b. A 10 to 14 range denoted potential low-level resistance to the ARV suggesting that the virus might have been exposed to the ARV at one point in time.

c. An estimate of 15 to 29 was indicative of low-level resistance depicting a reduced in-vitro drug susceptibility of these virus isolates. Patients infected with viruses in this category, are said to portray suboptimal virologic response to ARVs.
d. A 30 to 59 range showed an intermediate drug resistance level which was greater than low-level resistance but lower than high-level resistance
e. A level >60 depicted the high-level resistance and patients infected with viruses in this group, portray virologic failure to ARVs.

### 2.15.1 Analysis of HIV-1 drug resistance mutations over time

Eight patient plasma samples with high plasma viral load values at 24 weeks into therapy were selected and RNA was extracted from these samples. A cDNA synthesis was then performed and samples were then purified and commercially sequenced by Sanger sequencing. Resultant sequences generated from these patient samples were then aligned with wildtype reference sequences and edited using sequence chromatographs. DRMs were then determined by feeding edited sequences into the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu). In a few instances other expert list like the IAS-USA DRM panel (Johnson.VA et al., 2013) was consulted.

DRMs were determined for eight patients followed over the time points 24, 48, 64 and 72 weeks post ART. Since the plasma viral load is important in determining the probability with which a given HIV infected plasma sample could be amplified, the plasma viral load of each of the eight patient samples available at each time point was ≥1000 copies per ml of plasma (also the level generally considered as optimum for amplification of viral cDNA).

## 2.16 HIV-1 subtyping by phylogenetic tree analysis

Patient HIV-1 subtypes were determined from cDNA and DNA templates from the Stanford HIV Drug Resistance Database (Shafer, 2006) and then by phylogenetic tree analysis. The Molecular Evolution Genetic Analysis (MEGA) software version 6.0 was used to align DNA and cDNA sequences from drug naïve patients' samples, calculate distances, and construct phylogenetic trees (Tamura et al., 2007). All patient DNA and cDNA sequences encoding the

PR/RT region of HIV-1 pol viral genome were aligned against corresponding reference sequences of genetically characterized HIV-1 strains obtained from the Los Alamos database (Leitner.T et al., 2005) using the default setting on the MEGA software version 6.0 (Tamura et al., 2013). The Kimura two-parameter distance estimation approach with a transition/ transversion ratio of 2.0 was used to calculate pairwise distance matrices. Phylogenetic trees were then constructed by the neighbor-joining method of the CLUSTAL X programme in the MEGA software version 6.0. The consistency of the phylogenetic clustering was tested using bootstrap analysis with 1,000 replicates (Felsenstein, 1985). Bootstrap values above 70 were considered as enough for subtype designation and subtype designation was re-confirmed with subtypes as generated from Stanford HIV Drug Resistance Database output (Shafer, 2006).

## 2.17 Comparison of RNA- and DNA-associated pol sequences

HIV-1 pol nucleotide sequences derived from plasma viral RNA and DNA and covering the PR/RT region were compared to determine similarities existing between the two templates. These sequences were analyzed in the positions known to confer resistance to ARDs within the PR and RT regions of HIV-1, in order to establish whether genotypic drug resistance testing on the two types of sequences showed marked differences or remained the same.

## 2.18 Deep sequencing of drug naïve patient plasma samples

Polymerase chain reaction amplicons spanning the PR/RT of the HIV-1 pol region were generated from thirteen drug naïve paired patients RNA and DNA templates by RT-PCR and direct PCR respectively. Amplicon generation, gel electrophoresis, purification and quantification protocols used were the same as already described in sections 2.11-2.13. The primer sets used were also the same. The thirteen patients either had differences in DRMs within paired samples of their viral DNA and RNA derived sequences or ten or more differences within aligned nucleotide sequence pairs of RNA and DNA templates or had both differences in DRMs and aligned nucleotide sequence pairs of RNA and DNA templates.

Polymerase chain reaction products of all thirteen patients were then sent to the Institute of Immunology and Genetics, a commercial deep sequencing facility in Kaiserslautern, Germany where we purified the PCR products using the Agencourt AMPure XP PCR

purification system (Beckman Coulter, Inc., Beverly, Massachusetts, USA), which makes use of paramagnetic beads. The DNA was quantified with Quant-iT Picogreen dsDNA reagent and kits (Invitrogen, Life Technologies) using FLUOstar OPTIMA SP002 (BMG LABTECH, Ortenberg, Germany). The Picogreen kit includes a plasmid standard whose concentrations were used in determining the concentrations of the unknown PCR amplicons. After the quantifications, PCR amplicons were diluted and 1 ng DNA from each sample was used in generating DNA libraries. The Nextera DNA library preparation kit was used for generating DNA libraries by following the manufacturer's protocol (Illumina, Inc., San Diego, CA, USA). Shortly, DNA amplicons were fragmented into approximately 250bp at 55°C for 5 minutes in a process termed tagmentation. Unique index primers and adapters were then added in a short cycle PCR (12 cycles), a process termed barcoding. The process of fragmenting DNA and tagging them with index primers and adapters is termed Nextera tagmentation. After tagmentation, the PCR products were again purified, followed by normalization of the DNA library. The library for each DNA sample were all pooled and sequenced in the Illumina Miseq sequencer (Illumina, Inc., San Diego, CA, USA) overnight. Each new read contained both the DNA fragment sequence and the unique barcode.

Sequencing was completed overnight and sequence data was demultiplexed using an in-house bioinformatics pipeline, which sorted out identical sequences, by barcodes. The number of reads per sample was around 200,000 and the coverage per nucleotide position was between 10,000 - 20,000 times. Sequence data was then loaded into the HIV Grade Database (Obermeier et al., 2012) for alignments to reference HIV subtype B sequences, generation of consensus sequences for each patient sample and identification of DRMs and differences within paired nucleotide sequences of RNA and DNA templates. Minority variants in up to 1% of the viral quasispecies were determined.

## 2.19 Quantification of intracellular HIV-1 DNA load in drug-naïve patients samples by molecular-beacon-based real-time PCR

A molecular-beacon-based real-time PCR assay which had been improved to specifically quantify HIV-1 IC DNA that have completed both first and second template switches (STS) also called the HIV-1 STS DNA or strong stop DNA was used to quantify HIV-1 IC DNA levels in all drug naïve patient DNA samples. This assay was adapted from a

combination of protocols from Demetriou et al., 2010 and Kostrikis et al., 2002. The general procedure for quantifying single sequences with nucleotide-specific molecular beacons and real-time PCR was used in the assay (Tyagi and Kramer, 1996). Primers used in the real-time PCR assays spanned regions from the end of HIV-1 U5 in the LTR region to the region before the gag sequence (figure 11). The primers spanned a 165-nucleotide region starting from positions 623 to 788 in the HXB2 HIV sequence. The target recognition sequence of the molecular-beacon and the PCR primers were designed by an exhaustive alignment process of all HIV-1 sequences available within the HIV sequence database (Demetriou et al., 2010) to bind to HIV-1 group M conserved regions. Table 4 show the primers and the target recognition sequence of the molecular-beacon. The HIV-1 real time PCR assay was optimized for in-house detection of all HIV-1 IC DNA forms that had undergone both first and second DNA template switches, including integrated and unintegrated DNA forms within drug naïve patient DNA samples.

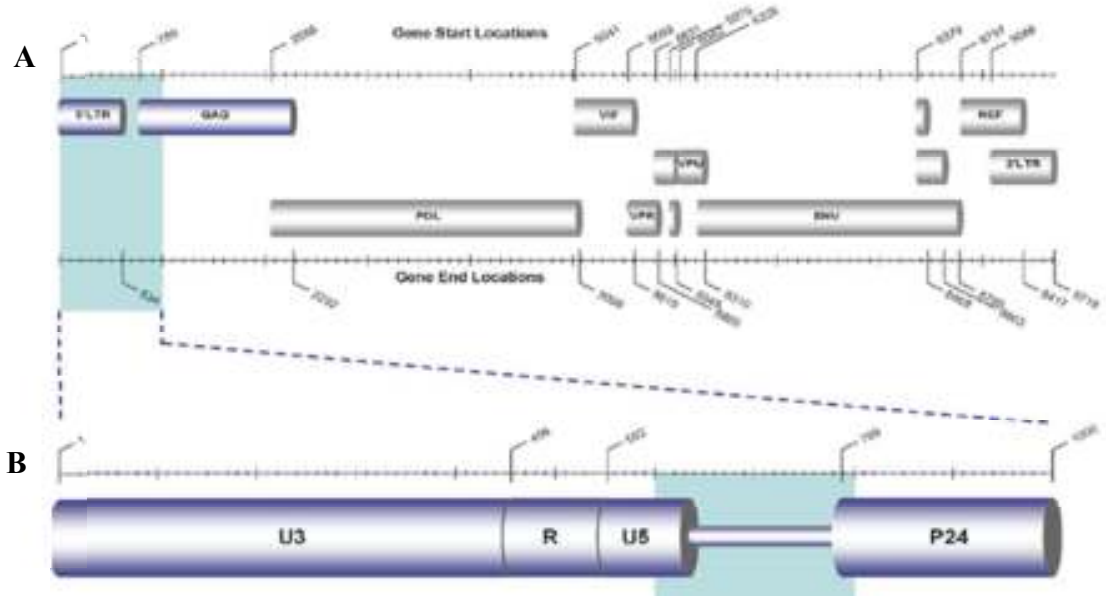



**Figure 11: Schematic diagram of the HIV-1 IC DNA region quantified from all drug naïve DNA samples.**

**(A) Indicates the entire HIV-1 genome architecture showing all gene locations and gene lengths. (B) Illustrates the magnified image of the LTR and p24 region of gag thereby detailing the exact region within the viral genome, which was targeted in the quantification of HIV-1 IC DNA forms.**

### 2.19.1 PCR amplification of HIV-1 intracellular and CCR5 DNA standards

Patient derived HIV-1 and CCR5 DNA were PCR amplified to generate PCR amplicons for plasmid production. These plasmids were then used as external standards in the real-time qPCR amplification of HIV-1 infected drug naïve patients DNA samples. The PCR products were generated with HIV-1 and CCR5 DNA amplification primers as indicated in table 4 above. The PCR amplification reaction conditions per 50 µl reaction mix and thermocycling conditions for both HIV-1 and CCR5 DNA were as shown in appendix IV. Electrophoresis of PCR amplicons were performed at 100 volts on 2% agarose gels stained with ethidium bromide. The expected PCR bands were 165 and 237 nucleotides for HIV-1 and CCR5 DNA (see appendix V) were then excised from gel and purified with the NucleoSpin Extract II PCR clean-up and gel extraction kit (Macherey- Nagel GmbH and Co.KG, Dueren, Germany) by following the manufacturers protocol. Amplicons were then used for cloning.

### 2.19.2 Cloning and transformation of HIV-1 and CCR5 plasmids into E.coli

Polymerase chain reaction amplicons of HIV-1 and CCR5 DNA were cloned into a pCR 4-TOPO vector obtained from the TOPO TA cloning kit (Invitrogen, Karlsruhe, Germany). The cloning reaction was set up according to the manufacturer's instructions as shown in appendix VI for HIV-1 IC DNA and CCR5 DNA. The pCR 4-TOPO plasmid vector (Invitrogen, Karlsruhe, Germany) was designed linearized with single 3′ deoxythymidine (T) overhangs. On the other hand, the Taq DNA polymerase within the PCR reaction mix has a non-template-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3′ ends of PCR products. This allows efficient ligation of PCR inserts with the linearized TOPO vector in the cloning reaction (http://redrecombineering.ncifcrf.gov/Protocols.html).

Transformation was carried out by adding 4 µl each of TOPO cloning reaction mix to two 50 µl vials of competent E. coli cells which had been thawed on ice and mixed gently. The mixtures were incubated on ice for 30 minutes and afterwards heat shocked at 42 °C in a water bath for 30s. The mixtures were then immediately cooled on ice for 2 minutes and cultured in 1 ml antibiotic-free LB medium for an hour in a 37°C incubator. Afterwards different concentrations of the mixture (20-50 µl) were streaked on pre-warmed LB-agar plates containing

50 µg/ml of selection antibiotics. The plates were subsequently incubated overnight at 37°C. Single colonies of eight to ten suspected to have the plasmid inserts of each sample, were selected and sub-cultured in 6 ml of LB medium containing 50 µg/ml of selection antibiotics overnight at 37°C overnight. Plasmid DNA was then extracted from 6 ml of these sub-cultures containing single or pure bacteria colonies using the QIAprep Spin MiniPrep kit (QIAGEN, Hilden, Germany) and extracted DNA were confirmed for expected HIV-1 or CCR5 DNA plasmid insert via sequencing with the TOPO TA sequencing primers (table 4) at the GATC Biotech (Konstanz, Germany). This was then followed with a second overnight sub-culture of 0.5ml pure colonies containing the required insert, at 37°C overnight in 400 ml of LB medium containing selection antibiotics. DNA extraction of cultures containing the plasmid insert were performed using QIAprep Spin MaxiPrep kit (QIAGEN, Hilden, Germany) to generate large volumes of the plasmids to serve as external standards in the real-time qPCR reactions. Plasmid DNA was suspended in Tris-EDTA (TE) buffer (components of the buffer are indicated in table 1) and stored at -20°C. Table 6 and 7 show the plasmid DNA inserts for HIV-1 and CCR5 sequence lengths (165 nucleotides and 237 nucleotides respectively) and the HBX2 locations of both inserts.

**Table 6: Query sequence of patient HIV-1 intracellular DNA inserted into a topo vector**

| | |
|---|---|
| Patient Derived HIV IC DNA | TCTCTCCTTCTAGCCTCCGCTAGtCAAAAATTTCTTGGCGTACTCACCGG |
| Sequence inserted Into Topo | TCGCCGCTCTCGCCTCTTGCTGTGTGCACCTCAGCAAGCCGAGTCCTGCG<br>TCGAGAGAACTTCTCTGGAACTTTCGCTTTCGAGTCCTTATTAACTTTCA<br>CTTTCGGGTCCCTGTTCGGGCGCCACTGCTAGAGATTT |
| Query Length: | **165** |
| HXB2 Location: | genome: **623←788** reverse complement |

**Table 7: Sequence of patient CCR5 amplicon inserted into a topo vector**

| | |
|---|---|
| Patient Derived HIV IC DNA | TCTCTCCTTCTAGCCTCCGCTAGtCAAAAATTTCTTGGCGTACTCACCGG |
| Sequence inserted Into Topo | TCGCCGCTCTCGCCTCTTGCTGTGTGCACCTCAGCAAGCCGAGTCCTGCG<br>TCGAGAGAACTTCTCTGGAACTTTCGCTTTCGAGTCCTTATTAACTTTCA<br>CTTTCGGGTCCCTGTTCGGGCGCCACTGCTAGAGATTT |
| Query Length: | **165** |
| HXB2 Location: | genome: **623←788** reverse complement |

Generated from http://www.hiv.lanl.gov/content/sequence/QUICK_ALIGNv2/QuickAlign.html

### 2.19.3 Plasmid quantifications and dilutions

Both the TOPO vector with either the HIV-1 or CCR5 plasmid DNA each consisted of 3956bp in size. The DNA concentrations of the plasmid HIV-1 and CCR5 DNA were 2268 ng/ul and 2638 ng/ul respectively. Plasmid DNA concentrations were quantified with the NanoPhotometer® P 300 (Implen, München, Germany). The DNA copy number per unit volume for each standard was calculated using the respective DNA concentrations measured. Tenfold serial dilutions of the HIV-1 and CCR5 plasmid DNA with a dynamic range of $10^1$- $10^6$ and $10^3$- $10^7$ DNA copies respectively, were freshly prepared and used for real-time qPCR. Low copy number DNA was maintained by using low DNA binding Eppendorf tubes for the dilutions, to ensure the highest recovery of DNA over time. Low CCR5 DNA copy numbers ($10^1$-$10^2$) could not be amplified.

### 2.19.4 Generation of HIV-1 and CCR5 standard curves with real-time PCR

Each real-time qPCR standard was carried out in triplicates in a 96 well plate. For the HIV-1 DNA standard, Each 50 µl real-time PCR mix consisted of 0.5 to 5.0µg of genomic DNA, 0.5 µl of 10 µM/µl HIV-1 MB684 molecular beacon, 2.5 µl of 10 µM/µl forward and reverse primers each, 10 mM dATP, 10 mM dCTP, 10 mM dGTP, 20 mM dTUP, 0.5 µl of 5 U/µl of AmpliTaq Gold DNA polymerase (Life Technologies, Darmstadt, Germany), 7 µl of 25mM $MgCl_2$, 5 µl of 10X TaqMan qPCR Buffer A (table 1 gives the constituents of the buffer). PCR grade water was added to the mix to make it up to 50µl per unit volume. The thermocycling conditions were set as follows; one cycle of denaturation (95°C for 10 minutes), followed by 40 cycles of amplification (denaturation at 95°C for 15 s, annealing and data collection at 62°C for 1 min, and polymerization at 72°C for 30 s in an ABI 7500 real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany). The molecular beacon had the FAM reporter and the quencher dabcyl. The target recognition sequence and primers are given in table 4.

The CCR5 DNA standard was generated for each 50 µl real-time PCR mix with the same reaction conditions as specified for the HIV-1 DNA qPCR with a further addition of the 0.50 µl of 1 U/µl of AmpErase Uracil-N-glycosylase to prevent amplification of dU-contaminating PCR products. The primers and the CCR5 molecular beacon containing the target

recognition sequence are listed in table 4. Also the thermocycling was carried out as follows; One incubation cycle for the Uracil-N-glycosylase (50°C for 2 minutes) and one cycle of denaturation at 95°C for 10 minutes, followed by 40 cycles of amplification as follows denaturation at 95°C for 15 s, annealing at 60°C for 1 min in an ABI 7500 real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany). Data was collected at all amplification stages and there was no polymerization cycle. Fluorescence emission was subsequently recorded at 521nm during the data-collection stage of each qPCR cycle.

HIV-1 and CCR5 standard curves were generated by constructing a scatter plot of known CCR5 and HIV-1 copies against the cycle threshold of each dilution point (Microsoft excel software 2010). The slope and co-efficient of determination $R^2$ of each of the two standard curves was calculated based on the equation of a straight line (CT=mx+b), where m and b represents the slope and intercept of the straight line respectively, x represents the independent variable and CT the dependent variable. The PCR efficiency, E, of the experimentally derived standards was computed from the equation, $E = (10^{-1/slope}-1)$.

## 2.19.5 Assessment of target specific binding to the HIV DNA template

The specific binding of HIV-1 DNA primers and molecular-beacons to the HIV-1 DNA template from the patient DNA samples was determined by using 10ul µl of 96 ng/µl of a non-template control (NTC) human HIV negative sample diluted in 90 ul of TE buffer and serially like the HIV-1 plasmid DNA standard. In this assay three experimental set-ups were made. The first consisted of qPCR mixes having the NTC and the HIV plasmid DNA template in the same qPCR mix per unit reaction. The dilution factor for both the HIV plasmid DNA and the NTC were the same. The qPCR mixes and thermocycling conditions were performed in the same manner as in the HIV plasmid DNA standard generation (section 2.19.4). Six dilutions points were setup with a dynamic range from $10^1$ to $10^6$ of HIV-1 DNA copies and for each dilution point, an NTC of the same dilution was added. The dilution points were set-up in triplicates. The second set-up consisted of the HIV-1 plasmid DNA only with a dynamic range from $10^1$ to $10^6$ also done in triplicates. The third reaction comprised of the human HIV negative control only, having again a dynamic range from a 1 in 10 dilution point to a 1 in 1000000 dilution point for each triplicate reaction. The three qPCR set-ups were run simultaneously on a 96 well plate.

### 2.19.6 Quantification of HIV-1 intracellular DNA levels in drug naïve patients DNA samples

Quantification of HIV-1 IC DNA of all drug naïve genomic DNA samples extracted from buffy coats was achieved with two molecular-beacon-based real-time qPCR assays. PCRs were carried out in triplicate in 96 well plates for the absolute quantification of CCR5 DNA copies and HIV-1 IC DNA copies for each patient sample, as described above. For each 96 well plate run, HIV-1 and human CCR5 DNA standard curves were also run alongside in triplicate using six serial dilution points as described above. Each 50 µl qPCR mix and thermocycling conditions were done as described in generating the standards. Each experimental qPCR output was assessed based on its own standard curve results. The co-efficient of determination ($R^2$) of each standard curve was determined by the slope and the intercept. For each patient DNA sample, the threshold cycles obtained from either the HIV-1 IC DNA or CCR5 qPCR were used to determine the HIV-1 IC DNA copies or the human CCR5 copy number. The HIV-1 IC DNA load was then defined as the number of HIV-1 IC DNA copies per one million cells. Cell numbers for each patient sample were obtained from the human CCR5 qPCR because it is already known that the CCR5 gene, exist as two copies in a cell.

## 2.20 Optimizations of absolute real-time quantitative PCR reactions

### 2.20.1 Optimizations of HIV-1 intracellular DNA quantification protocol

In order to quantify HIV-1 IC DNA from genomic DNA samples of infected drug naïve patients, an absolute qPCR reaction protocol employing the use of molecular beacons was optimized to specifically quantify HIV-1 IC DNA that has completed both first and second template switches.

#### 2.20.1.1 Quantification of HIV-1 intracellular DNA with the platinum qPCR SuperMix

Standards were prepared with pCHIV transformed into E.coli, cultured and plasmid purified from pure cultures. Known plasmid concentrations of $10^6$-$10^1$ pCHIV DNA copies were amplified with the platinum qPCR SuperMix at an annealing and data collection temperature of

55°C. The resultant amplifications ($10^6$-$10^5$) copies were non-specific and the other standard points gave no amplification.

A patient derived HIV-1 DNA plasmid was generated from PCR amplicons using the HIV-1 DNA primers, which were also used in the quantification of HIV-1 IC DNA. The qPCR reactions were performed with the platinum qPCR SuperMix-UDG (Life Technologies, Darmstadt, Germany) in the ABI 7500 real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany). Each reaction mix consisted of a total volume of 25 µl, containing 5 µl of extracted genomic DNA, 1 µl of 20 pmol/µl each of forward and reverse primer, 12.5 µl of 1X Platinum quantitative qPCR Supermix-UDG (Invitrogen, Carlsbad, CA, USA), 1 µl of 6 pmol/µl of the HIV DNA molecular-beacon (MB684) and 4.5 µl of nuclease free water. The cycling conditions were as follows: one incubation cycle at 50°C for 2 minutes, followed by one cycle of denaturation at 95°C for 10 min and 50 cycles of amplification consisting of a denaturation at 95°C for 15 s, annealing and data collection at 55°C for 30 s (annealing temperature was increased by 1°C at a time until an annealing temperature of 62°C was reached and annealing time was also increased to 1 min), polymerization was set at 72°C for 30 s. Specific amplifications were seen for $10^6$-$10^3$ HIV-1 DNA copies at an annealing temperature of 60°C for 1minute and the other two standard points gave non-specific amplification.

#### 2.20.1.2 Quantification of HIV-1 intracellular DNA with AmpliTaq Gold DNA polymerase

Patient derived HIV-1 DNA plasmid was used again as a standard for the absolute quantification of HIV-1 DNA copies using a molecular-beacons designed to hybridize with the HIV-1 templates in the M group (Demetriou et al., 2010). The dynamic range of the standard was $10^6$-$10^1$ HIV-1 DNA copies. The 50 µl reaction mix per unit was prepared for the qPCR with AmpliTaq Gold DNA Polymerase (Life Technologies, Darmstadt, Germany). All reactions were carried out in triplicates in 96 well plates and sealed with qPCR plate sealers. In the initial set of reactions, each 50 µl mix included 10 µl of genomic DNA, 0.5 µl of 10 µM/µl HIV-1 MB684, 2.5 µl of 10 µM/µl forward and reverse primers each, 0.5 µl of 25 mM dNTPs, 7 µl of 25mM $MgCl_2$, 5 µl of 10X TaqMan qPCR Buffer A (table 1 gives the constituents of the buffer). Nuclease free water was added to the mix to add up to 50µl volume. The thermocycling conditions were set as follows; one cycle of denaturation (95°C for 10 min), followed by 40

cycles of amplification (denaturation at 95°C for 15 s, annealing and data collection was set at 55°C for 1 min and increased gradually by 1°C in each experimental set-up to 62°C but the annealing time was kept constant. Extension was set at 72°C for 30. The ABI 7500 real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany) was used for all the HIV-1 DNA quantification reactions. Five dilution points ($10^6$-$10^2$ HIV-1 DNA copies) were quantified and the co-efficient of determination achieved was 0.95 which was lower than the co-efficient (0.99) achieved by Demetriou et al. 2010 and Kostrikis et al. 2002 and also the PCR efficiency achieved was above 100%.

In order to improve the qPCR efficiency and also quantify 10 copies of HIV-1 DNA, efficiently, the dNTPs was changed to single dNTPs and qPCR mix was prepared as already described in section 2.19.4 with the same thermocycling conditions. The annealing temperature was adjusted as well from 60°C to 62°C and the best results were obtained at the 62°C annealing temperature. All HIV-1 DNA dilution points ($10^6$-$10^1$ copies) were efficiently quantified with these conditions.

## 2.20.2 Optimizations of human CCR5 DNA quantification protocol

To quantify the numbers of cell input in the HIV-1 DNA, a molecular-beacon-based qPCR, which uses primers flanking the positions 478-623 bp of the human CCR5 gene, was used. A patient derived plasmid CCR5 DNA was used as the standard.

### 2.20.2.1 CCR5 DNA amplification with the platinum qPCR SuperMix

The standard was initially amplified with platinum qPCR SuperMix-UDG (Life Technologies, Darmstadt, Germany). For the real-time PCR reaction, each 25 µl reaction mixture contained 5 µl of extracted genomic DNA, 1 µl of 20 pmol/µl each of forward and reverse primer respectively, 12.5 µl of 1X Platinum quantitative qPCR SuperMix-UDG (Invitrogen, Carlsbad, CA, USA) and nuclease free water added to make the 25 µl volumes. The molecular-beacon (LK155) used in the 25 µl reaction mix was 13 pmol/µl. The cycling conditions were the same as in the HIV-1 IC DNA reaction and performed in ABI 7500 real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany). Data was collected

at 55 °C for 30s and in the repeat experiment data was collected at 55 °C for 1 min. Annealing temperature was increased by 1°C at a time, keeping all other conditions constant until an annealing temperature of 62°C was reached. The FAM and JOE fluorescence emission wavelengths 518 nm and 548 nm respectively were used instead of the TET filter (emission wavelength 538 nm) which is the reporter dye for the CCR5 molecular-beacon. The reason was that the ABI 7500 real-time spectrofluorometric thermal cycler does not have the TET filter. Amplification at 50 cycles gave non-specific bands. Quantitative PCR was again tried in the ABI 480 real-time spectrofluorometric light cycler (Thermo Scientific, Darmstadt, Germany). The higher standard points from $10^6$-$10^4$ CCR5 copies had non-specific amplification with varied CTs within each standard point and the lower standard points ($10^3$-$10^1$ copies) showed no amplification. Experiments were repeated twice and results were the same. Standards were afterwards prepared from $10^8$-$10^3$ CCR5 copies and specific amplification was detected for $10^8$-$10^7$ CCR5 copies only.

#### 2.20.2.2 CCR5 DNA amplification with AmpliTaq gold DNA polymerase

The CCR5 plasmid standards ($10^8$-$10^3$ CCR5 copies) were again amplified with the AmpliTaq Gold DNA Polymerase (Life Technologies, Darmstadt, Germany). The 50 µl reaction mix was performed as described in section 2.19.4 and thermal cycling conditions were also the same with no incubation cycle for the Uracil-N-glycosylase because the reaction mix did not contain this enzyme. The qPCR was run in the ABI 7500 real-time spectrofluorometric thermal cycler and also in the ABI 480 real-time spectrofluorometric light cycler (Thermo Scientific, Darmstadt, Germany). At the 62°C annealing temperature, three standard points were specifically amplified from $10^8$-$10^6$ CCR5 copies. Amplification products were run on 2% agarose gels at each run to detect the presence of specific and non-specific bands. Keeping qPCR reaction mix and reaction conditions the same the experiments were repeated using the ABI 7900 HT real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany) which has the TET reporter. Data was collected at 62°C annealing temperature. Amplification products obtained were non-specific at all data points. A 0.50 µl of 1 U/µl of AmpErase Uracil-N-glycosylase and the additional 50°C incubation step was added and data again collected at the annealing temperature and again non- specific amplifications were obtained.

The next round of amplifications were carried out at an annealing temperature of 60°C for 1 min in the ABI 7900 HT real-time spectrofluorometric thermal cycler (Thermo Scientific, Darmstadt, Germany) and data collection was performed at all thermal cycling points. There was no polymerization step added. These conditions eventually gave specific amplifications of $10^8$-$10^3$ CCR5 copies and this was used to quantify and the external standards were also used to measure the unknown CCR5 DNA copies of patients.

## 2.21 Statistical analysis

The Stanford HIV drug resistance algorithm was used to determine TDRM having an effect on HIV drug susceptibility for a particular class of ARVs. TDRMs to RT and PR inhibitors were displayed in percentages. The prevalence of transmitted resistance was determined with the Wilson score interval at 95% confidence interval (CI). A bar chart was used to display the distribution of HIV-1 IC DNA or HIV-1 STS DNA load among HIV-1 infected drug naïve patients. Descriptive statistics of median and interquartile ranges (IQR) were used for displaying HIV-1 IC DNA load/$10^6$ cells, plasma viral RNA load/ml of plasma, $CD4^+$ T-cell count/ml of blood and other relevant clinical parameters among infected patients. Pearson's correlation coefficient at 95% CI was used to evaluate the associations between HIV-1 IC DNA levels, plasma HIV-1 viral RNA levels and $CD4^+$ T-cell counts. The Mann-Whitney test at 95% CI, was used to compare HIV-1 IC DNA load, plasma HIV-1 viral RNA load, $CD4^+$ T-cell counts and all other clinical parameters among drug resistant and drug sensitive patients. A one-way ANOVA was used to demonstrate the association between age of patients and HIV-1 IC DNA load. Consequently the HIV-1 IC DNA levels among predominant HIV-1 subtypes depicted by analysis done with the Kruskal-Wallis test and the Bartlett's test for equal variances were also displayed. All statistical analysis were done with Microsoft excel 2010 or with the GraphPad Prism version 6.00 (Graph Pad Inc., San Diego, California, USA) softwares.

# 3 Results

## 3.1 HIV-1 infected patient demographics and clinical parameters

This section describes the characteristics of all drug naïve patients from Nouna Burkina Faso recruited for this study. The drug naïve HIV-1 infected patients consisted of 61 (71%) females, 24 (28%) males and 1(1%) with unknown gender. The median ages of females 34 (30–41) and males 42 (34.25–48.00) were significantly different (p=0.0076). The median $log_{10}$ plasma viral load and the $CD4^+$ T-cell count for all drug naïve patients were 5.757 copies per ml of plasma and 5.301 cells per ml of blood respectively. The total number of patients with records on the presence or absence of secondary bacterial infections was 60 and the remaining 26 had no records on their infection status. Among the 60 patients with infection history, 34 harboured secondary bacterial infections and the remaining 26 had no bacterial infections. None of the patients had secondary parasitic infections. Sixty-one patients had data on HBV co-infection status and the remaining 25 had no records. Among these 61 patients, those harbouring HBV as a co-infection with HIV-1 were six and those without HBV co-infection were 55. Fifty-five males and females had records on their Hb levels while the remaining 31 patients had no history on Hb levels. The median Hb levels for males and females were 11.7g/dl and 10.25g/dl respectively. With respect to the Centers for Disease Control and Prevention (CDC) classification system for HIV infection, 57 patients had their respective clinical staging categories recorded and per the record, a majority of patients 56 were within CDC category B and 1 patient was within the CDC category A. The remaining 29 patients had no records on their clinical staging status. Table 8 describes demographics and clinical characteristics of HIV-1 infected patients in this study.

**Table 8: Baseline characteristic details of drug naïve patients attending the Nouna District hospital in Nouna Burkina Faso**

The Mann Whitney t-test for unpaired data was used to estimate the median and IQR**.** Abbreviations IQR represent Interquartile Range; CDC represents the Centers for Disease Control and Prevention

| | **Drug NaïvePatients** |
|---|---|
| **Number of patients** (n=86, median age in years, IQR) | 35 (30–43) |
| Females {n=61 (71 %), median age in years, IQR} | 34 (30–41) |
| Males {n= 24 (28%), median age in years, IQR} | 42 (34–48) |
| Unknown gender (n, %) | 1 (1) |
| **Median $log_{10}$ transformed plasma RNA load in copies per ml of plasma (n=86, median, IQR)** | 5.76 (4.99 - 6.16) |
| **Median $log_{10}$ transformed $CD4^+$ T-cell count in cells per ml (n=86, median, IQR)** | 5.30 (5.08 -5.45) |
| **Patients' records on CDC Classification System for HIV Infection** | CDC-A/B/C (n=57) |
| CDC-Category A (n) | 1 |
| CDC-Category B2, B3 (n) | 55 |
| CDC-Category B1 (n) | 1 |
| **Recorded number of patients with infection status (n)** | 60 |
| Presence of Secondary bacterial or parasitic infection (n) | 34 |
| Absence of Secondary bacterial or parasitic infection (n) | 26 |
| **Recorded patients' HBV co-infection status (n)** | 61 |
| Presence of HBV Co-infection (n) | 6 |
| Absence of HBV Co-infection (n) | 55 |
| **Recorded patients BMI in $kg/m^2$ (n=58, median, IQR)** | 17.72 (20.59-15.76) |
| **BMI Categories:** | |
| Males Underweight, Below 18.5$kg/m^2$ (n=10, median, IQR) | 15.87 (14.74-17.41) |
| Females Underweight Below 18.5$kg/m^2$ (n=22, median, IQR) | 16.18 (14.54-17.28) |
| Males Normal Weight, 18.5–24.9$kg/m^2$ (n=4, median, IQR) | 22.51 (20.90-23.86) |
| Females Normal Weight, 18.5–24.9$kg/m^2$ (n=20, median, IQR) | 20.57 (20.36-21.80) |
| Overweight Female, 25–29.9$kg/m^2$ (n=1, actual BMI) | 26.17 |
| Obese Female, 30.0$kg/m^2$ and Above (n=1, actual BMI) | 35.16 |
| **Recorded patients Hb in g/dl (n=55, median, IQR)** | 10.60 (9.70-11.80) |
| Hb of Males in g/dl (n=15, median, IQR) | 11.70 (10.60-12.60) |
| Hb of females in g/dl (n=40, median, IQR) | 10.25 (9.63-11.23) |

## 3.2 HIV-1 subtyping of drug naïve patient RNA and DNA sequences by phylogenetic tree analysis

In order to determine the HIV-1 subtypes found among drug naïve patients, edited patient sequences were fed into the Stanford HIV Drug Resistance Database

(http://hivdb.stanford.edu) for subtype determination. Subtypes were again confirmed using phylogenetic trees as shown in figures 12A, 12B, 12C, 12D, 12E and 12F. Phylogenetic trees were generated with patient RNA and DNA sequences encoding the PR/RT of HIV-1 pol gene. The predominant HIV-1 subtype found among patient sequences was CRF02_AG- like viruses. These formed 70% (58/83) of the entire patient sequence pool analyzed, CRF06_cpx -like viruses were the second abundant group of viruses 14.5% (12/83), the third was CRF01_AE-like 7.2% (6/83), the fourth virus subtype was G-like viruses 5% (4/83), the fifth subtype was 02_A1-like viruses 2.4% (2/83), the sixth and final was subtype D- like viruses 1.0% (1/83). There were no major differences between RNA and DNA virus subtypes of the same patient.

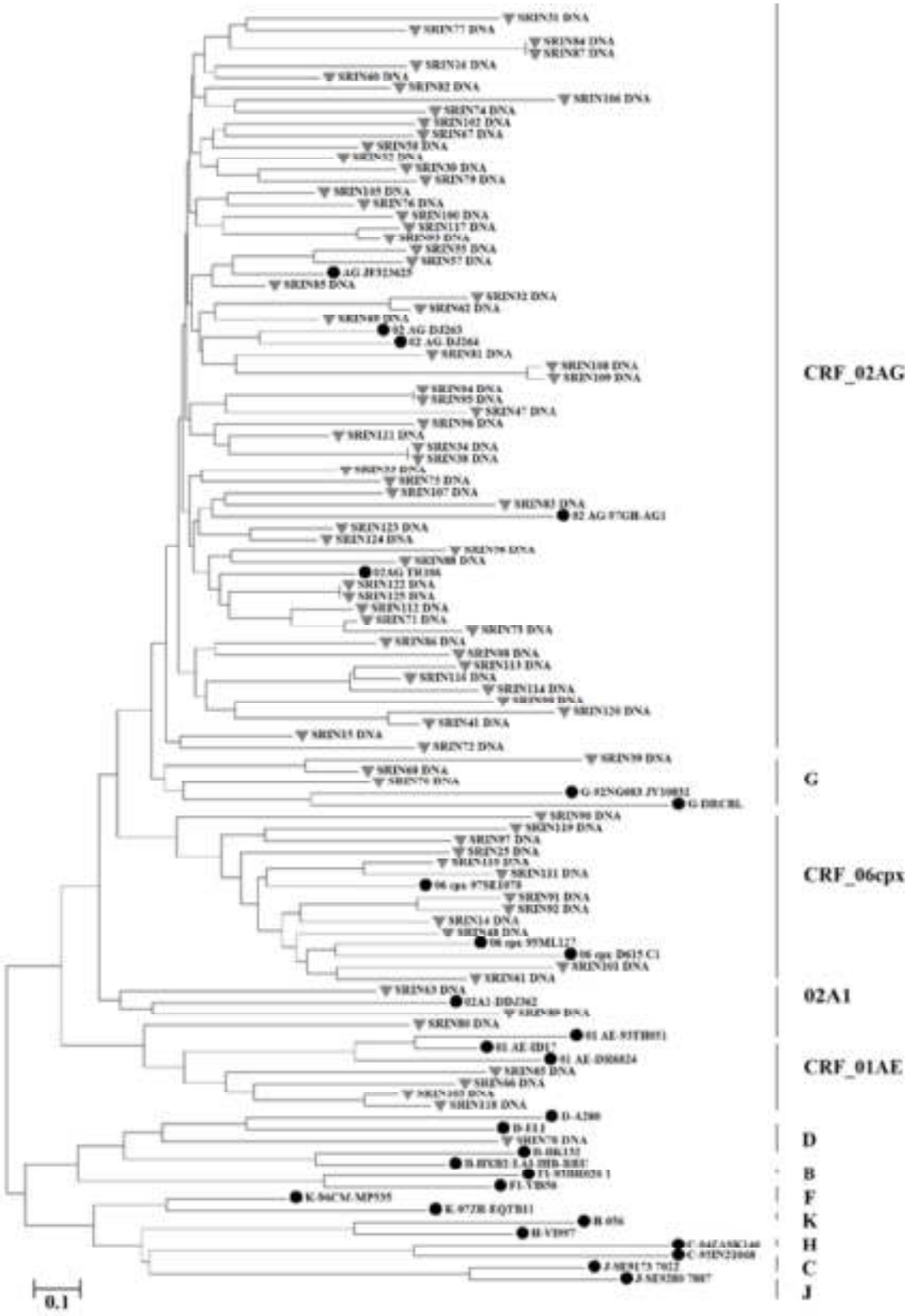



**Figure 12A: Phylogenetic tree from DNA as starting template. See details on pages 67-68.**

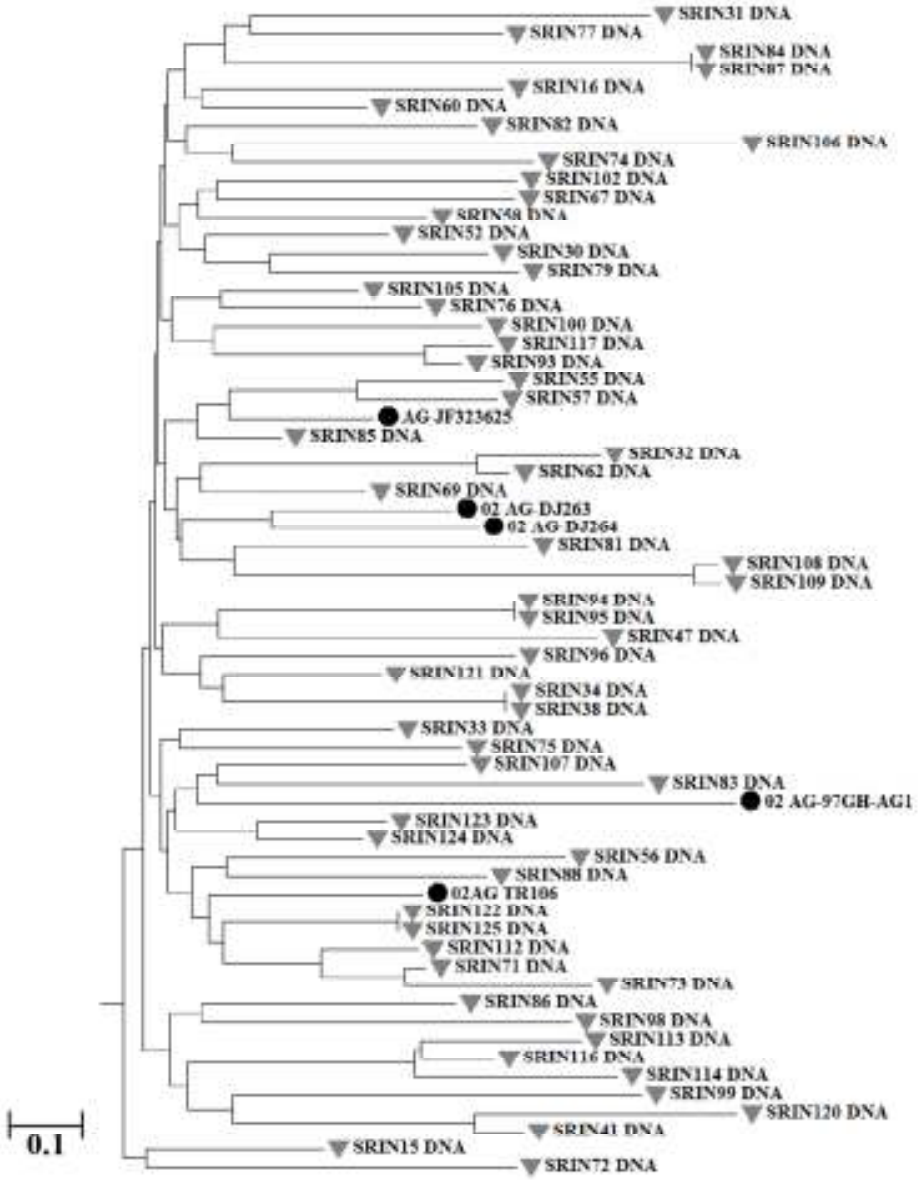



**Figure 12B: Enlarged phylogenetic tree from 1A showing CRF02_AG subtypes with DNA as starting template.**

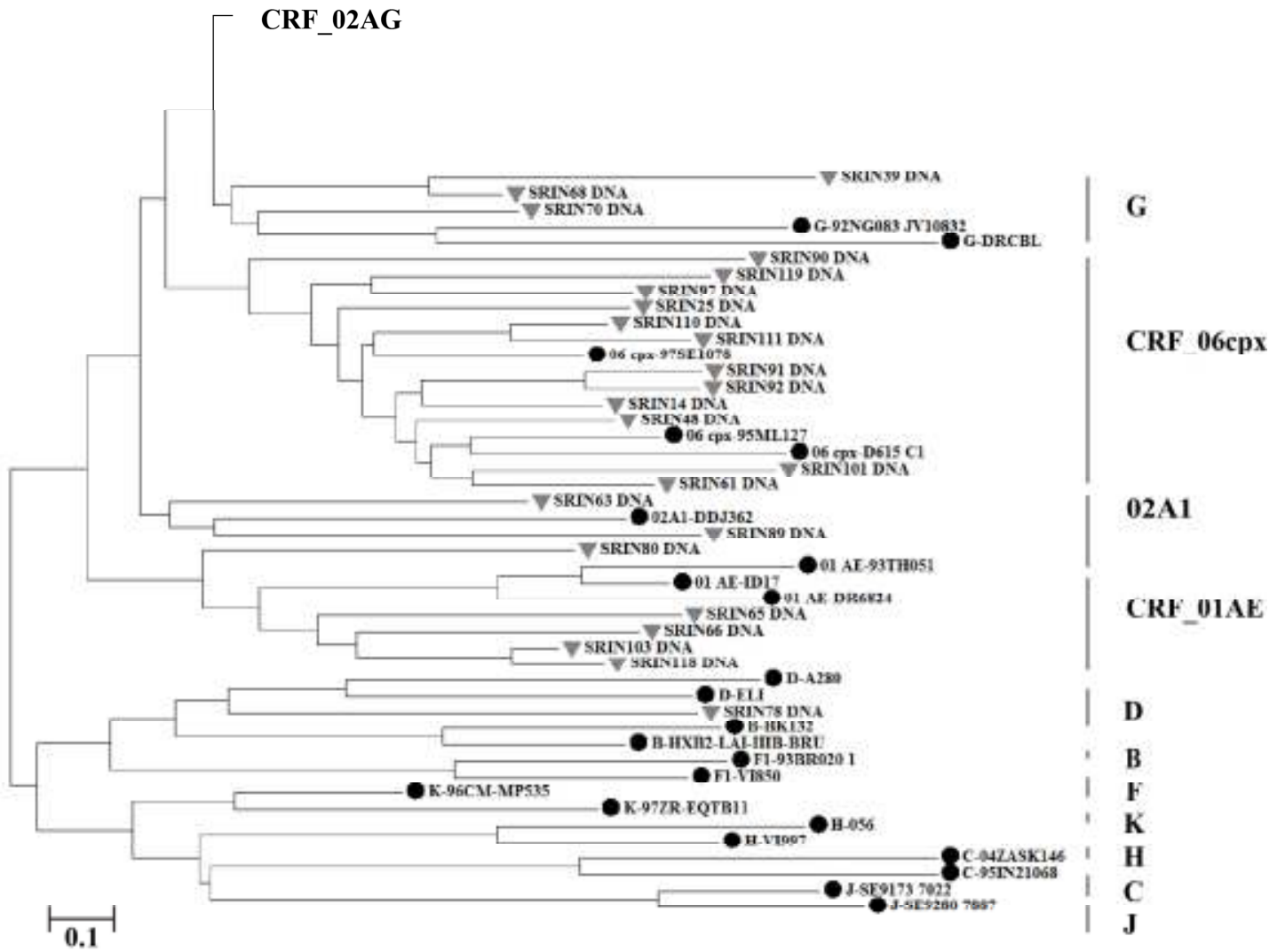



**Figure 12C: Enlarged phylogenetic tree from 1A showing all other subtypes apart from CRF02_AG, with DNA as starting template.**

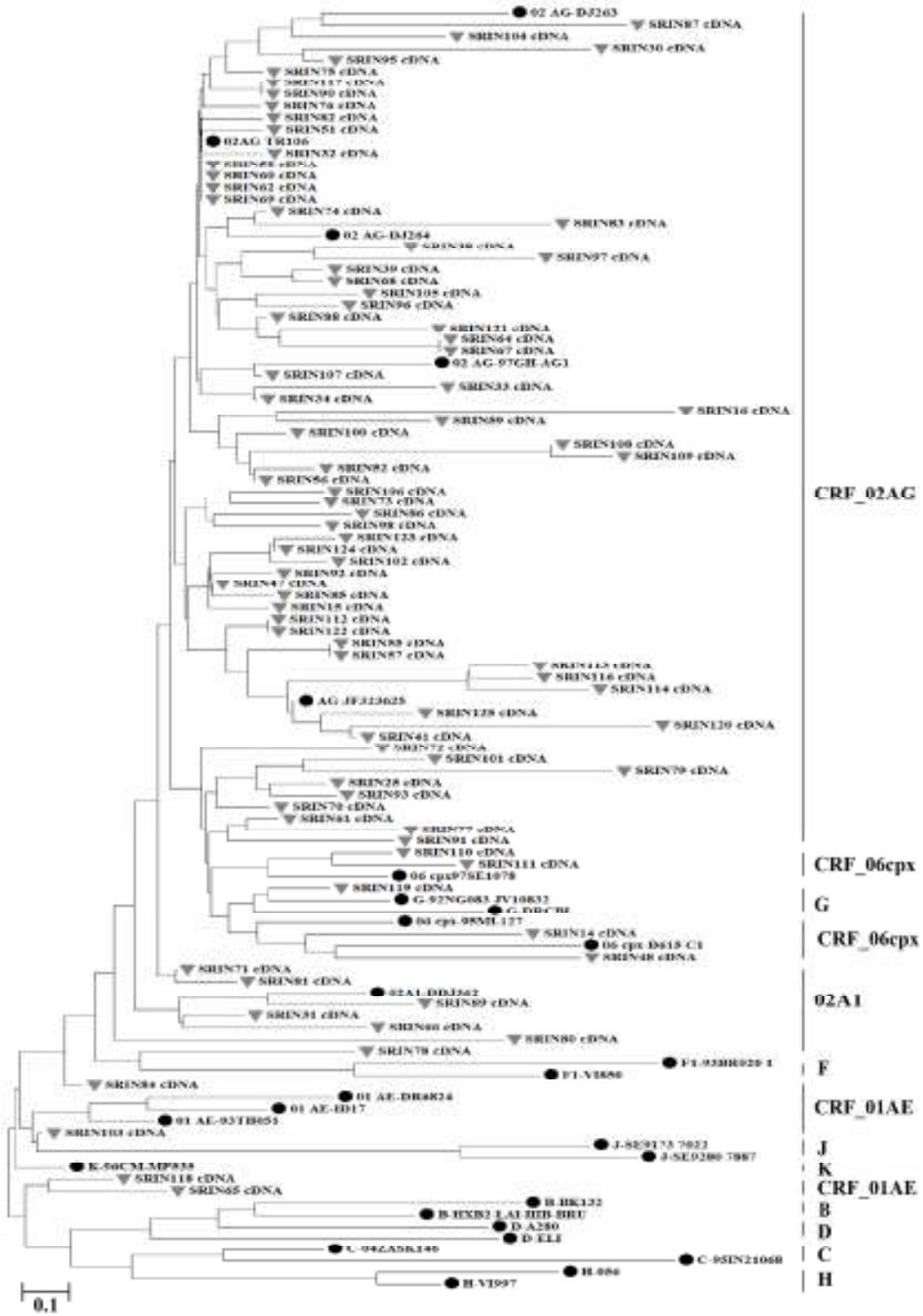



**Figure 12D: Phylogenetic tree from RNA as starting template. See details on pages 67-68.**

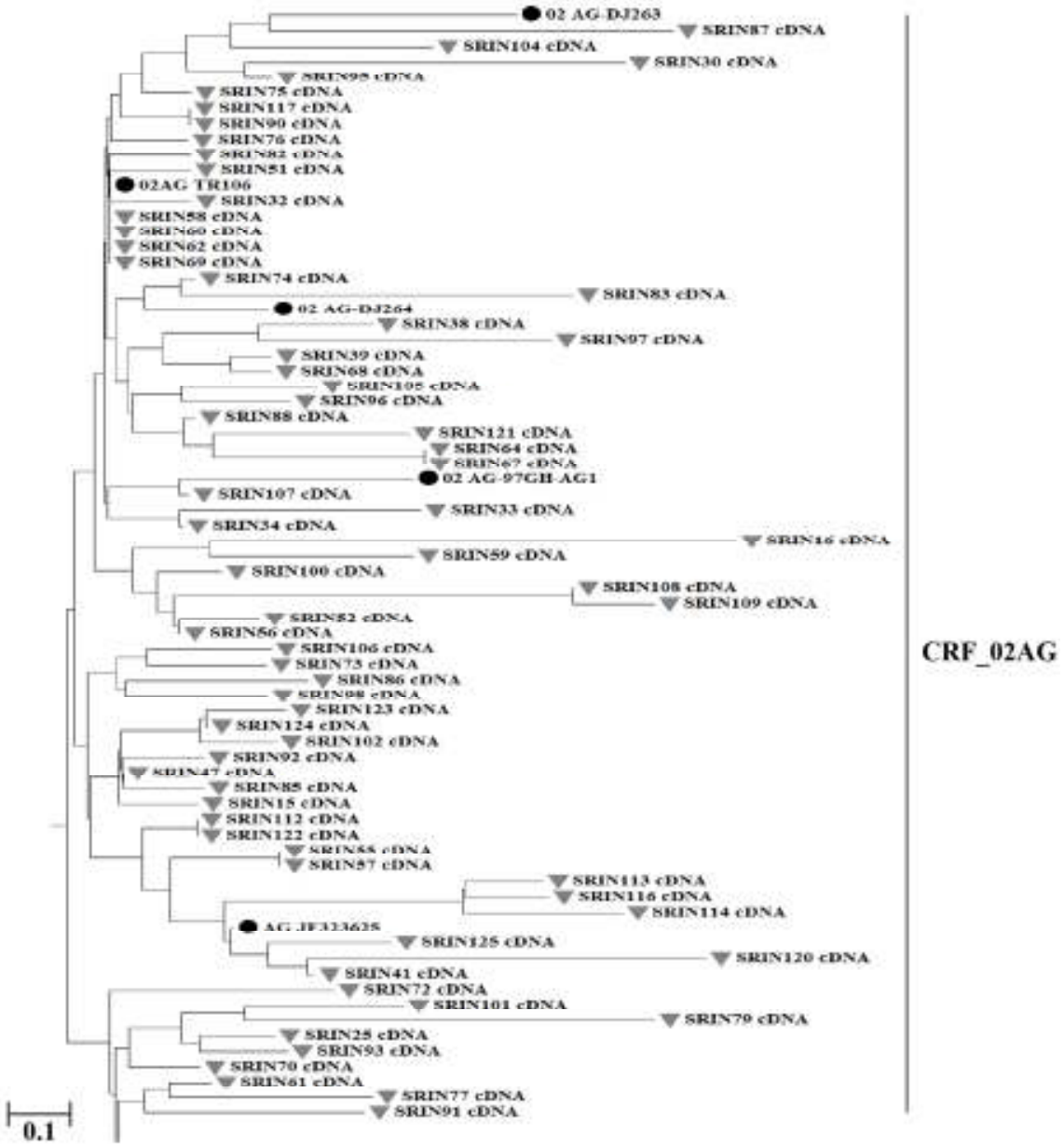



**Figure 12E: Enlarged phylogenetic tree from 1D showing CRF02_AG subtypes with RNA as starting template.**

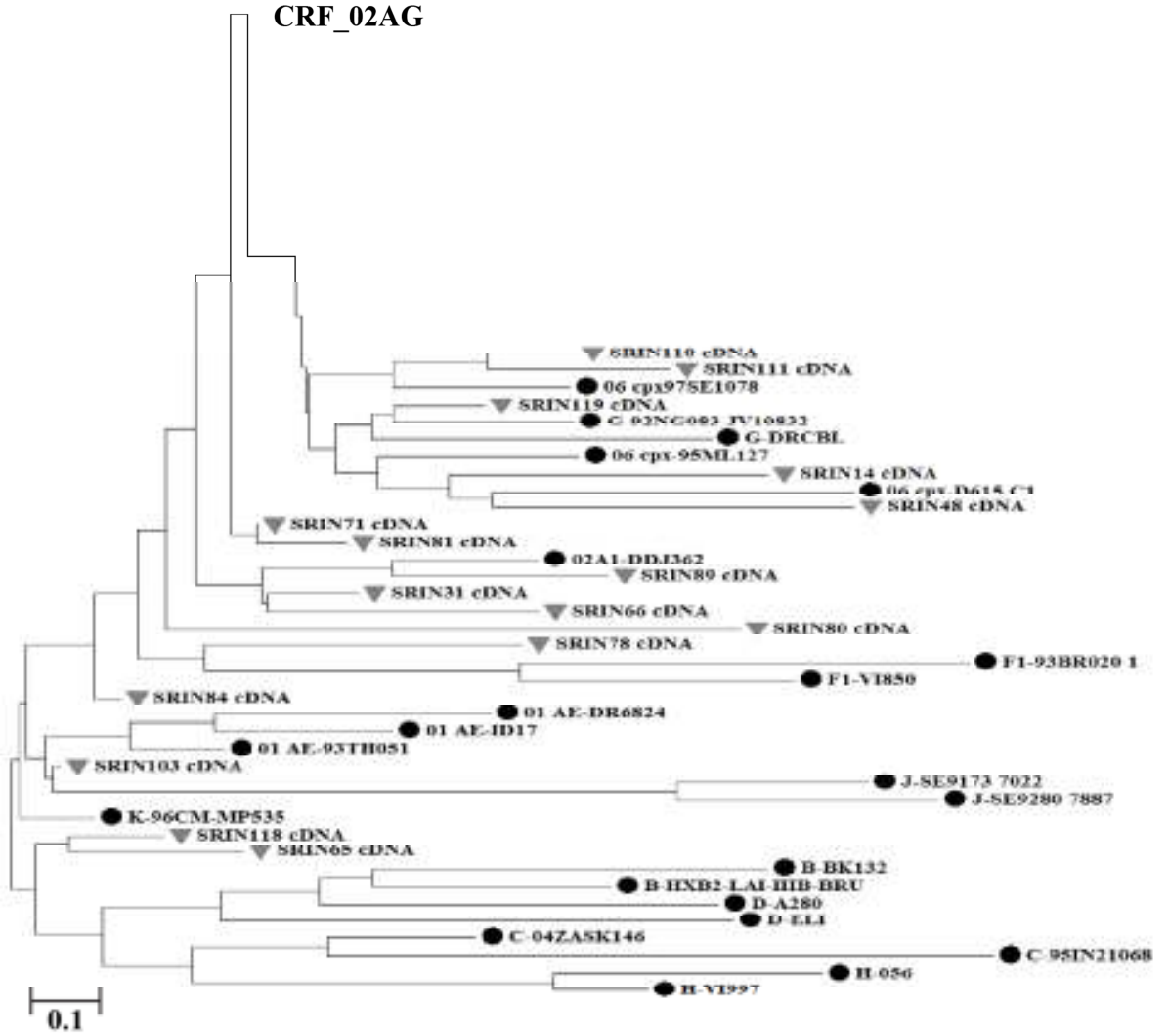



**Figure 12F: Enlarged phylogenetic tree from 1D showing all other subtypes apart from CRF02_AG, with RNA as starting template.**

**Figure 122: Neighbor-joining phylogenetic tree comprising of the PR/RT region of HIV-1 sequences from drug naïve patients residing in Nouna Burkina Faso.**

**The trees were generated from eighty three viral RNA and DNA template pairs isolated from the same patient. (A) Phylogenetic trees from DNA as starting template. (B) Enlarged phylogenetic tree from 1A showing CRF02_AG subtypes. (C) Enlarged phylogenetic tree from 1A showing all other subtypes apart from CRF02_AG.**

**(D) Phylogenetic tree from RNA as starting template. (E) Enlarged phylogenetic tree from 1D showing CRF02_AG subtypes. (F) Enlarged phylogenetic tree from 1D showing all other subtypes apart from CRF02_AG. Representative reference sequences of HIV-1 subtypes (A–K) and CRF02_AG, CRF01_AE and CRF06_cpx indicated as dark shaded circles were included from the National Center for Biotechnology Information (NCBI) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) via a megablast for highly similar HIV-1 sequences to the query sequence. Inverted triangles representing patient HIV-1 query sequences clustered to various references, which were used to determine the respective subtypes of the query sequences. The divergence existing between any two sequences was determined by the sum of the horizontal branch length, using the scale at the lower left. The straight lines on the right end of the trees indicate the subtype/CRF clustering. CRF02 AG was the predominant virus subtype in both DNA and cDNA trees.**

## 3.3 Genotyping of viral sequences from drug naïve and follow-up patients

### 3.3.1 Transmitted drug resistance mutations among drug naïve patients and further comparisons of viral RNA and DNA templates from the same patient

Genomic RNA and DNA were extracted from plasma and buffy coat samples of 86 HIV-1 infected drug naïve Burkinabés and used in drug resistance genotyping for the detection of TDRMs in the PR/RT region of HIV-1. Among the drug naïve patients, sequence alignment, editing and DRM analysis for mutations conferring resistance to HIV-1 was done for 83 paired patient derived HIV-1 RNA/DNA sequences out of a total number of 86 sequence pairs. The remaining three paired patient sequences were eliminated from the analysis because the sequence chromatographs could not be read.

Seventy-five patients harboured the drug susceptible strain of the virus and these had a Wilson's score mean of 0.91 (91%) and a Wilson's score confidence interval (CI) range of (0.81-0.95) at 95% CI. The remaining eight patients harboured TDRM that are known to affect drug susceptibility (as shown in figure 13B) with a Wilson's score mean of 0.09 (9%) and a CI range of 0.05- 0.17 at 95% CI. The drug resistance categories (see table 9 and appendix VII) were determined from the overall drug resistance score for a particular class of ARVs as given by the Stanford HIV drug resistance database (Rhee et al. 2003). Patients with no TDRM affecting drug susceptibility for a particular class of ARVs were 75 (90.4%) and those that had TDRM affecting drug susceptibility for a particular class of ARVs were 8 (9.6%) (figure 13B). Other minor and accessory resistance mutations were also seen but these had no effect on overall drug susceptibility and therefore nothing was recorded for the drug resistance mutations per resistance score category as shown again in table 9 and appendix VII. Among patients harbouring DRM that affected drug susceptibility, 7 (87.5%) had mutations affecting the susceptibility to reverse transcriptase inhibitors (RTIs) while 1 (12.5%) had a mutation affecting the susceptibility to protease inhibitors (PIs) as indicated in figure 13C. Consequently, among those harbouring RTI mutations, 6(86%) had DRMs to NNRTIs and only 1 (14%) had mutations affecting NRTI

susceptibility (see figure 13D). The general DRM profile per resistance score as seen in figure 13E, showed mainly low-level resistance to NNRTIs 4 (50%) followed by intermediate/high level resistance to NNRTIs 2 (25%), low-level resistance to NRTIs 1 (12.5%) and low-level resistance to PIs 1 (12.5%).

The major DRMs among drug naïve patients were mainly mutations to NNRTIs and NRTIs. Table 9 and appendix VII show the details of the DRM found among patient derived HIV-1 sequences. Table 9 gives the summary of mainly the mutations that are known to affect drug susceptibility and appendix VII gives details of DRM found within all 83 individual drug naïve patients. PI associated mutations and polymorphisms were L10V/I/F/L, K20I/R/V, L89M/I, V11I, K43T, A71T, T74S, K33F, and I93M. The K20I polymorphism was found in 81(98%) of all 83 patients and the K20R mutation was found in the remaining 2 (2%). Major NNRTI mutations included K103N and Y181C, these were found in 4 (5%) out of 83 patients. Minor NNRTI mutations, V108I, V179I, V90I, E138A, V106I and K238K/R were found in 20 (24%) out of all the 83 patients. A major NRTI mutation L210W was found in 1 (1%) of all 83 patients. Minor NRTI mutations found in 4 (5%) out of the 83 patients were mainly T69N, E44D, V118I/V and K219N mutations.

Furthermore, similarities between viral RNA and DNA sequences in PR/RT regions conferring resistance to HIV-1 were also assessed among drug naïve patients. A high similarity 78(94%) existed between RNA and DNA templates of the same patient among drug naïve patients and differences between templates were few 5 (6%) as shown in figure 13A. Appendix VII gives a detailed profile on how the similarities and differences in figure 13A were achieved. In table 9 and appendix VII, all yellow coloured regions in the tables indicate same mutations found in RNA and DNA templates of the respective patients and orange coloured regions indicate patients whose RNA and DNA templates differed by the presence of one or more mutations in RNA or DNA or vice versa. Within the orange coloured regions, the light blue segments highlight the differences in mutations. An example is the patient with lab ID SRIN48 in table 9. This patient had the K219N mutation in DNA and not in RNA.

**Table 9: Summary of drug resistance profile of HIV-1 infected drug naïve patients from Nouna Burkina Faso**

| Lab ID | Viral RNA load copies/ml | CD4$^{+}$T -cell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN31 | 172739 | 271 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN32 | 430515 | 177 | RNA | CRF02_AG - Like | *K20I* | | *V118IV* | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | *V118I* | | |
| SRIN48 | 2416935 | 147 | RNA | CRF06_cpx - Like | *K43T* *K20I* | **Y181C** | | | EFV,ETR,NVP,RPV |
| | | | DNA | CRF06_cpx - Like | *K43T* *K20I* | **Y181C** | *K219N* | | EFV,ETR,NVP,RPV |
| SRIN63 | 86731 | 316 | RNA | 02A1-Like | *K20I* | *V108IV* *V90I* | | NVP | |
| | | | DNA | 02A1-Like | *K20I* | *V108IV* *V90I* | | NVP | |
| SRIN66 | 227199 | 140 | RNA | CRF01_AE- like | *L10V* | *E138A* *V179I* | | RPV | |
| | | | DNA | CRF01_AE- like | *L10V* | *E138A* *V179I* | | RPV | |

**DR** represents drug resistance; **LR** represents level resistance Type of template: RNA; **Region analyzed**: PR 1-99 codons and RT (2/3) 1-335 codons. Primary resistance mutations are **bolded,** primary resistance mutations in other expert list are **bolded and underlined** and minor and accessory resistance mutations are ***italicized***.

**Table 9 continued.**

| Lab ID | Viral RNA load copies/ml | CD4⁺T -cell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN69 | 515487 | 428 | RNA | CRF02_AG - Like | *L10LV*<br>*K20I* | no mutation | | | |
| | | | | | | | L210LW | AZT, D4T | |
| | | | DNA | CRF02_AG - Like | *L10V*<br>*K20I* | no mutation | | | |
| | | | | | | | L210LW | AZT, D4T | |
| SRIN81 | 590908 | 323 | RNA | CRF02_AG - Like | *K20I, T74S* | *K238KR* | | NFV | |
| | | | DNA | CRF02_AG - Like | *K20I, T74S* | *K238R* | | NFV | |
| SRIN106 | 3589550 | 176 | RNA | CRF02_AG - Like | *K20I* | *V108I*<br>*V90I* | no mutation | NVP | |
| | | | DNA | CRF02_AG - Like | *K20I* | *V108I*<br>*V90I* | no mutation | NVP | |
| SRIN120 | 1400866 | 504 | RNA | CRF02_AG - Like | *K20I*<br>*V11I*<br>*L10I* | **K103N** | no mutation | | NVP, EFV |
| | | | DNA | CRF02_AG - Like | *K20I*<br>*V11I*<br>*L10I* | **K103KN** | no mutation | | NVP, EFV |
| SRIN123 | 976517 | 301 | RNA | CRF02_AG - Like | *K20I* | E138A | no mutation | RVP | |
| | | | DNA | CRF02_AG - Like | *K20I* | E138A | no mutation | RVP | |
| SRIN124 | 225569 | 467 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN125 | 1126782 | 33 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |

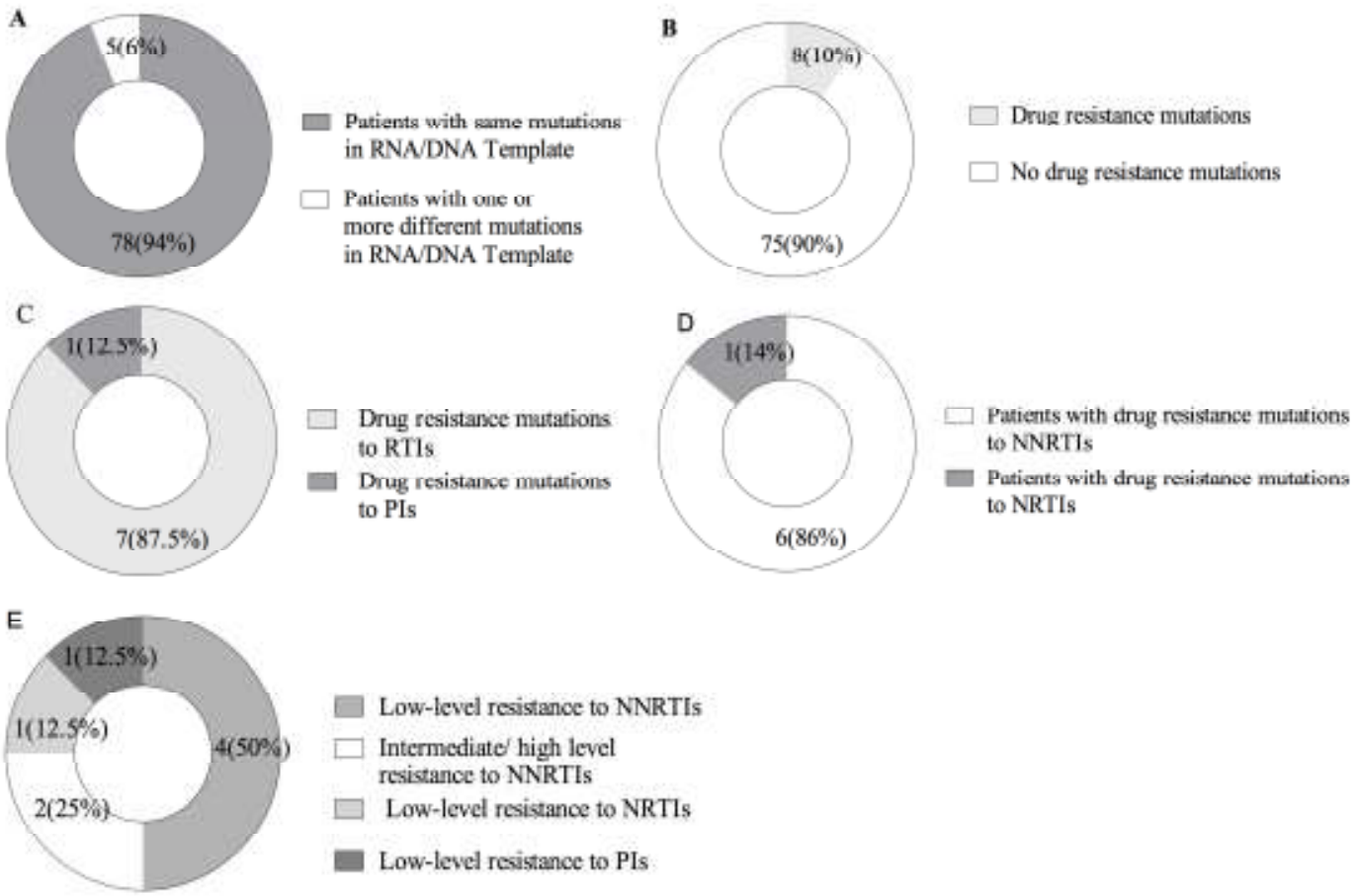



**Figure 13: Summary of HIV-1 genotyping profile of the PR/RT region of the HIV-1 pol gene of drug naive patients from Nouna Burkina Faso**

**(A) A pie chart demonstrating similarities and/or differences between RNA/DNA templates in regions conferring resistance to RTIs and PIs. (B) Comparison of patients with or without TDRM in PR/RT. (C) Among patients harbouring DRM that affect drug susceptibility, as indicated in B, 7 (87.5%) had mutations affecting the susceptibility of RTIs while 1 (12.5%) had mutations affecting the susceptibility of PIs. (D) Profile of RTI mutations only among patients with TDRM. High prevalence of mutations to NNRTI mutations n= 6 (86%) and low prevalence of NRTI mutations n=1 (14%). (E) DRM categories per resistance scores in patient derived PR/RT viruses.**

## 3.4 Nucleotide and amino acid changes within HIV-1 RNA and DNA paired sequences of the same patient

Among the drug naïve patients, paired sequences from the PR/RT of the HIV-1 pol region, obtained from RNA and DNA starting templates of the same patients were aligned to each other to determine any nucleotide differences present within these paired sequences. The sequences were made up of the entire HIV-1 protease (1-99 amino acids) and 2/3 of RT. The numbers of positions having nucleotide differences were noted and subsequent codons from both templates specifying an amino acid were written.

Nucleotide differences were synonymous when they did not change the resultant amino acid. For instance, a base substitution occurring at position 942 in the paired nucleotide sequence of patient SRIN15 from C (in RNA) to A (in DNA) changed the codon ACC (in RNA) to ACA (in DNA). The resulting amino acid in both cases was lysine, which is denoted by K, and therefore the change was synonymous. In instances where the amino acid changed from for example lysine (ACC) in RNA, to threonine (ACG) in DNA, the change was described as non-synonymous because the nucleotide difference resulted in changing the amino acid. A change was described as either synonymous or non-synonymous when a degenerate code occurred at a position that allowed possibilities of both a synonymous and a non-synonymous change to occur at the same position. Every ambiguity nucleotide (also termed the degenerate code) either in the viral DNA or RNA or both, were given a value of one, the total number of ambiguities was determined by adding up the number of ambiguities in a given sequence and the resultant value was divided by two and added to already obtained values under the sections synonymous and non-synonymous. Table 10 summarizes the results of the nucleotide changes found among all drug naive paired RNA/DNA templates. Templates with more than 10 nucleotide changes were deep sequenced as a way of confirming the results obtained. Paired templates having no nucleotide differences were not included.

Forty one percent of the patients had nucleotide differences in paired viral templates, while 59% had no nucleotide differences and there was no significant difference ($p=0.60$) between the number of synonymous changes ($n=170$) and the number of non-synonymous

nucleotide changes (n=198) (table 11). Among patients who had nucleotide differences, 38% of them had at least ten nucleotide differences within paired viral templates.

**Table 10: Summary of RNA/DNA paired nucleotide sequence alignments from drug naïve patients**

| Patient Lab ID | No of postions with nucleotide changes | Chracteristics of change based on amino acid | |
|---|---|---|---|
| | | Synonymous | Non- synonymous |
| Srin 14 day 0 | 2 | 2 | 0 |
| Srin 16 day 0 | 23 | 7 | 16 |
| Srin 31 day 0 | 26 | 18 | 8 |
| Srin 38 day 0 | 53 | 29 | 23 |
| Srin 39 day 0 | 23 | 5 | 17 |
| Srin 47 day 0 | 5 | 3 | 2 |
| Srin 48 day 0 | 2 | 1 | 1 |
| Srin 52 day 0 | 1 | 1 | 0 |
| Srin 55 day 0 | 3 | 3 | 0 |
| Srin 56 day 0 | 6 | 3 | 3 |
| Srin 60 day 0 | 8 | 4 | 4 |
| Srin 63 day 0 | 3 | 3 | 0 |
| Srin 66 day 0 | 5 | 5 | 0 |
| Srin 69 day 0 | 7 | 3 | 4 |
| Srin 70 day 0 | 1 | 1 | 0 |
| Srin 71 day 0 | 3 | 2 | 1 |
| Srin 72 day 0 | 2 | 1 | 1 |
| Srin 74 day 0 | 1 | 1 | 0 |
| Srin 75 day 0 | 2 | 1 | 0 |
| Srin 77 day 0 | 1 | 1 | 0 |
| Srin 80 day 0 | 7 | 1 | 6 |
| Srin 81 day 0 | 2 | 0 | 2 |
| Srin 82 day 0 | 2 | 2 | 0 |
| Srin 83 day 0 | 1 | 1 | 0 |
| Srin 84 day 0 | 2 | 0 | 0 |
| Srin 85 day 0 | 2 | 1 | 1 |
| Srin 86 day 0 | 14 | 5 | 9 |
| Srin 87 day 0 | 2 | 0 | 2 |
| SRIN93 day 0 | 10 | 1 | 9 |
| SRIN94 day 0 | 10 | 2 | 8 |
| SRIN95 day 0 | 12 | 2 | 10 |

Patient sequences with nucleotide changes of ≥10 are highlighted in grey.

**Table 11: Summary of nucleotide substitution analysis for all drug naïve patient derived viral sequences via Sanger sequencing**

| Paired templates | Number of paired templates | Nucleotide substitutions | | | | Mann–Whitney U test |
|---|---|---|---|---|---|---|
| | | Type | Number | Median (IQR) | Maximum | |
| Number of patients with changes in RNA and DNA sequence pairs | n=34 | Synonymous | 170 | 2(1-5) | 30 | p= 0.60 |
| | | Non-synonymous | 198 | 2(0-8.5) | 30 | |
| Number of patients with no changes in RNA and DNA | n=49 | | | | | |

### 3.4.1 Patterns of HIV-1 drug resistance mutations among patients followed over time

Follow-up plasma samples from all patients 24 weeks post HAART with high plasma viral load > 1000 copies per ml of plasma were genotyped and sequenced for HIV-1 DRM analysis. Patients were put on a first line regime of 2 NRTI (3TC, d4T or AZT, d4T) and 1NNRTI (NVP or EFV). A total of eight patients were followed-up. Six out of the eight patients had DRMs affecting drug susceptibility for a particular class of ARVs as described earlier for drug naïve patients in section 3.3.1. The remaining two patients had no resistance mutations affecting drug susceptibility. These six patients harboured DRMs that affected the susceptibility to RTIs. Furthermore, the six patients had mainly major DRMs to NNRTIs and three out of these six patients again had additional major DRMs to NRTIs (table 12).

As indicated in table 12 the patient SRIN31, had a major DRM K103N developing to NNRTI within 48 weeks of therapy, causing high-level resistance to NVP and EFV. This mutation was still present at 72 weeks post ART. Patient SRIN48 on the other hand, harboured the major TDRM to NNRTIs, the Y181C mutation and the NRTI accessory TAM, K219N, at the drug naïve stage. The accessory TAM disappeared after 2 weeks of therapy but the Y181C mutation persisted. At 12 weeks into therapy, a major NRTI mutation M184V emerged in addition to the Y181C mutation and at 36 weeks into therapy, multiple mutations to NNRTIs (V108I and H221HY) emerged in addition to the already existing ones. All these mutations disappeared at 64 weeks into therapy leaving the accessory resistance mutation V108I which causes low-level resistance to NNRTIs. Furthermore, patient SRIN72 had the V179I mutation, which was no longer present at 48 weeks into therapy. Consequently, patient SRIN98 had a

major DRM to NNRTIs, thus the K103N that causes high-level resistance to NVP and EFV. This mutation developed at 24 weeks into therapy. At 36 weeks into therapy an additional DRM to NRTI, the M184V developed with the K103N mutation. The former disappeared and the NNRTI mutation E138Q that is known to cause high-level resistance to NVP and EFV and low level resistance to RPV appeared in addition to the K103N mutation at 48 weeks into therapy. Patient SRIN103 developed multiple major DRMs G190A, and M230L to NNRTI and M184V to NRTIs within 24 weeks into therapy. These cause high-level resistance to 3TC, FTC, EFV, NVP and RPV and intermediate level resistance to ETR and also low-level resistance to ABC. At 36 weeks into therapy the M230L mutation disappeared leaving the G190A and the M184V mutations. Subsequently, patient SRIN106 developed the major NNRTI DRM K103N in addition to the NNRTI polymorphism V90I at 24 weeks of therapy and this mutation persisted at 48 weeks into therapy in addition to the reappearance of the V108I mutation. The V108I and the V90I polymorphism persisted from the drug naïve stage until 12 weeks into therapy when the V108I mutation disappeared. Patient SRIN117 had no DRM that affected drug susceptibility. Finally patient SRIN120 had the major NNRTI mutation K103N at the drug naïve stage and within 24 weeks of therapy, this patient had the NRTI mutation M184V emerging in addition to the NNRTI mutation K103N. Both major DRMs K103N and M184V were still present at 48 weeks of therapy together with minor PI mutations and polymorphisms that were present at the drug naïve stage. The specific scenarios above indicate that major DRMs developed during the course of HIV infection usually persist a while and may either disappear or lead to multi drug DRMs. Also in some instances high viral loads seen in patients undergoing HAART may be due to other reasons (for example non- compliance) either than resistance mutations as seen in patients SRIN72 and SRIN117.

**Table 12: Drug resistance profile in plasma samples of HIV-1 infected patients over time**

| Lab ID | Viral RNA load copies/ml | $CD4^+$T -cell counts | Subtype PR-RT | Weeks of follow-up | PI Mutations | NNRTI Mutation | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | Intermediate-LR | High level-LR |
| SRIN31 | 98000 | 271 | CRF02_AG - Like | 0 | *K20I* | no mutation | | | | |
| | 2000 | ND | CRF02_AG - Like | 2 | *K20I* | no mutation | | | | |
| | ND | 564 | CRF02_AG - Like | 4 | *K20I* | no mutation | | | | |
| | < 40 | 450 | CRF02_AG - Like | 12 | *K20I* | no mutation | | | | |
| | 39000 | 462 | CRF02_AG - Like | 48 | *K20I* *L10V* | **K103N** | no mutation | | | NVP, EFV |
| | 2600 | ND | CRF02_AG - Like | 72 | | **K103N** | no mutation | | | NVP, EFV |
| SRIN48 | 2400000 | 147 | CRF06_cpx - Like | 0 | *K43T* *K20I* | **Y181C** | *K219N* | | EFV,ETR | NVP,RPV |
| | 9200 | ND | CRF06_cpx - Like | 2 | *K20I* *K43T* | **Y181C** | | | EFV,ETR,RPV | NVP |
| | 1900 | 498 | CRF06_cpx - Like | 12 | *K20I* *K43T* | **Y181C** | **M184V** | ABC | EFV,ETR,RPV | NVP,3TC,FTC |
| | 22000 | 344 | CRF06_cpx - Like | 36 | *K20I* *K43T* | **V108I, Y181C, H221HY** | **M184V** | ABC | EFV,ETR,RPV | NVP,3TC,FTC |
| | 6400 | 310 | CRF06_cpx - Like | 64 | | *V108I* | | NVP | | |
| SRIN72 | 97000 | 333 | G - Like | 0 | *K20I* | *V179I* | | | | |
| | 8800 | 436 | G - Like | 48 | *K20I* | | | | | |
| SRIN98 | 10000000 | 63 | CRF02_AG - Like | 0 | *K20I* *L10V* | *no mutation* | | | | |
| | 1900 | 498 | CRF02_AG - Like | 12 | *K20I* *L10V* | no mutation | | | | |
| | 40000 | 217 | CRF02_AG - Like | 24 | | **K103N** | | | | NVP, EFV |
| | | | CRF02_AG - Like | 36 | *K20I* *L10V* | **K103N** | **M184V** | ABC | | 3TC,FTC,EFV,NVP |
| | 650000 | 85 | CRF02_AG - Like | 48 | | **K103N E138Q** | | RPV | | NVP, EFV |
| SRIN103 | 183000 | 397 | CRF01_AE- like | 0 | *L10V* | *V179I* | no mutation | | | |
| | 8000 | ND | CRF01_AE- like | 2 | *L10V* | *V179I* | | | | |
| | 900 | 682 | CRF01_AE- like | 4 | *L10V* | *V179I* | | | | |
| | 15000 | 608 | CRF01_AE- like | 12 | *L10V* K20R | | | | | |
| | 150000 | 701 | CRF01_AE- like | 24 | *L10V* | **G190A M230L** *V179I* | **M184V** | ABC | ETR | 3TC,FTC,EFV,NVP,RPV |
| | 220000 | 681 | CRF01_AE- like | 36 | *L10V* | **G190A** *V179I* | **M184V** | ABC, ETR, RPV | EFV | 3TC,FTC,NVP |

**DR** represents drug resistance; **LR** represents level resistance Type of template: RNA; **Region analyzed**: PR 1-99 codons and RT (2/3) 1-335 codons. Primary resistance mutations are **bolded,** primary resistance mutations in other expert list are **bolded and underlined** and minor and accessory resistance mutations ***are italicized***. $CD4^+$ T-cell counts were measured in cells/ul. **ND** means not determined.

**Table 12 continued.**

| Lab ID | Viral RNA load copies/ml | $CD4^+$T -cell counts | Subtype PR-RT | Weeks of follow-up | PI Mutations | NNRTI Mutation | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | Intermediate-LR | High level-LR |
| SRIN106 | 3600000 | 176 | CRF02_AG - Like | 0 | *K20I* | *V108I* *V90I* | no mutation | NVP | | |
| | 35000 | ND | CRF02_AG - Like | 2 | *K20I* | *V108I* *V90I* | no mutation | NVP | | |
| | ND | 353 | CRF02_AG - Like | 4 | *K20I* | *V108I* *V90I* | no mutation | NVP | | |
| | 19000 | 386 | CRF02_AG - Like | 12 | *K20I* | *V108I* *V90I* | no mutation | NVP | | |
| | 59000 | 309 | CRF02_AG - Like | 24 | *K20I* | *V90I* **K103N** | no mutation | | | NVP, EFV |
| | 163000 | ND | CRF02_AG - Like | 48 | | *V108I* *V90I* **K103N** | no mutation | | | NVP, EFV |
| SRIN117 | 90000 | 341 | CRF02_AG - Like | 0 | *K20I* | no mutation | | | | |
| | 8000 | 409 | CRF02_AG - Like | 24 | | no mutation | | | | |
| SRIN120 | 1400000 | 504 | CRF02_AG - Like | 0 | *K20I* *V11I* *L10I* | **K103N** | no mutation | | | NVP, EFV |
| | 5000 | 589 | CRF02_AG - Like | 4 | *K20I* *V11I* *L10I* | | no mutation | | | |
| | 13000 | 704 | CRF02_AG - Like | 24 | *K20I* *V11I* | **K103N** | **M184V** | ABC | | 3TC,FTC,EFV |
| | 18000 | ND | CRF02_AG - Like | 36 | *K20I* *V11I* | **K103N** | **M184V** | ABC | | 3TC,FTC,EFV |
| | ND | ND | CRF02_AG - Like | 48 | *K20I* *V11I* | **K103N** | **M184V** | ABC | | 3TC,FTC,EFV,NVP |

**DR** represents drug resistance; **LR** represents level resistance Type of template: RNA; **Region analyzed**: PR 1-99 codons and RT (2/3) 1-335 codons. Primary resistance mutations are **bolded,** primary resistance mutations in other expert list are **bolded and underlined** and minor and accessory resistance mutations ***are italicized***. $CD4^+$ T-cell counts were measured in cells/ul. **ND** means not determined.

## 3.5 Deep sequencing of thirteen drug naïve patient templates

Thirteen patients derived HIV-1 RNA and DNA templates from HIV-1 infected drug naïve patients were deep sequenced via a process called Nextera tagmentation using PCR amplicons. These patients were selected based on the fact that after Sanger sequence analysis, they either had differences in DRMs between individual RNA and DNA pairs or had ten or more differences in RNA and DNA nucleotide pairs that needed to be confirmed by a more robust method.

### 3.5.1 Transmitted drug resistance mutations profile of all thirteen drug naïve patient paired RNA and DNA templates from deep sequencing compared to Sanger sequencing.

In deep sequencing, mutational frequencies were given for TDRMs from 1 to 100% frequency cut-offs (Table 13). Minority variants were classified as TDRMs detected at a frequency of 1% to less than 20% of the viral quasispecies (Colson et al., 2014). The minority variants represented TDRMs, which could not be detected via Sanger sequencing but were detected in deep sequencing. Sanger sequencing only detected mutations occurring in deep sequencing, at a frequency of 20% or higher, while deep sequencing detected mutations at all frequencies from 1 to 100%. In deep sequencing, the coverage per nucleotide position for each patient sample was between 20,000-10,000 reads and the number of reads for each patient RNA or DNA template was approximately 200,000 reads unlike in Sanger sequencing where the templates were read only once as described by (Chabria et al., 2014). Of the thirteen paired patient templates, ten harboured TDRMs to PIs and RTIs at frequencies defined as minority variants whereas in the previous Sanger sequencing these mutations were not seen. The additional mutations detected in only deep sequencing as minority variants and not in Sanger sequencing were D30N, I47V and G73S to PIs and mutations to RTIs included K103N, T69D, M230I, M184I, E138K, K101E, V179I, V90I and F22FL (Table 13). Above the frequency of 20%, all DRMs detected by deep sequencing were also detected by Sanger sequencing.

Deep sequenced viral RNA and DNA templates were compared for concordance or discordance in terms of detecting DRMs within paired sequences by frequency, as shown in table 13. We defined concordance as the identification of a mutation in both paired viral RNA and DNA templates at a frequency of 20% and above (also the range for Sanger sequencing) or below 20% for the minority viral population in deep sequencing. Discordance on the other hand was defined as the inability of identifying a mutation at a frequency of 20% and above or below 20% in either viral RNA or DNA templates. The detection limit for all mutations considered was a frequency of 1-100%. What defined discordance was the ability of the mutation being picked up in one template and not in the other. Since Sanger sequencing only detects mutations at a frequency of 20% or higher, we decided to assess how DNA and RNA detection compared with each other above 20%. We observed a high level of concordance (95%) between DNA and RNA pairs for mutation frequencies 20% and above. Discordance observed in this group was only 5%. Also we assessed the same concept of discordance or concordance in paired RNA and DNA sequences in the minority viral population at a frequency below 20%, in contrast to mutation frequencies 20% and above, there was a high rate of discordance (74%) in paired RNA and DNA samples in the minority viral population with a percentage concordance of 26%.

**Table 13: Detection of drug resistance mutations from thirteen drug naïve patient RNA/DNA template pairs via frequency cut-offs in deep sequencing**

| HIV-1 target region | Type of DRM | Frequencies of DRM in patient template (%) | |
|---|---|---|---|
| | | **SRIN16-RNA** | **SRIN16-DNA** |
| PR | K20I | 100 | 100 |
| PR | L89M | 99 | 99 |
| RT | V189I | | 42 |
| | | **SRIN31-RNA** | **SRIN31-DNA** |
| PR | K20I | 100 | 100 |
| PR | D30N | | 1 |
| PR | M46I | 2 | 3 |
| PR | L89I | 3 | |
| PR | L10V | 37 | 33 |
| PR | L89M | 96 | 99 |
| RT | V189I | 5 | |
| RT | M230I | | 6 |
| | | **SRIN33-RNA** | **SRIN33-DNA** |
| PR | I47V | | 3 |
| PR | V11I | 4 | |
| PR | K20I | 100 | 100 |
| PR | L10V | 65 | 100 |
| PR | L89M | 99 | 99 |
| | | **SRIN38-RNA** | **SRIN38-DNA** |
| PR | L89M | 78 | 100 |
| PR | K20I | 100 | 100 |
| RT | M184I | | 1 |
| | | **SRIN39-RNA** | **SRIN39-DNA** |
| PR | M46I | | 11 |
| PR | V82I | 100 | 100 |
| PR | L89I | 2 | |
| PR | L89M | 97 | 99 |
| PR | K20I | 100 | 100 |
| RT | M230I | | 15 |
| | | **SRIN48-RNA** | **SRIN48-DNA** |
| PR | K43T | 99 | 100 |
| PR | K20I | 100 | 99 |
| PR | L89M | 99 | 100 |
| RT | K219N | 2 | 27 |
| RT | K103N | 1 | |
| RT | Y181C | 63 | 57 |
| | | **SRIN56-RNA** | **SRIN56-DNA** |
| PR | K20R | 2 | |
| PR | L89I | 34 | 17 |
| PR | L10V | 73 | 35 |
| PR | K20I | 99 | 97 |
| PR | L89M | 65 | 83 |
| RT | E138K | | 2 |
| RT | V179I | 9 | 9 |
| RT | K101E | | 4 |
| RT | M184I | | 4 |
| RT | M230I | | 8 |

Freqency cut-offs of DRMs from 1% to less than 20% of the viral quasispecies represent minority variants while frequency cut-offs from 20% to 100% represent majority variants, which are also detected by Sanger or bulk

sequencing. Empty cubicles represent DRMs in less than 1% frequency cut-off of the virus population. Here frequency values are not shown.

**Table 13 continued.**

| HIV-1 target region | Type of DRM | Frequencies of DRM in patient template (%) | |
|---|---|---|---|
| | | **SRIN58-RNA** | **SRIN58-DNA** |
| PR | I47V | | 2 |
| PR | L89I | 18 | 5 |
| PR | K20I | 100 | 100 |
| PR | V11I | 16 | 6 |
| PR | L10V | 74 | 74 |
| PR | L89M | 81 | 95 |
| RT | V90I | 10 | 6 |
| RT | F227L | 1 | |
| | | **SRIN78-RNA** | **SRIN78-DNA** |
| PR | V11I | | 1 |
| PR | K20I | 99 | 100 |
| PR | L89M | 27 | |
| RT | D67N | | 1 |
| RT | M230I | | 1 |
| | | **SRIN86-RNA** | **SRIN86-DNA** |
| PR | L10V | 2 | 2 |
| PR | L10I | 97 | 98 |
| PR | K20I | 99 | 99 |
| PR | L89M | 99 | 100 |
| | | **SRIN93-RNA** | **SRIN93-DNA** |
| PR | G73S | | 1 |
| PR | K20I | 100 | 100 |
| PR | L10V | 35 | 8 |
| PR | L89M | 99 | 99 |
| RT | M184I | | 2 |
| RT | M230I | | 2 |
| | | **SRIN94-RNA** | **SRIN94-DNA** |
| PR | K20R | 89 | 96 |
| PR | L89I | 18 | 15 |
| PR | L10V | 89 | 100 |
| PR | L10I | 11 | |
| PR | K20I | 32 | 12 |
| PR | L89M | 81 | 84 |
| RT | V90I | 88 | 100 |
| | | **SRIN95-RNA** | **SRIN95-DNA** |
| PR | V11I | | 1 |
| PR | K20I | 98 | 99 |
| PR | L10I | 98 | 99 |
| PR | L89M | 99 | 99 |
| RT | T69D | 1 | |
| RT | V90I | 26 | 6 |
| RT | E138K | | 2 |

### 3.5.2 Comparison of RNA and DNA templates of thirteen selected samples using bulk or Sanger and deep sequencing data at the 20 percent cut-off

Bulk sequencing detected differences in mutations among RNA and DNA templates of five patients, which were later, confirmed by deep sequencing. These patient templates were part of the thirteen paired patient samples that were deep sequenced. Deep sequencing results confirmed the differences in resistance mutations seen in bulk sequencing as indicated in table 14. One mutation, the K219N mutation, seen in DNA and not in RNA templates of patient derived HIV-1 sequences in Sanger sequencing was also confirmed in deep sequencing. In Sanger sequencing, two minor mutations and two polymorphisms were found in RNA that were not present in DNA. These were the PI mutation L89I and the RTI mutation V90I and the PI polymorphisms K20I and L10V respectively. The same pattern was also seen in deep sequencing thereby showing that at a frequency of 20%, mutation profiles of both deep and Sanger sequencing are the same.

**Table 14: Differences and similarities in mutations between RNA and DNA templates of thirteen selected drug naïve patients using Sanger and deep sequencing data at a 20 percent frequency cut-off.**

| | | Sanger Sequencing | | | Deep Sequencing 20% Variants | | |
|---|---|---|---|---|---|---|---|
| **Lab ID** | **Type of Template** | **PI Mutations** | **NNRTI Mutations** | **NRTI Mutations** | **PI Mutations** | **NNRTI Mutations** | **NRTI Mutations** |
| SRIN16 | RNA | *K20I* | no mutation | | *K20I* | no mutation | |
| | DNA | *K20I* | no mutation | | *K20I* | no mutation | |
| SRIN31 | RNA | *K20I* | no mutation | | *K20I* | | |
| | DNA | *K20I* | no mutation | | *K20I* | | |
| SRIN33 | RNA | *L10V* | no mutation | | *L10V* | no mutation | |
| | | *K20I* | | | *K20I* | | |
| | DNA | *L10V* | no mutation | | *L10V* | no mutation | |
| | | *K20I* | | | *K20I* | | |
| SRIN38 | RNA | *K20I* | no mutation | | *K20I* | no mutation | |
| | DNA | *K20I* | no mutation | | *K20I* | no mutation | |
| SRIN39 | RNA | *V82I* | no mutation | | *V82I* | no mutation | |
| | | *K20I* | | | *K20I* | | |
| | DNA | *V82I* | no mutation | | *V82I* | no mutation | |
| | | *K20I* | | | *K20I* | | |
| SRIN48 | RNA | *K43T* | | | *K43T* | | |
| | | *K20I* | | | *K20I* | | |
| | | | **Y181C** | | | **Y181C** | |
| | DNA | *K43T* | | | *K43T* | | |
| | | *K20I* | | | *K20I* | | |
| | | | **Y181C** | | | **Y181C** | |
| | | | | *K219N* | | | *K219N* |

Brown coloured sections with grey colours within represent the five patient templates that showed one or more different mutations in RNA or DNA templates in Sanger and deep sequencing. Green and yellows coloured sections represent sequences, which had the same mutations in RNA and DNA templates.

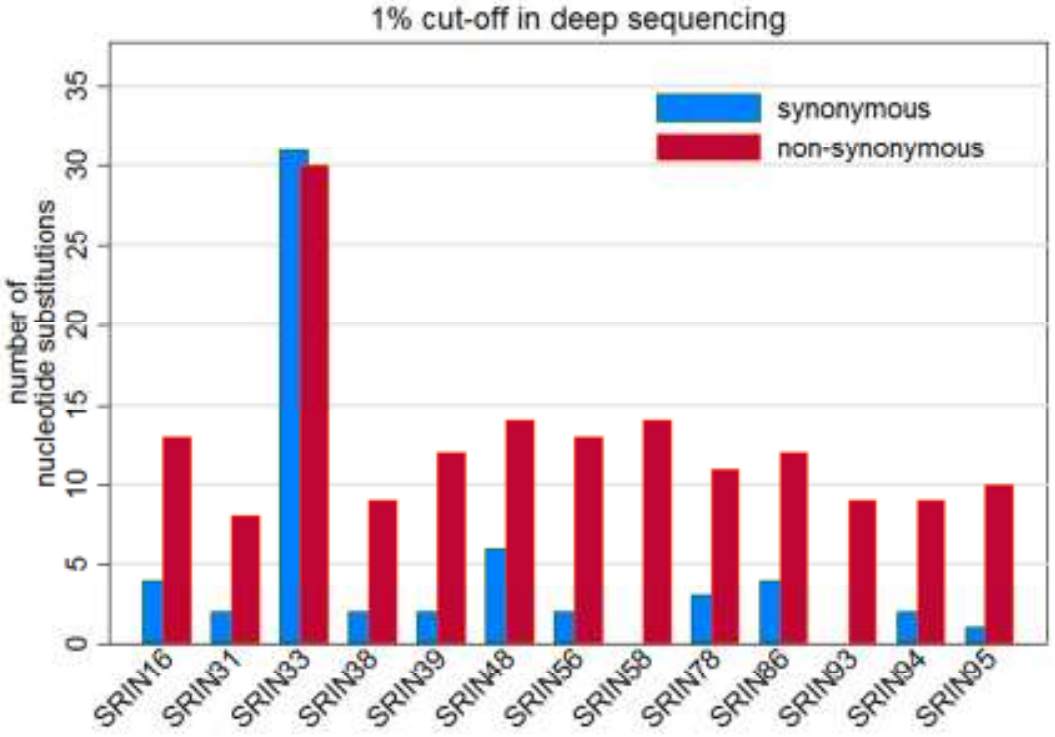
**Table 14 continued**

| Lab ID | Type of Template | PI Mutations | NNRTI Mutations | NRTI Mutations | PI Mutations | NNRTI Mutations | NRTI Mutations |
|---|---|---|---|---|---|---|---|
| SRIN56 | RNA | *K20I*<br>*L10V*<br>*L89I* | no mutation | | *K20I*<br>*L10V*<br>*L89I* | no mutation | |
| | DNA | *K20I*<br>*L10V*<br>no mutation | no mutation | | *K20I*<br>*L10V*<br>no mutation | no mutation | |
| SRIN58 | RNA | *K20I*<br>*L10V* | no mutation | | *K20I*<br>*L10V* | no mutation | |
| | DNA | *K20I*<br>*L10V* | no mutation | | *K20I*<br>*L10V* | no mutation | |
| SRIN78 | RNA | no mutation | | | no mutation | | |
| | DNA | no mutation | | | no mutation | | |
| SRIN86 | RNA | *L10I*<br>*K20I* | no mutation | | *L10I*<br>*K20I* | no mutation | |
| | DNA | *L10I*<br>*K20I* | no mutation | | *L10I*<br>*K20I* | no mutation | |
| SRIN93 | RNA | *K20I*<br>*L10V* | *V90I* | | *K20I*<br>*L10V* | *V90I* | |
| | DNA | *K20I*<br>no mutation | *V90I* | | *K20I*<br>no mutation | *V90I* | |
| SRIN94 | RNA | *K20I*<br>*L10V* | *V90I* | | *K20I*<br>*L10V* | *V90I* | |
| | DNA | no mutation<br>*L10V* | *V90I* | | no mutation<br>*L10V* | *V90I* | |
| SRIN95 | RNA | *K20I*<br>*L10I* | *V90I* | | *K20I*<br>*L10I* | *V90I* | |
| | DNA | *K20I*<br>*L10I* | no mutation | | *K20I*<br>*L10I* | no mutation | |

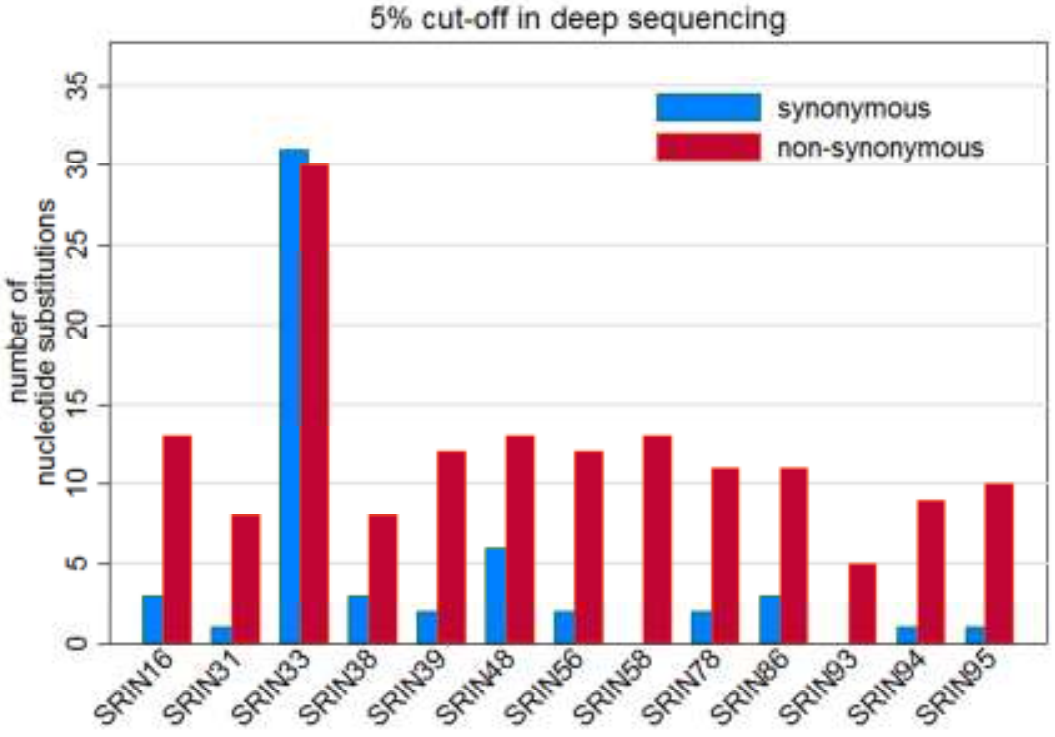
### 3.5.3 Patterns of synonymous and non-synonymous nucleotide differences compared among individual subjects from the thirteen HIV-1 infected drug naïve patients, at various sensitivity cut-offs in deep and bulk sequencing

Heterogeneity in RNA/DNA sequence pairs were assessed by aligning nucleotide sequence pairs of RNA/DNA in each of the 13 patient sequences for deep sequencing sensitivity cut-offs of 1, 2, 5, 10 and 20% and also in Sanger or bulk sequencing sensitivity cut-off which is usually 20%. In all deep sequencing sensitivity cut-offs, more non-synonymous nucleotide changes were recorded for twelve patients (SRIN16, 31, 38, 39, 48, 56, 58, 78, 86, 93, 94 and 95) except one (SRIN 33), who had more synonymous nucleotide changes than non-synonymous changes (figure 14A, 14B, 14C and 14D). In the case of bulk or Sanger sequencing, eight patients (SRIN16, 33, 39, 78, 86, 93, 94 and 95) had more non-synonymous nucleotide changes then synonymous changes, two (SRIN31, 38), had more synonymous nucleotide changes and the remaining three patients, (SRIN48, 56, 58), had equal number of synonymous and non-synonymous changes (figure 14E).

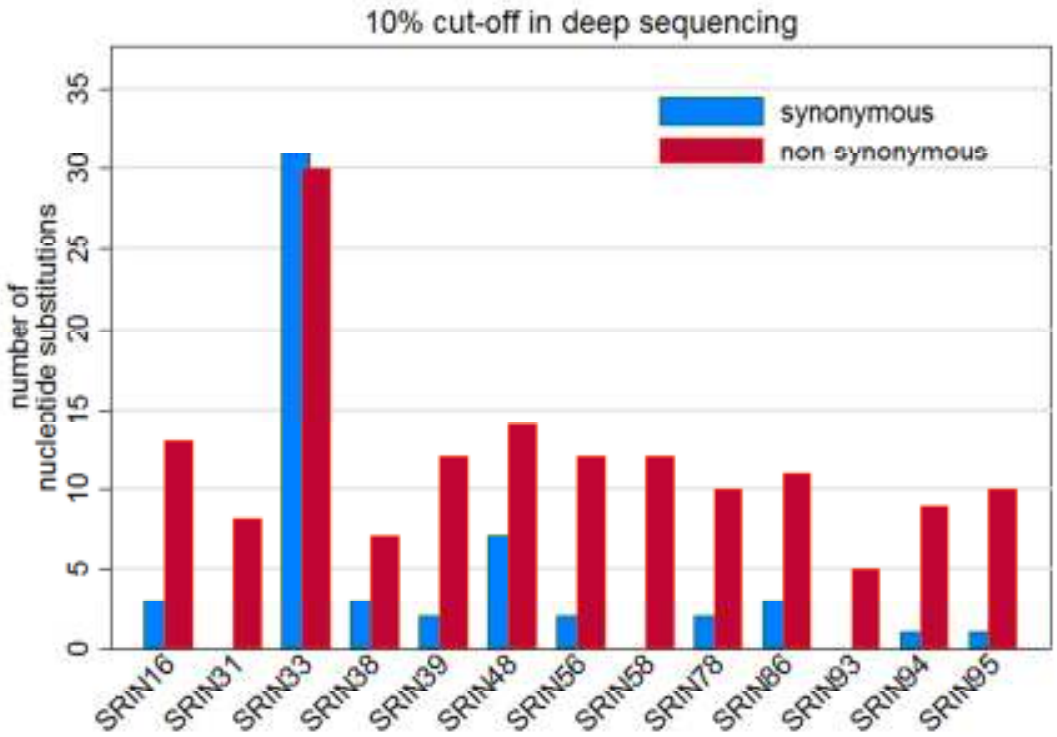
**A.**


1% cut-off in deep sequencing
synonymous
non-synonymous
number of nucleotide substitutions
0
5
10
15
20
25
30
35
SRIN16
SRIN31
SRIN33
SRIN38
SRIN39
SRIN48
SRIN56
SRIN58
SRIN78
SRIN86
SRIN93
SRIN94
SRIN95

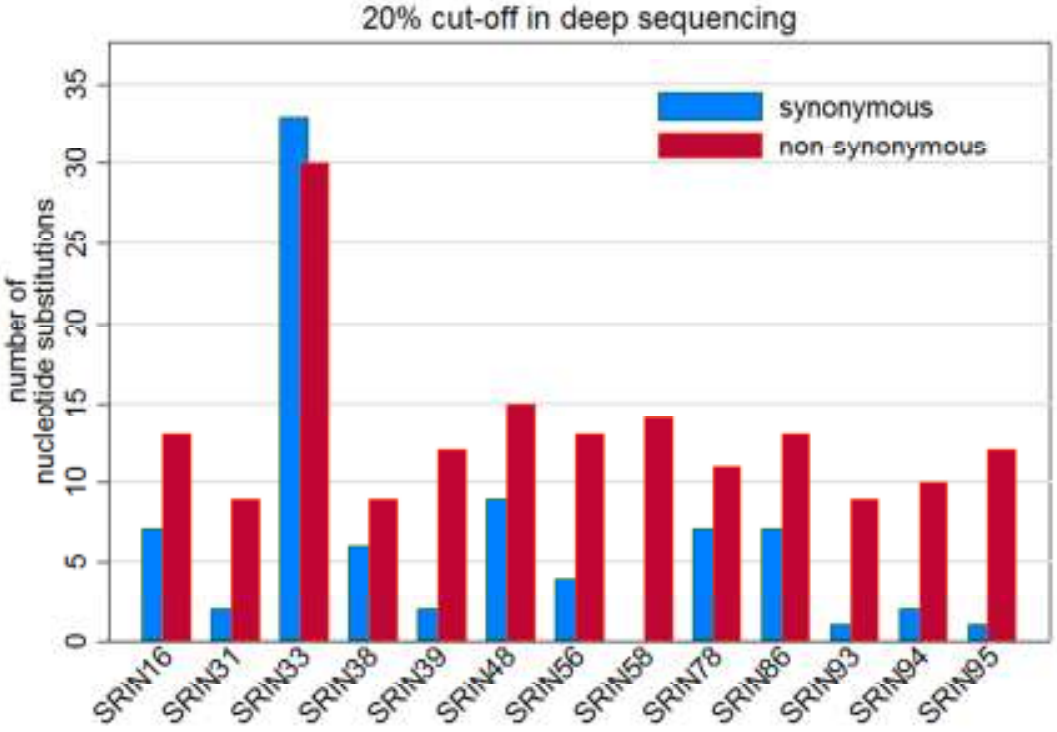
**B.**


5% cut-off in deep sequencing
synonymous
non-synonymous
number of nucleotide substitutions
0
5
10
15
20
25
30
35
SRIN16
SRIN31
SRIN33
SRIN38
SRIN39
SRIN48
SRIN56
SRIN58
SRIN78
SRIN86
SRIN93
SRIN94
SRIN95

**C.**


10% cut-off in deep sequencing
synonymous
non synonymous
number of nucleotide substitutions
0
5
10
15
20
25
30
35
SRIN16
SRIN31
SRIN33
SRIN38
SRIN39
SRIN48
SRIN56
SRIN58
SRIN78
SRIN86
SRIN93
SRIN94
SRIN95

**D.**


20% cut-off in deep sequencing
synonymous
non synonymous
number of nucleotide substitutions
0
5
10
15
20
25
30
35
SRIN16
SRIN31
SRIN33
SRIN38
SRIN39
SRIN48
SRIN56
SRIN58
SRIN78
SRIN86
SRIN93
SRIN94
SRIN95

**E.**

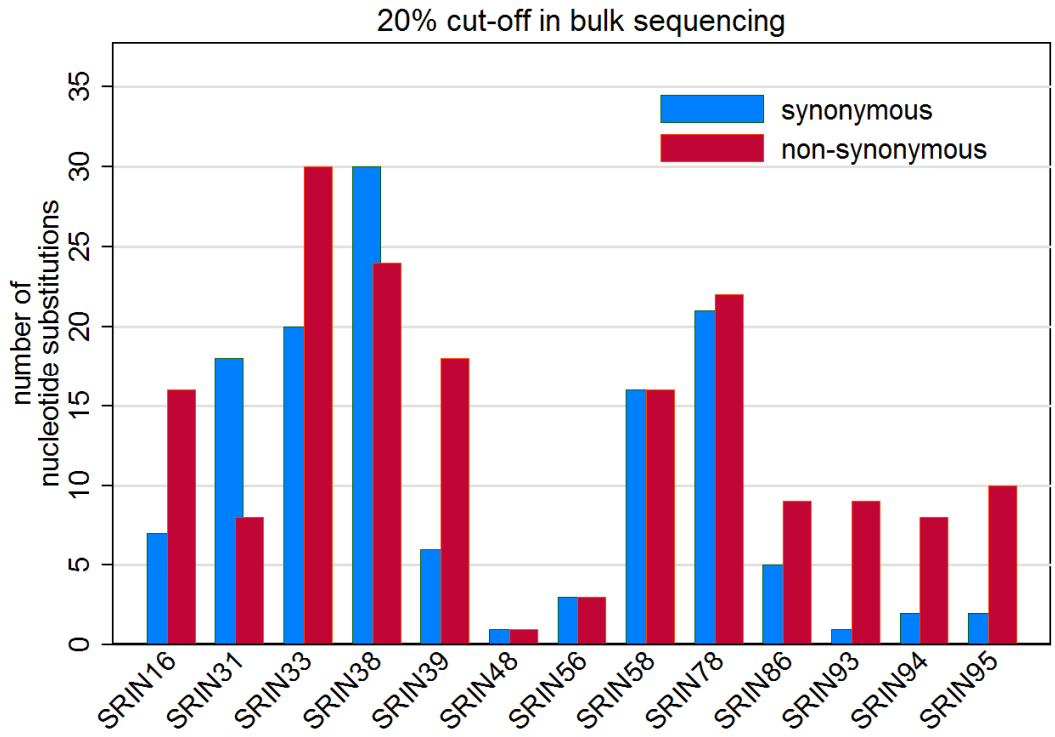



**Figure 14: Bar charts depicting synonymous and non-synonymous nucleotide differences in RNA and DNA templates of individual patients from the list of thirteen drug naive patient sequences as seen in both deep and bulk or Sanger sequencing of amplified PCR products.**

**In deep sequencing the cut-offs include 1, 2, 5, 10 and 20% and in Sanger or bulk sequencing the cut-off is approximately 20%. Synonymous nucleotide changes are indicated as blue bars and non-synonymous nucleotide changes are shown as red bars.**

### 3.5.4 Statistical evaluation of the pattern of synonymous and non-synonymous nucleotide differences compared among various cut-offs in deep and bulk sequencing

Nucleotide differences examined within each paired sequence of viral RNA and DNA of the thirteen drug naïve patient samples indicated that, at the one percent cut-off in deep sequencing, synonymous and non-synonymous nucleotide changes were significantly different ($p=0.0003$), with approximately two-fold increase in the number of nucleotide changes seen

between the synonymous and the non-synonymous groups. The same pattern of fold increase was observed in the five percent (p=0.0003), ten percent (p=0.0004) and twenty percent (p=0.0003) cut-offs. On the contrary, the difference between synonymous and the non-synonymous nucleotide differences seen in bulk or Sanger sequencing did not reach statistical significance (p=0.23) although more non-synonymous nucleotide differences were present in bulk sequencing of paired viral RNA and DNA sequences as compared to synonymous nucleotide differences. The non-synonymous group had more nucleotide differences than the synonymous group in all cut-offs in deep sequencing and also in bulk or Sanger sequencing (figure 15). Statistical significance was attained between all synonymous and non-synonymous differences for all deep sequencing cut-offs with more non-synonymous differences seen in each instance. This indicates that The heterogeneity seen within RNA/DNA template pairs, most likely lead to more divergence in the amino acid sequence of the protein structure. The results indicate that synonymous differences that generally conserve the protein structure were less frequently seen in deep sequencing. Also minority mutations existed that could not be determined with bulk sequencing.

Table 15 also gives the summary of nucleotide differences as observed in all thirteen patient samples. It indicates the number of synonymous and non-synonymous differences in each cut–off and shows the median and the IQR for each cut-off. The number of nucleotide differences in bulk sequencing either synonymous (132 nucleotide differences) or non-synonymous (173 nucleotide differences) was higher than differences observed in all the other cut-offs in deep sequencing although the synonymous and non-synonymous differences when compared within patient RNA/DNA pairs were not statistically significant. Noticeably non-synonymous nucleotide differences at the 20% cut-off in deep (170 nucleotide differences) and bulk sequencing (173 nucleotide differences) were similar, with median and IQRs of 12 (9.5-13.5) and 10 (8-20) respectively. However, with respect to synonymous changes the values were quite different. In the 20% cut-off in deep sequencing, 81 nucleotide differences were seen, while in bulk sequencing 132 nucleotide differences were present. The mean and IQRs of synonymous differences for the 20% cut-off in deep and bulk sequencing were 4 (1.5-7) and 6 (2-19) respectively. There was nearly two times more synonymous differences seen in bulk sequencing at a cut-off of 20% as compared to the same sensitivity cut-off in deep sequencing

indicating that, there was more heterogeneity in bulk sequenced samples at the 20% cut-off compared to deep sequencing probably due to the high degree of sensitivity in deep sequencing although these changes are ones that maintain protein structure since they do not lead to a change in the amino acid sequence. One reason for the heterogeneity seen in bulk sequencing data may be because it is read only once while deep sequenced data is read many times.

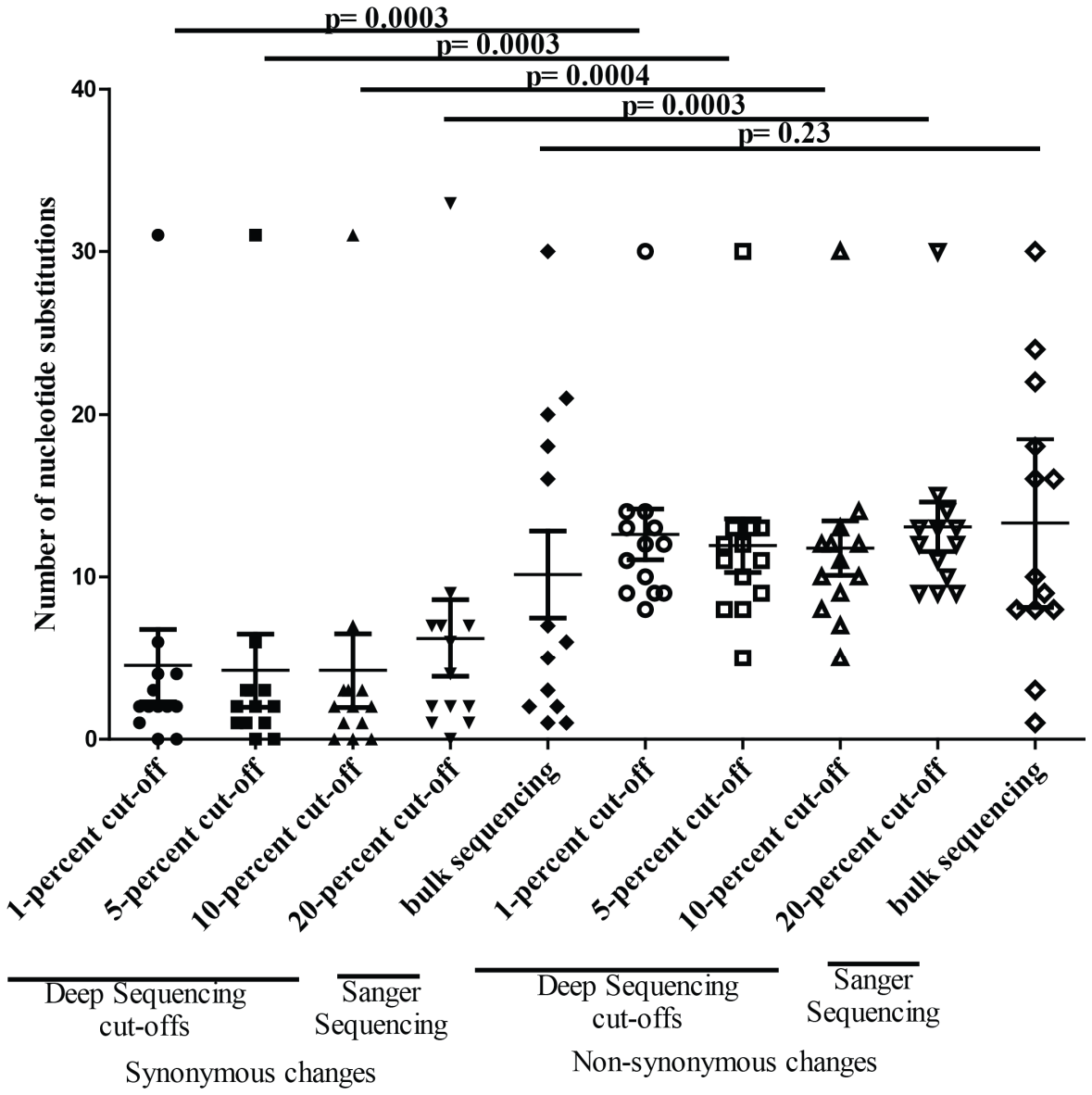



**Figure 15: Column plots depicting differences in synonymous and non-synonymous nucleotide differences among the thirteen drug naïve HIV-1 infected patients.**

**One percent cut-off in deep sequencing represents codons occurring at a frequency of 1 to 100% within each patient derived viral sequence. Five percent cut-off represents codons at a frequency of 5 to 100% of patient**

**derived viral sequences. Ten and 20% cut-offs denote codons from 10 to 100% and from 20 to 100% respectively within each patient derived viral sequence. Bulk or Sanger sequencing represents patient derived viral sequences occurring at a frequency of approximately 20 to 100% of patient derived viral sequences (Chabria et al., 2014). P values of synonymous and non-synonymous nucleotide changes between the same cut-offs are indicated on the graph. A one-way ANOVA showed that values of synonymous and non-synonymous nucleotide changes were different ($p< 0.001$).**

**Table 15: Summary of nucleotide substitution analysis for thirteen drug naïve patient derived viral sequences via deep and bulk sequencing**

| Synonymous nucleotide changes | Number of patients sequences with | Changes | | |
|---|---|---|---|---|
| | | Number | Median (IQR) | Maximum |
| 1-percent cutt-off | n=13 | 59 | 2(1.5-4) | 31 |
| 5-percent cutt-off | | 55 | 2(1-3) | 31 |
| 10-percent cutt-off | | 55 | 2(0.5-3) | 31 |
| 20-percent cutt-off | | 81 | 4(1.5-7) | 33 |
| 20-percent bulk seq | | 132 | 6(2-19) | 30 |
| **Non- synonymous nucleoide changes** | | | | |
| 1-percent cutt-off | n=13 | 164 | 12(9-13.5) | 30 |
| 5-percent cutt-off | | 155 | 11(8.5-13) | 30 |
| 10-percent cutt-off | | 153 | 11(8.5-12.5) | 30 |
| 20-percent cutt-off | | 170 | 12(9.5-13.5) | 30 |
| 20-percent bulk seq | | 173 | 10(8-20) | 30 |

## 3.6 Quantification of HIV-1 intracellular DNA of drug naïve patient samples using patient derived HIV-1 and CCR5 plasmid standards

Total HIV-1 IC DNA that has completed the first and second template switches also called HIV-1 STS DNA were measured from DNA samples of drug naïve patients using a real-time PCR assay. This was done to investigate any correlations existing between HIV-1 IC DNA levels of infected patients and primary markers of disease progression, mainly plasma RNA viral load, and $CD4^+$ T –cell count as well as correlations with other surrogate markers of disease progression.

### 3.6.1 Real-time PCR assay characteristics of HIV-1 intracellular DNA and CCR5 plasmid standards

Patient derived target HIV-1 IC DNA also called HIV-1 STS DNA and CCR5 PCR amplicons were successfully cloned into topo plasmids. The plasmids were further used to generate HIV-1 IC DNA and CCR5 DNA standards for the quantifications of all drug naïve patients DNA. HIV-1 IC DNA and CCR5 DNA inserts were 165 and 237 nucleotides in length. There was a 99% identity of the patient derived CCR5 DNA insert to human CCR5 reference sequences in the National Center for Biotechnology Information (NCBI) Database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). The HIV-1 IC DNA insert starts from the end of U5 to just before the gag sequence as shown in figure 16.

Figures 17A and 17B are the standard curves generated from the real- time PCR assays showing the quantifications of the HIV-1 IC DNA and CCR5 copies respectively. The specificity of the HIV-1 IC DNA assay was determined by accurately detecting 10 HIV-1 IC DNA copies with a qPCR dynamic range of $1x10^1$ to $1x10^6$ DNA copies. The sensitivity for CCR5 was determined by detecting 1000 CCR5 DNA copies accurately with a dynamic range of $1x10^3$ to $1x10^8$ DNA copies. The slopes of the standard curves were -3.3 cycles/$\log_{10}$ DNA templates for the HIV-1 IC DNA assay and that for the CCR5 assay was -3.4 cycles/$\log_{10}$ DNA templates. These slopes corresponded to PCR efficiencies of greater than 99.99% for both HIV-1 IC DNA and CCR5 assays. The optimum temperatures for the annealing and data collection of the HIV-1 IC DNA and CCR5 assays were optimized to 62°C and 60°C respectively.

The ability of the HIV-1 IC DNA assay to detect the M group HIV-1 strains was demonstrated by the assay's potential to quantify the HIV-1 IC DNA of all HIV-1 drug naïve patients DNA samples from Nouna Burkina Faso with HIV-1 genetic subtypes G, 02A1 and recombinants CRF01_AE, CRF02_AG and CRF06_cpx (Demetriou et al., 2010).

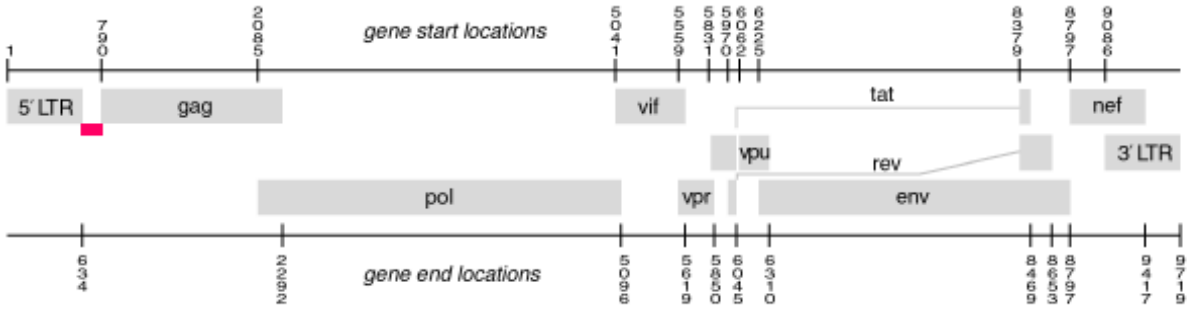



**Figure 16: Schematic representation of the patient derived HIV-1 STS DNA insert sequence, within the HIV-1 Genome.**

**The red mark shows the location of the sequence within the HIV-1 genome. The figure shows that the target starts from the end of U5 to just before the gag sequence with a sequence length of 165 nucleotides. The gene map was obtained from the HIV Los Alamos database (http://www.hiv.lanl.gov/cgi-bin/QUICK_ALIGNv2/convent.cgi).**

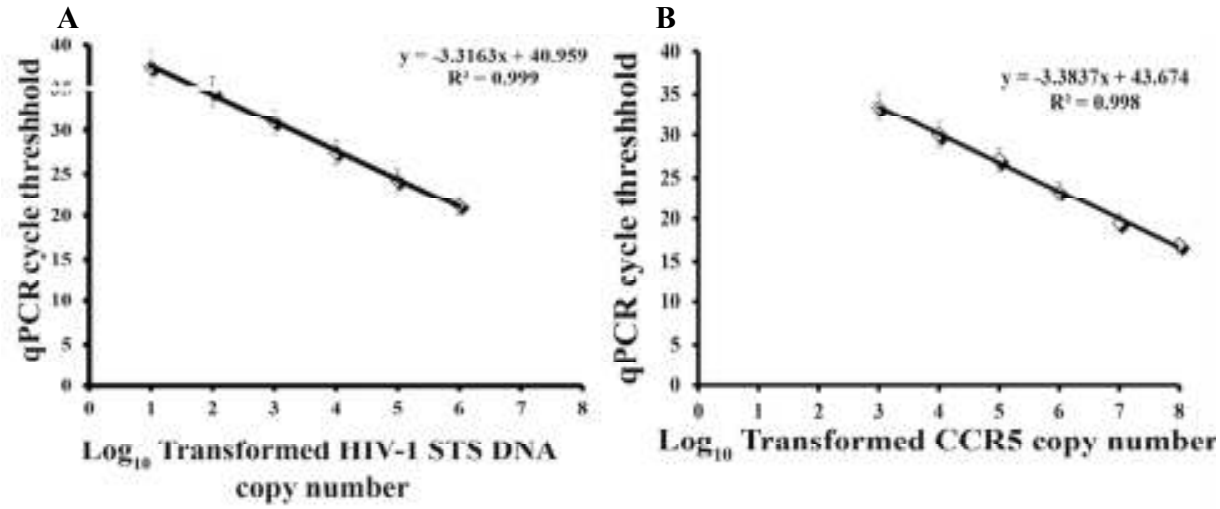



**Figure 17: Standard curves used for the HIV-1 IC DNA and CCR5 DNA quantifications.**

(A) A scatter plot representing HIV-1 STS DNA plasmid standard curve (B) A scatter plot representing human CCR5 DNA plasmid standard curve. Both curves had a co-efficient of determination $r^2 > 0.99$ and were both generated from molecular beacon real-time PCRs.

### 3.6.2 Specificity of binding of HIV-1 IC DNA primers and molecular-beacons to the target template

The assay specificity of the forward and reverse primer sets and the target specific molecular-beacon to the HIV-1 IC DNA was tested against HIV negative human DNA using a molecular-beacon-real-time qPCR approach. The real-time qPCR showed no detectable HIV-1 IC DNA in the HIV-1 negative human DNA under the regular HIV-1 IC DNA qPCR conditions of 40 cycles. Figures 18A and 18B demonstrates the standard curves obtained from the HIV-1 DNA standard mixed with the HIV negative human DNA and the HIV-1 IC DNA standard only. In both cases the PCR efficiencies were >99%. The HIV negative human DNA only real-time qPCR assay gave no cycle threshold value after 40 qPCR amplification cycles.

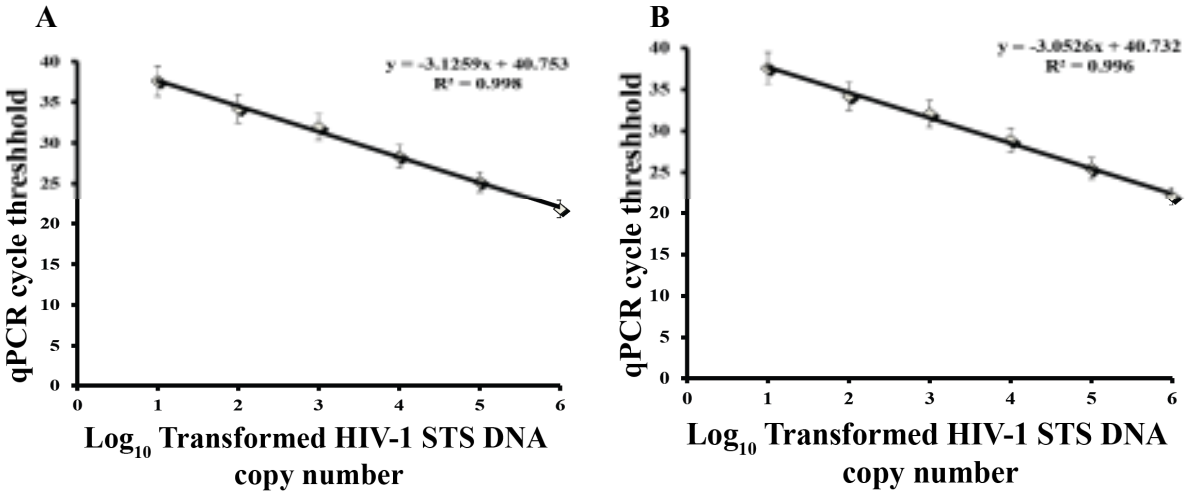



**Figure 18: A scatter plot representing the specificity of binding of HIV-1 STS DNA primers and beacons to the plasmid standard in the presence of an HIV negative human DNA control.**

**(A) HIV-1 STS DNA plasmid standard only. (B) HIV-1 STS DNA plasmid standard mixed with the HIV negative human DNA or non-template control (NTC). Tenfold serial dilutions of the purified plasmid and NTC of known concentrations were used as templates to generate the standard curves for the qPCR assay. HIV negative human DNA of known concentrations corresponding to $10^6$ HIV DNA copies contained 96 ng of DNA and five more dilution points were obtained. The last dilution point corresponding to 10 HIV DNA copies contained 9.6 x $10^{-4}$ ng of HIV negative human DNA.**

## 3.7 Distribution of HIV-1 IC DNA load among HIV -1 drug naïve patients

The HIV-1 IC DNA load and CCR5 copy numbers were determined for all 86 HIV-1 infected drug naïve patient DNA samples from Nouna Burkina Faso. The HIV-1 IC DNA load for each patient was expressed as per 1 million buffy coat cells. The number of cells in the input DNA used for the quantification of HIV-1 IC DNA was determined from the CCR5 copy number, which is known to exist as two copies in a cell (Samson et al., 1996).

According to Figure 19, all patient HIV-1 IC DNA load per 1 million cells were mainly between 100 to 1,000 copies/$10^6$ cells 50 (58%) followed by 1,000 to 10,000 copies/$10^6$ cells 22 (26%). The median $\log_{10}$ transformed HIV-1 IC DNA copies/ $10^6$ cells was 2.645 copies/ $10^6$ cells (IQR 2.259-3.055).

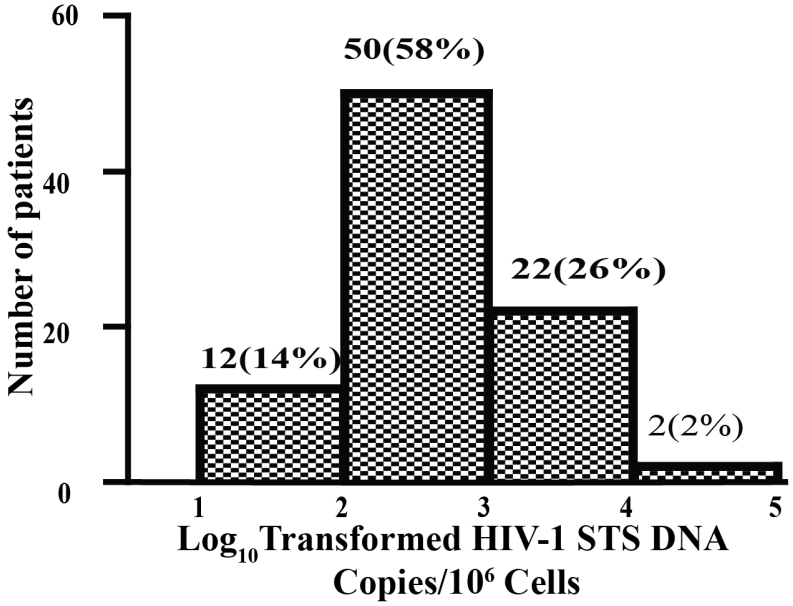



**Figure 19: A skewed curve showing HIV-1 STS DNA load among HIV -1 infected drug naïve patients from Nouna Burkina Faso.**

**The HIV-1 STS DNA load of patients was mainly between 100 and 1,000 copies/$10^6$ cells. The HIV-1 STS DNA load was calculated as HIV-1 STS DNA copies per 1million buffy coat cells. The total number of patient samples analyzed was 86 and the median $\log_{10}$ transformed HIV-1 IC DNA load was 2.645 (IQR 2.2596 – 3.055).**

## 3.8 Correlation of HIV-1 intracellular DNA load on known markers of disease progression

This section describes the correlation of HIV-1 IC DNA load and plasma viral load, $CD4^+$ T -cell counts and CDC clinical staging, based on $CD4^+$ T -cell counts of drug naive patients.

### 3.8.1 HIV-1 intracellular DNA load correlated with plasma viral load and $CD4^+$ T-cell counts

A potential correlation between HIV-1 IC DNA load and plasma viral load and that between HIV-1 IC DNA load and $CD4^+$ T -cell counts were assessed among all drug naïve patients as shown in figures 20A and 20B. In figure 20A there was no correlation between HIV-1 IC DNA load and plasma viral load ($r^2$=0.0032; p=0.6077). Also in figure 20B there was no significant correlation between HIV-1 IC DNA load and $CD4^+$ T-cell count ($r^2$=0.0172; p=0.2292). The individual patient values used in the correlation are shown in table 16.

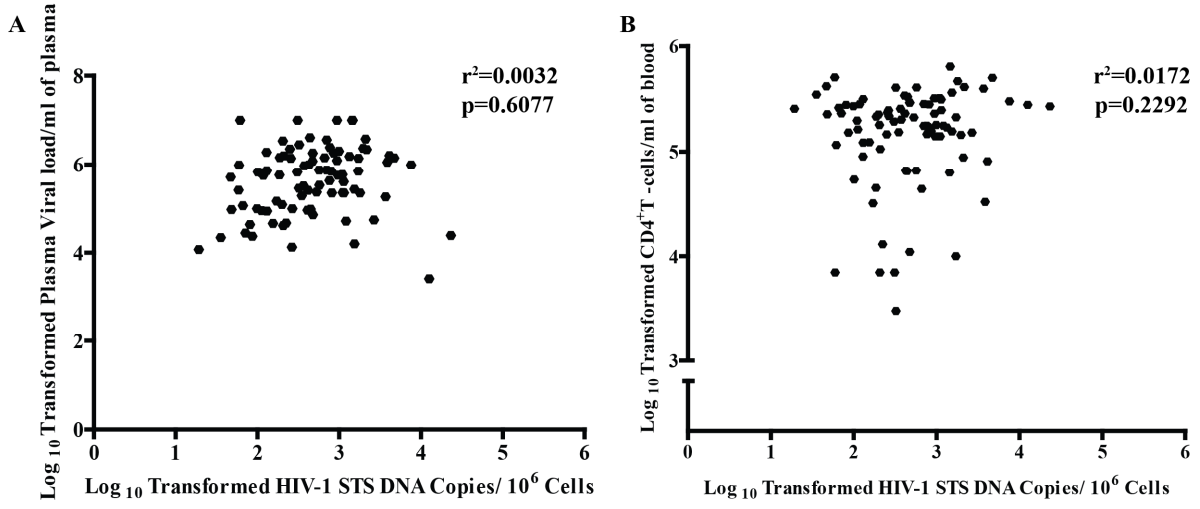



**Figure 20: Correlations of HIV-1 STS DNA load against plasma viral load and $CD4^+$ T –cells among HIV-1 infected drug naive patients from Nouna Burkina Faso.**

**(A) Correlation of plasma viral load/ml of plasma against HIV-1 STS DNA load/$10^6$ cells, (B) Correlation of $CD4^+$T -cells/ml of blood against HIV-1 STS DNA load/$10^6$ cells. At 95% confidence interval, there was no correlation between HIV-1 STS DNA and plasma viral load and HIV-1 STS DNA and $CD4^+$ T-cell counts**

**Table 16: HIV-1 STS DNA, $CD4^+$ T-cell counts and plasma viral load values of drug naïve patients used for correlations**

| Lab ID | HIV DNA Copies/ 10^6 Cells | CD4 count/ml of blood | Viral load/ml of plasma |
|---|---|---|---|
| SRIN14 | 373.96 | 202000 | 939988 |
| SRIN15 | 59.47 | 7000 | 971579 |
| SRIN16 | 1703.59 | 10000 | 707181 |
| SRIN25 | 2662.58 | 152000 | 54837 |
| SRIN30 | 112.14 | 163000 | 89125 |
| SRIN31 | 99.10 | 271000 | 98426 |
| SRIN32 | 759.81 | 177000 | 430515 |
| SRIN33 | 204.31 | 180000 | 3399058 |
| SRIN34 | 322.99 | 3000 | 2801732 |
| SRIN37 | 794.99 | 172000 | 707210 |
| SRIN38 | 1142.02 | 248000 | 223957 |
| SRIN39 | 66.23 | 262000 | 115615 |
| SRIN41 | 1964.83 | 144000 | 2314931 |
| SRIN47 | 567.88 | 66000 | 743697 |
| SRIN48 | 766.43 | 147000 | 2416935 |
| SRIN51 | 129.96 | 122000 | 722623 |
| SRIN52 | 348.33 | 153000 | 194086 |
| SRIN55 | 48.01 | 227000 | 93865 |
| SRIN56 | 4107.70 | 81000 | 1593082 |
| SRIN57 | 110.11 | 198000 | 653455 |
| SRIN58 | 414.57 | 231000 | 256724 |
| SRIN59 | 940.93 | 230000 | 1229920 |
| SRIN60 | 199.34 | 226000 | 121581 |
| SRIN61 | 702.55 | 284000 | 749061 |
| SRIN62 | 989.57 | 180000 | 2005003 |
| SRIN63 | 130.27 | 316000 | 86731 |
| SRIN64 | 306.21 | 193000 | 677363 |
| SRIN65 | 171.00 | 32000 | 145511 |
| SRIN66 | 1103.67 | 140000 | 227199 |
| SRIN67 | 470.74 | 293000 | 1195067 |
| SRIN68 | 2158.53 | 411000 | 2201354 |
| SRIN69 | 46.95 | 418000 | 515487 |
| SRIN70 | 223.52 | 13000 | 46747 |
| SRIN71 | 365.78 | 256000 | 325435 |
| SRIN72 | 445.47 | 333000 | 95911 |
| SRIN73 | 856.60 | 157000 | 1818940 |
| SRIN74 | 206.62 | 7000 | 40625 |
| SRIN75 | 263.20 | 247000 | 13291 |
| SRIN76 | 1210.15 | 177000 | 51695 |
| SRIN77 | 154.99 | 123000 | 45750 |
| SRIN78 | 23216.97 | 269000 | 24482 |
| SRIN79 | 19.22 | 256000 | 11678 |
| SRIN80 | 86.45 | 152000 | 23449 |
| SRIN81 | 1091.12 | 323000 | 590908 |
| SRIN82 | 963.52 | 140000 | 569328 |
| SRIN83 | 80.76 | 279000 | 43140 |

**Table 16 continued**

| Lab ID | HIV DNA Copies/ 10^6 Cells | CD4 count/ml of blood | Viral load/ml of plasma |
|---|---|---|---|
| SRIN84 | 1531.37 | 156000 | 15793 |
| SRIN85 | 187.71 | 215000 | 1434896 |
| SRIN86 | 249.71 | 146000 | 2233224 |
| SRIN87 | 437.89 | 65000 | 4040062 |
| SRIN88 | 185.29 | 45000 | 574163 |
| SRIN89 | 101.39 | 54000 | 677363 |
| SRIN90 | 255.56 | 236000 | 1379797 |
| SRIN91 | 425.18 | 66000 | 977237 |
| SRIN92 | 117.75 | 282000 | 561916 |
| SRIN93 | 1456.33 | 643000 | 10000000 |
| SRIN94 | 61.46 | 116000 | 10000000 |
| SRIN95 | 310.80 | 7000 | 10000000 |
| SRIN96 | 129.41 | 90000 | 1892979 |
| SRIN97 | 209.51 | 106000 | 1581674 |
| SRIN98 | 1431.20 | 63000 | 10000000 |
| SRIN99 | 936.53 | 322000 | 10000000 |
| SRIN100 | 475.62 | 292000 | 71380 |
| SRIN101 | 572.97 | 408000 | 339531 |
| SRIN102 | 2096.61 | 88000 | 3790705 |
| SRIN103 | 3679.25 | 397000 | 182608 |
| SRIN105 | 265.99 | 218000 | 98211 |
| SRIN106 | 701.95 | 176000 | 3589550 |
| SRIN107 | 321.07 | 406000 | 290269 |
| SRIN108 | 71.18 | 232000 | 27833 |
| SRIN109 | 1134.23 | 313000 | 405359 |
| SRIN110 | 661.71 | 44000 | 1420627 |
| SRIN111 | 529.86 | 212000 | 236592 |
| SRIN112 | 1717.59 | 213000 | 1373030 |
| SRIN113 | 472.80 | 11000 | 1803350 |
| SRIN114 | 35.59 | 349000 | 21777 |
| SRIN116 | 58.90 | 508000 | 262060 |
| SRIN117 | 410.15 | 341000 | 90498 |
| SRIN118 | 12586.42 | 279000 | 2579 |
| SRIN119 | 1529.91 | 364000 | 272998 |
| SRIN120 | 4701.27 | 504000 | 1400866 |
| SRIN121 | 811.20 | 281000 | 224037 |
| SRIN122 | 1330.56 | 170000 | 1531228 |
| SRIN123 | 7585.02 | 301000 | 976517 |
| SRIN124 | 1792.91 | 467000 | 225569 |
| SRIN125 | 3856.28 | 33000 | 1126782 |

## 3.9 Correlation of HIV-1 intracellular DNA load with plasma viral load and CD4$^+$ T-cell counts based on CDC clinical Stage B of the disease

The relationship between HIV-1 IC DNA load, plasma viral load and CD4$^+$ T-cell counts was assessed among HIV-1 infected drug naïve patients based on the Centers for Disease Control and Prevention (CDC) clinical Stage B of the disease as depicted in figures 21A and 21B. Correlations were considered based on CDC clinical disease stage because it was important to find out if HIV-1 IC DNA had any effect on disease staging. Clinical stage B was correlated with HIV-1 IC DNA levels at baseline because almost all the patients with clinical stating status recorded were in clinical stage B (n=56) except one patient who was in clinical stage A (table 8) There was no correlation between HIV-1 IC DNA load and plasma viral load ($r^2$=0.0020; p=0.7429) and between HIV-1 IC DNA load and CD4$^+$ T-cell counts ($r^2$=0.0119; p=0.4242).

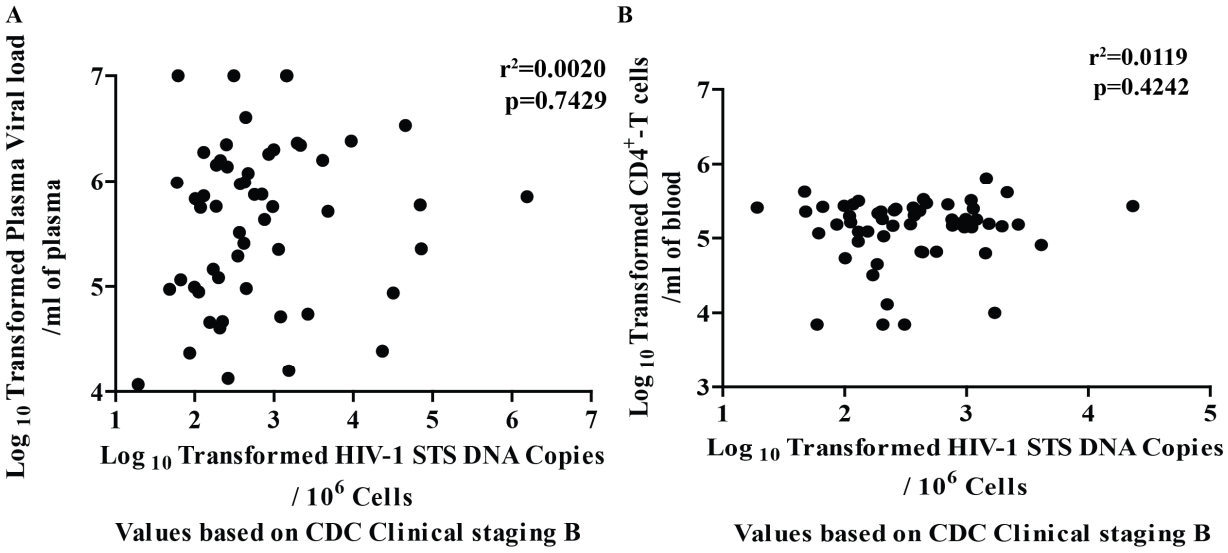



**Figure 21: HIV-1 STS DNA load based on CDC Staging of disease against plasma viral load and CD4$^+$ T-cells among drug naïve patients.**

(A) An XY correlation between plasma viral load/ml of plasma and HIV-1 STS DNA load/$10^6$ cells based on CDC clinical Stage B (B) An XY correlation between CD4$^+$ T-cell counts/ml of blood against HIV-1 STS DNA load/$10^6$ cells.

## 3.10 Assessing disease progression after six months of therapy based on HIV-1 intracellular DNA levels

### 3.10.1 Comparison of HIV-1 IC DNA levels in patients with virologic failure at twenty four weeks of therapy

In assessing disease progression based on baseline HIV-1 DNA levels, follow-up patients were grouped into those who achieved virologic suppression under HAART and those who had virologic failure. Virologic failure was defined as viral RNA levels of ≥ 1,000 copies/ ml of plasma detected in follow-up patient samples after 24 weeks or 6 months of HAART and virologic suppression was defined as the attainment of viral RNA levels of < 1,000 copies/ ml of plasma detected in follow-up patient samples after 24 weeks or 6 months under HAART (Dionisio et al., 2001). The baseline HIV-1 IC DNA load of patients who experienced virologic failure at twenty-four weeks of therapy (n=5) was compared to that of those who experienced virologic suppression (n=55). There was a significantly higher baseline HIV-1 IC DNA load among patients who experienced virologic failure than those who did not (p=0.045). The corresponding median levels were 3.16 $\log_{10}$ copies/$10^6$ cells (IQR 2.75- 3.62) and 2.63 $\log_{10}$ copies/$10^6$ cells (IQR, 2.12 -3.04) respectively (figure 22).

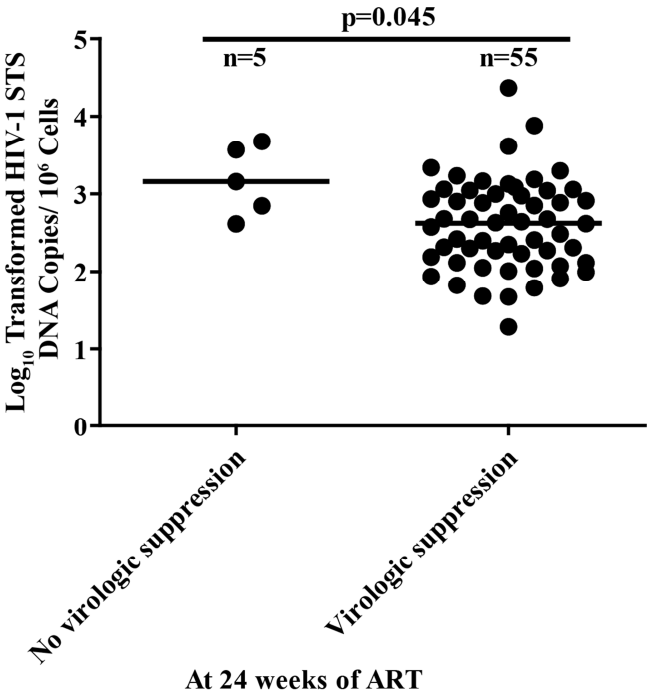



**Figure 22: HIV-1 STS DNA levels of patients experiencing virologic failure and those with no virologic failure, 24 weeks post HAART. Comparisons of plasma viral levels of infected patients with and without virologic failure**

The plasma viral load and $CD4^+$ T-cell counts of patients experiencing virologic failure and those in whom viral suppression was achieved were compared at the drug naïve stage and at 24 weeks of therapy. At the drug naïve stage, the results showed no difference in viral load and $CD4^+$ T-cell counts between patients who experienced virologic suppression post therapy and those who had virologic failure (figure 23C and 23D). The median viral load levels/ ml of plasma among those who had virologic failure and those who did not were 6.15 $\log_{10}$ (5.14-6.83) and 5.82 $\log_{10}$ (4.97-6.16) respectively. The median $CD4^+$ T-cell counts were 5.53 $\log_{10}$ (5.08-5.65) and 5.26 $\log_{10}$ (5.09-5.43) respectively between the two groups of patients. However, at 24 weeks into therapy, there was a strong difference in plasma viral load (approximately 2.5 $\log_{10}$ difference, $p<0.001$) observed among patients who achieved virologic suppression and those who did not (figure 23A). The median viral load levels were 4.11 $\log_{10}$ (IQR 3.41-5.02) and 1.60 $\log_{10}$ (IQR 1.60-1.99) respectively. On the contrary, there was no significant difference observed in $CD4^+$ T-cell counts among these patients (figure 23B). The median $CD4^+$ T-cell count at 24 weeks of therapy was 5.53 $\log_{10}$ (5.37-5.60) and 5.61 $\log_{10}$ (5.48-5.85) among patients who experienced virologic suppression and those who had virologic failure respectively.

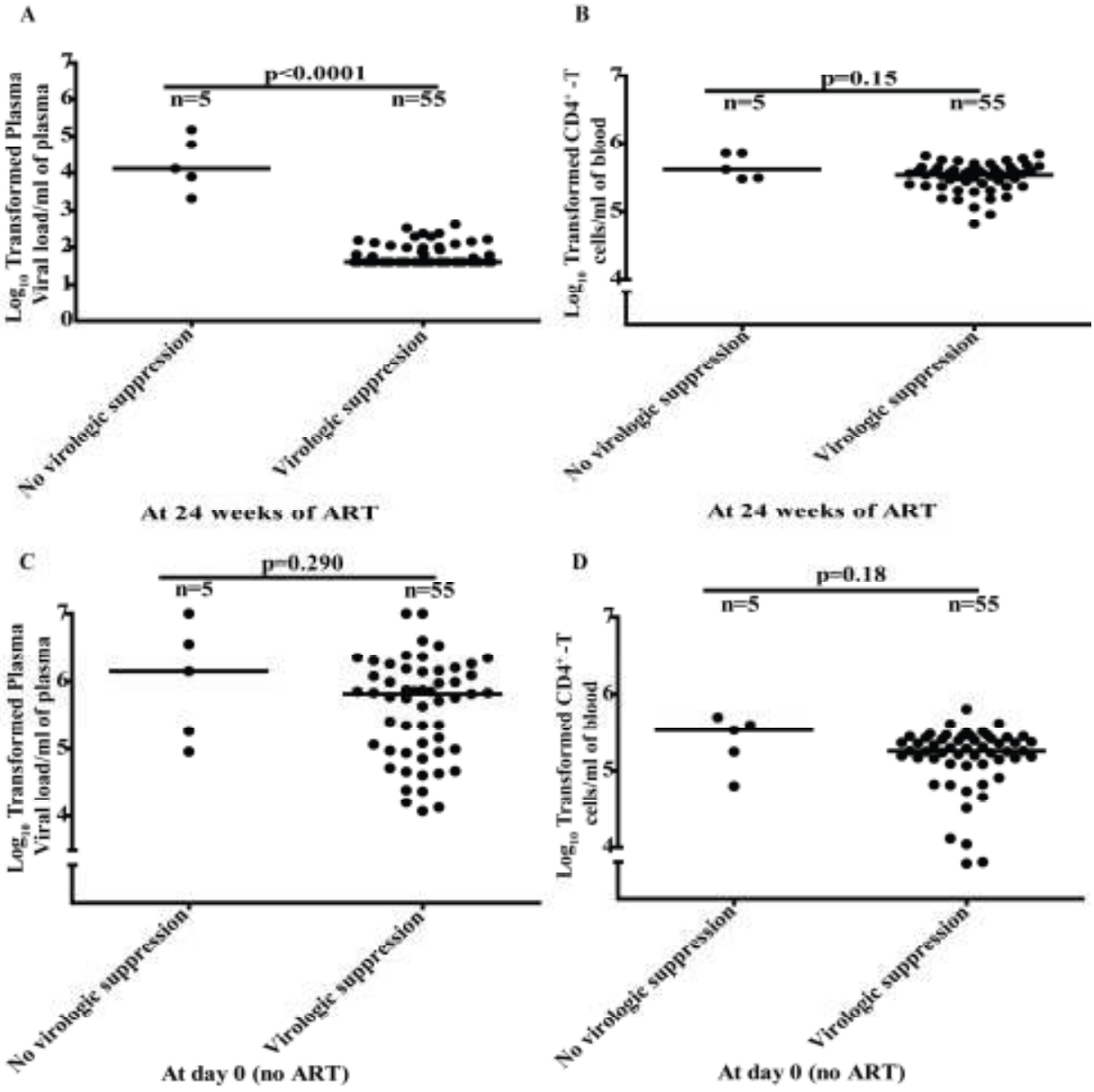



**Figure 23: Plasma viral load and $CD4^+$ T-cell counts comparisons among patients experiencing virologic failure and those who achieved virologic suppression after 24 weeks on HAART.**

## 3.11 Influence of other factors associated with HIV-1 disease progression on HIV-1 intracellular DNA load

This section describes the influence of other factors associated with disease progression such as the presence or absence of hepatitis B (HBV) co-infection, number of DRMs that have an impact on drug susceptibility, predominant HIV-1 subtypes, age Hb levels and the presence or absence of secondary bacterial or parasitic infections (opportunistic infections) on HIV-1 IC DNA load.

### 3.11.1 Hepatitis B co-infection

In figure 24A, there is a trend towards a higher HIV-1 IC DNA load among drug naïve patients with HBV co-infection as compared to those without HBV co-infection but this difference did not reach statistical significance (p= 0.1496). As illustrated in figure 24B and 24C, there was no difference in baseline $CD4^+$ T-cell counts (p= 0.6806) and plasma viral load levels (p= 0.3638) in drug naïve patients with HBV co-infection and those without HBV co-infection.

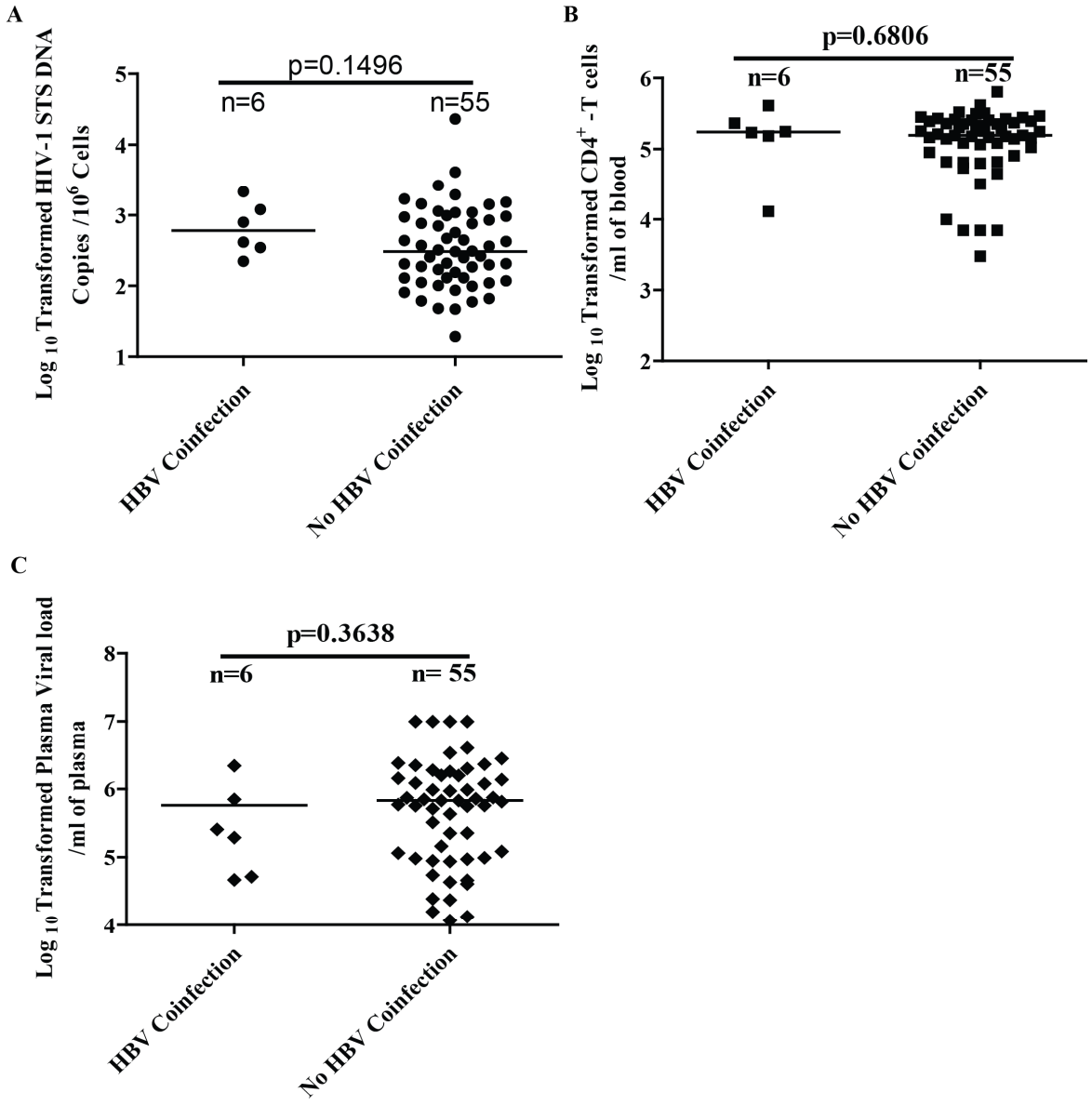



**Figure 24: Hepatitis B co-infections among drug naïve patients.**

**Boxes and whiskers of (A) HIV-1 STS DNA load levels (p= 0.1496), (B) $CD4^+$ T-cell counts (p= 0.6806) and (C) Plasma viral load (p= 0.3638) among drug naïve patients with HBV co-infection and those without HBV co-infection.**

### 3.11.2 Age, predominant HIV-1 subtypes, number of mutations and hemoglobin levels

Figure 25A shows an analysis that assessed the relationship between HIV-1 IC DNA levels and the age of drug naïve patients in years. There was no correlation between HIV-1 IC DNA load and the age of patients ($r^2$= 0.0224; p= 0.1692). Using the Kruskal-Wallis test, differences in HIV-1 IC DNA levels among drug naïve patients having predominant HIV-1 subtypes namely CRF02_AG, CRF06_cpx and CRF01_AE were analyzed and there was no significant difference (p=0.8162) seen between all subtypes compared (Figure 25B). We proceeded to determine the variances of HIV-1 IC DNA levels between subtypes using the Bartlett's test for equal variances as shown in table 17 and demonstrated that, the spread of HIV-1 IC DNA levels between HIV-1 subtypes analyzed are unequal (p<0.0001). There was no relationship ($r^2$= 0.1056; p= 0.4323) between HIV-1 IC DNA and number of mutations affecting drug susceptibility as shown in figure 25C.

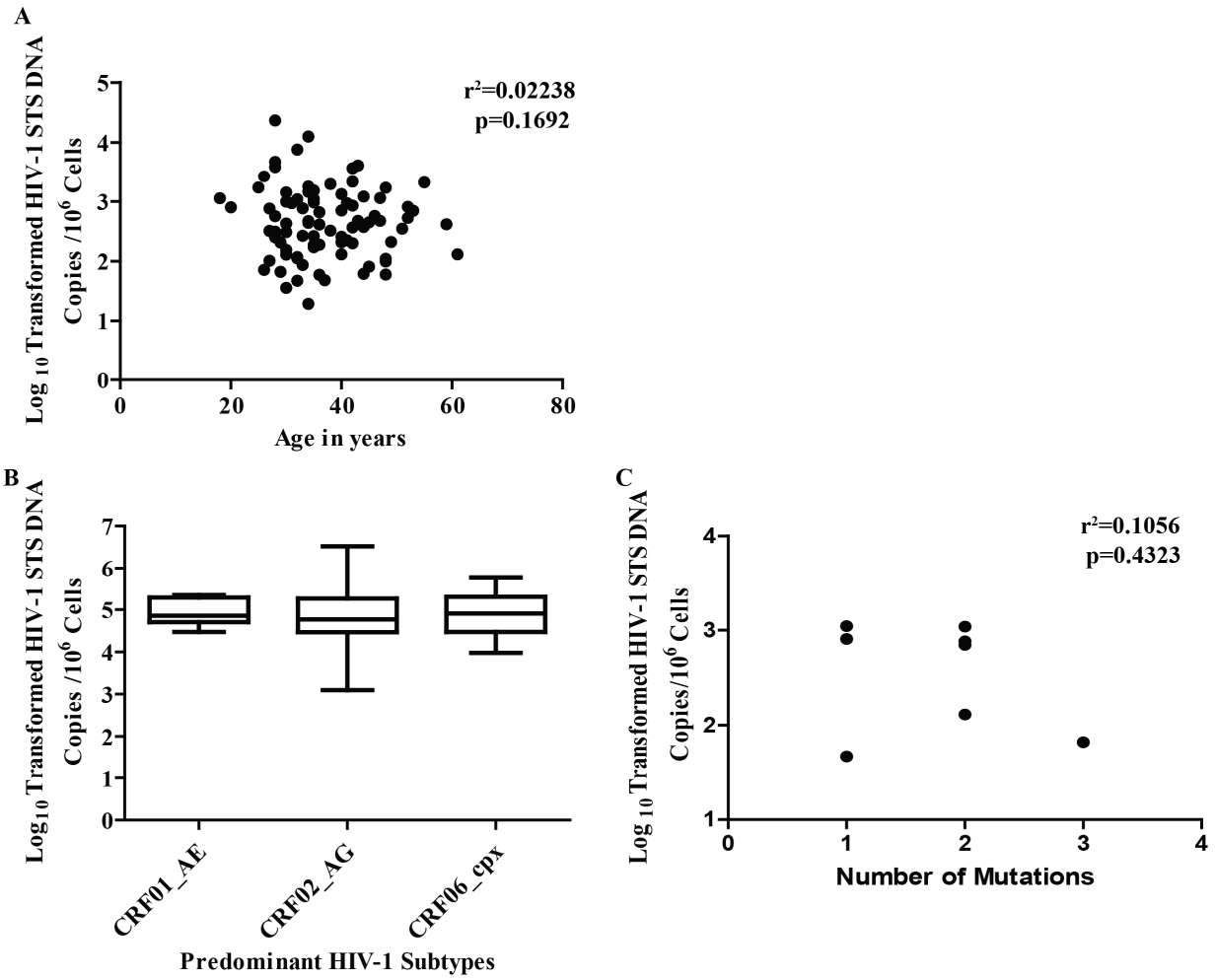



**Figure 25: HIV-1 STS DNA load compared with age of patients, virus subtype and number of DRMs.**

**(A) A correlation of HIV-1 STS DNA load against age of drug naïve patients, (B) A box and whisker plot of HIV-1 STS DNA load with the predominant HIV-1 subtypes and (C) A correlation of HIV-1 STS DNA load to number of mutations affecting drug susceptibility.**

**Table 17: The Bartlett's test for equal variances among predominant HIV-1 subtypes**

| HIV -1 Subtype | Variable | Observations | Mean | Std Deviation | Variance | Mimmun value | Maximum value |
|---|---|---|---|---|---|---|---|
| CRF01_AE -Like | HIV-1 STS DNACopies | 5 | 3525356 | 5269.46 | 2,777E+11 | 8644538 | 12586.42 |
| CRF02_AG -Like | HIV-1 STS DNACopies | 63 | 8772681 | 1284.052 | 1,649E+12 | 1922446 | 7585018 |
| CRF06_cpcx -Like | HIV-1 STS DNACopies | 12 | 733997 | 7094512 | 5,033E+13 | 1177456 | 2662583 |

## 3.12 Comparison of HIV-1 intracellular DNA load, plasma viral load and $CD4^+$ T-cell counts of susceptible and resistant HIV-1 strains among drug naïve patients

These consisted of drug naïve patients harbouring drug sensitive strains (n=75) or drug resistant strains (n=8) of HIV-1. In figures 26A, 26B and 26C, there were no significant differences in terms of HIV-1 IC DNA load (p=0.2633), viral load (p=0.3585) and $CD4^+$ T-cell counts (p=0.0723) among drug resistant and drug sensitive strains of HIV-1 infected drug naïve patients.

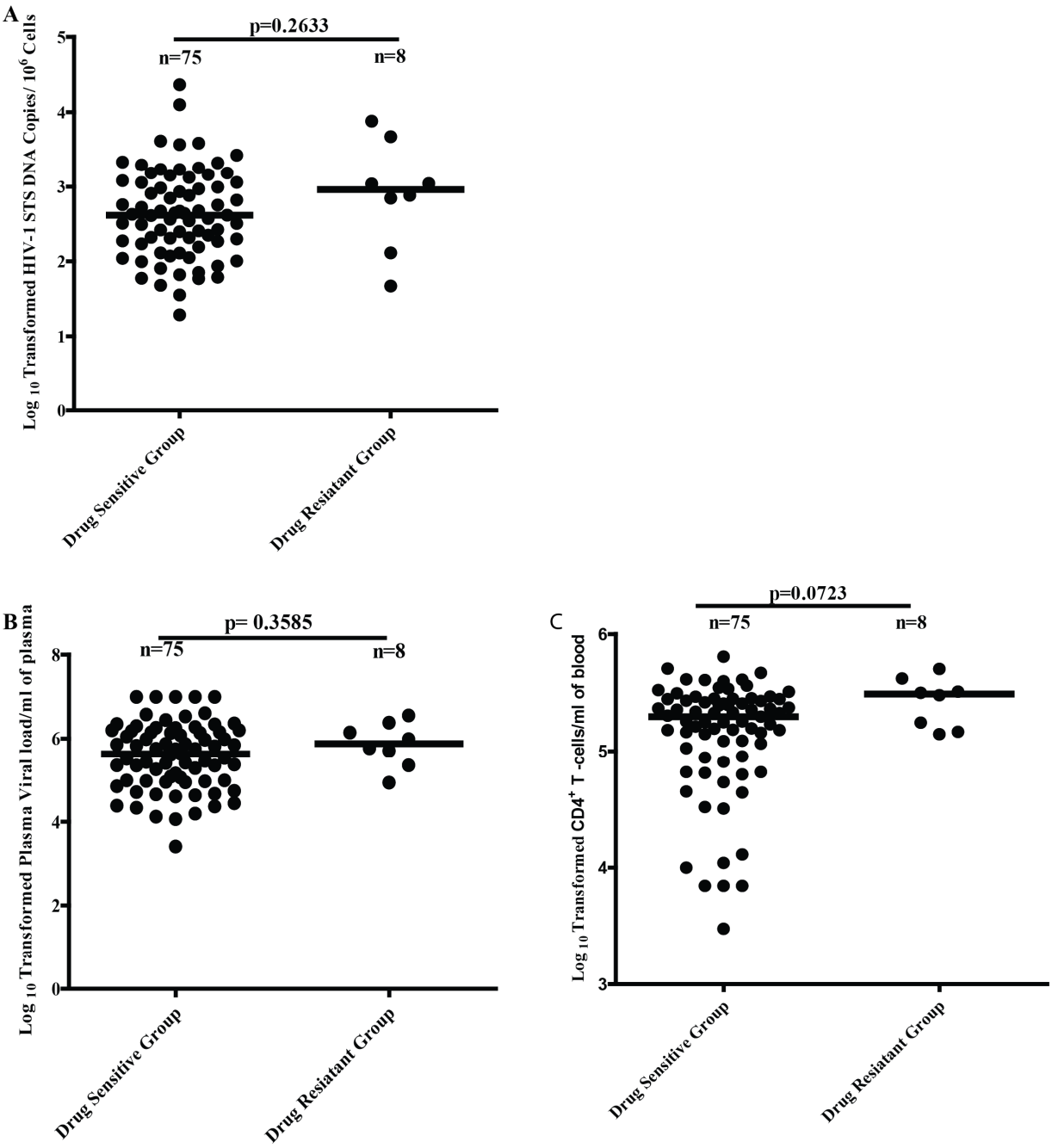



**Figure 26: Vertical scatter plots of HIV-1 STS DNA load, plasma viral load and $CD4^+$ T-cell counts of drug resistant and drug sensitive strains of HIV-1 infected drug naive patients from Nouna Burkina Faso.**

**The numbers of patient samples analyzed were 83. (A) HIV-1 STS DNA load in drug resistant and drug sensitive strains. The median $log_{10}$ transformed HIV-1 STS DNA/$10^6$ cells and IQR in resistant and sensitive strains were 2.961 (IQR 2.298- 3.515) and 2.618 (IQR 2.233- 3.058) respectively, (B) Plasma viral load in drug resistant and drug sensitive strains. The median $log_{10}$ transformed viral load and IQR in resistant and sensitive strains were 5.881 (IQR 5.445- 6.324) and 5.634 (IQR 4.982- 6.185) respectively and (C) $CD4^+$ T-cell**

**counts in drug resistant and drug sensitive strains. The median $log_{10}$ transformed $CD4^+$ T-cell counts and IQR in resistant and sensitive strains were 5.489 (IQR 5.187- 6.5.593) and 5.297 (IQR 4.954- 5.446) respectively.**

### 3.12.1 Categorization of HIV-1 intracellular DNA load plasma with plasma viral load and $CD4^+$T -cells among HIV-1 drug resistant and drug sensitive patients by CDC clinical staging

A total number of 57(98%) drug naïve patients had their CDC clinical staging recorded and among these patients, 56 were within the clinical stage B and only 1(2%) patient was within the clinical stage A category. This section of the analysis was done with only clinical stage B patients as a dominant group to determine whether clinical staging of HIV-1 influences baseline HIV-1 IC DNA levels, $CD4^+$ T-cell counts and plasma viral load among drug resistant and drug susceptible strains of HIV. These patients were divided into two groups namely the drug resistant group and the drug sensitive group. A column plot was then used to assess the two groups with respect to their HIV-1 IC DNA copies/$10^6$ cells, plasma viral load/ml of plasma and $CD4^+$ T-cell counts/ml of blood as indicated in figures 27A, 27B and 27C respectively. The median HIV-1 IC DNA values among resistant and sensitive strains were 2.885 (IQR 1.948- 3.040 $log_{10}$ copies /$10^6$ cells) and 2.493 (IQR 2.114- 2.984 $log_{10}$ copies /$10^6$ cells) respectively. The median plasma viral load copies for resistant and sensitive viruses were 5.712 (IQR 5.1959- 6.177 $log_{10}$ copies /ml of plasma) and 5.815 (IQR 4.982- 6.202 $log_{10}$ copies /ml of plasma) respectively. Finally the median $CD4^+$ T-cell counts/ml of blood among resistant and sensitive virus groups was 5.500 (IQR 5.157- 5.569 $log_{10}$ $CD4^+$ T-cell counts/ml of blood) and 4.501 (IQR 4.214- 4.698 $log_{10}$ $CD4^+$ T-cell counts/ml of blood) respectively. According to figure 27C, resistant strains had significantly higher $CD4^+$ T-cell counts as when compared to drug sensitive patents ($p=0.0003$). No significant changes were observed in HIV-1 IC DNA loads and plasma viral load of drug resistant and drug sensitive strains as shown in figures 27A and 27B.

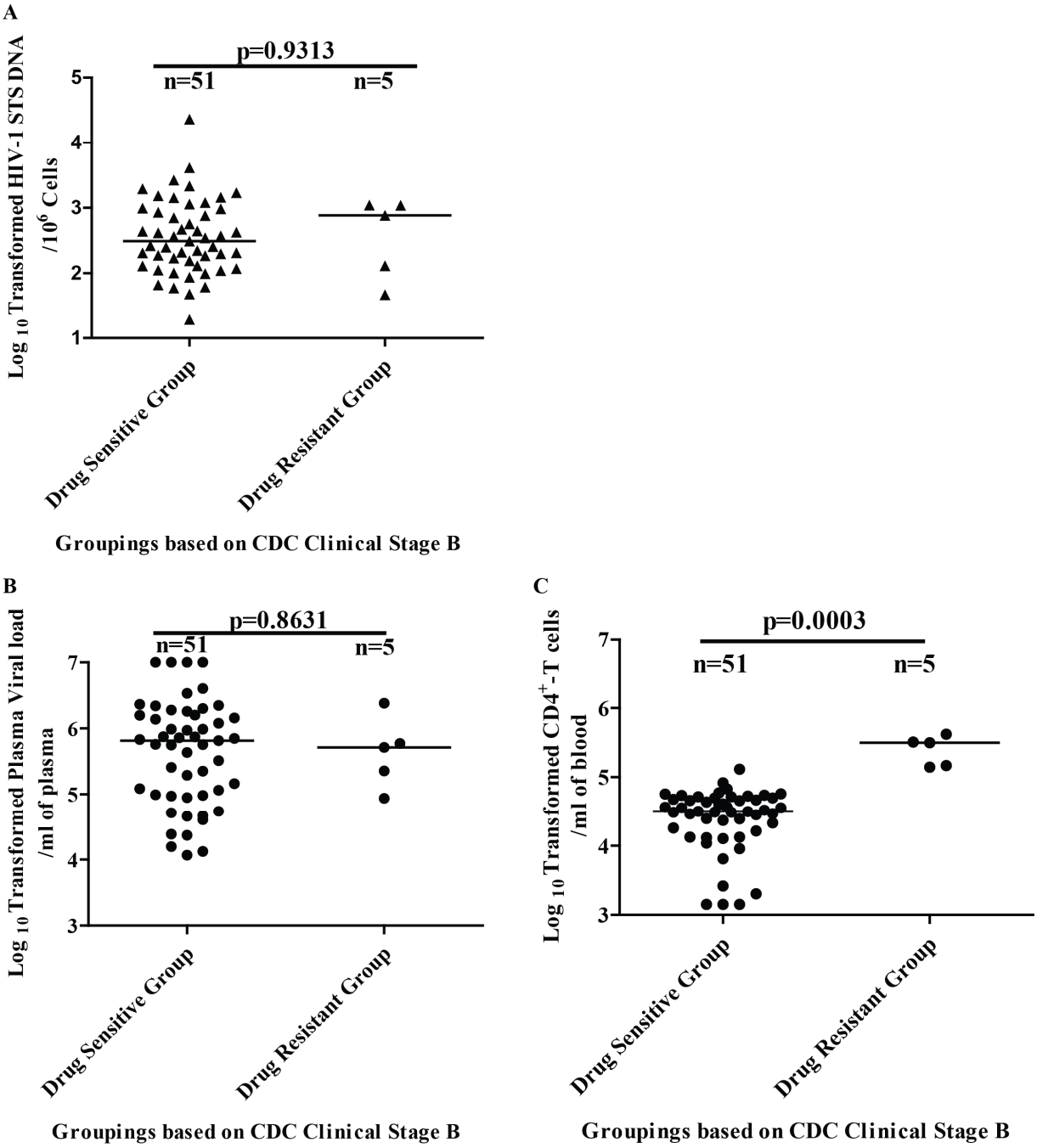



**Figure 27: Column plots of HIV-1 STS DNA load against plasma viral load and $CD4^+$ T-cell counts of drug resistant and drug sensitive strains using CDC clinical stage B patients only.**

**(A) HIV-1 STS DNA load of drug resistant and drug sensitive strains (B) Plasma viral load in drug resistant and drug sensitive strains (C) $CD4^+$ T-cell counts in drug resistant and drug sensitive strains.**

## 3.13 Principal component analysis for disease progression among drug naïve patients

Principal component analysis (PCA) is a data reduction method used when one has several related variables that measure different aspects of a common component (Vyas and Kumaranayake, 2006) in our case disease progression. In PCA, one comes up with a measure that holistically describes the different variables forming the component. It is used to transform a set of related data into a set of linearly unrelated components in which the first component gives the most variance or the largest variability existing in the data. Patients' experimental and biodata were extracted for the analysis. The indicators used to explain PCA could be continuous or categorical (Vyas and Kumaranayake, 2006). PCA was done for nine variables among 82 drug naïve patients (table 18). All but one patient had data for all nine variables. This patient was eliminated from the analysis. Also four variables that had a number of missing patient information were also eliminated. Twenty-five percent of variants in all nine variables assessed within each patient explained influence to HIV-1 disease progression as shown by component 1 in table 18. Variables that stood out as dominant contributors to HIV-1 disease progression in PCA in descending order were the HIV drug resistance status of a patient (either susceptible or resistant to HIV-1) and the number of DRMs in the viral genome of an infected patient, $CD4^+$ T-cell counts and the HIV-1 subtype, HIV-1 IC DNA load and gender, secondary parasitic or bacterial infection status of an infected patient, plasma viral load and age (table 18). The values contributing to diseases progression according to PCA are in the range +1 to -1, where proportionality increases or value of contribution increases as the variable gets closer to +1 and decreases when it approaches -1.

**Table 18: Variable contribution to disease progression**

| Variable | Component 1 (25%) |
|---|---|
| Drug resistance status | 0.6 |
| Age | -0.1 |
| Gender | 0.2 |
| HIV-1 IC DNA copies per 1 million cells | 0.2 |
| CD4+T –cell counts | 0.3 |
| Number of mutations | 0.6 |
| HIV-1 subtype | 0.3 |
| Secondary infection status | 0.03 |
| Plasma viral load/ml of plasma | -0.1 |

# 4 Discussion

One important area of challenge in the fight against HIV is the development of drug resistant mutations and their transmission, which can undermine effective drug treatment (Ndembi et al., 2011; Price et al., 2011b). Also, in clinical practice, new methods in monitoring disease progress are needed in order to complement the already existing ones to the ends of predicting and improving patient outcomes. These challenges are particularly relevant in sub-Saharan Africa since most of the world's HIV infected people are located there.

Using samples from HIV infected patients from Nouna, Burkina Faso in West-Africa, this study analyzed HIV-1 DRMs that are found in paired RNA and DNA templates of HIV-1 infected drug naïve patients and also assessed DRMs that can be found in RNA templates of follow-up patients who harboured persistently high plasma viral loads after 24 weeks of HAART. The determination of DRMs was performed with Sanger or bulk sequencing and selected samples having discrepancies in resistance mutations or in paired nucleotide sequences of RNA and DNA were resolved with deep sequencing. Furthermore, results obtained from Sanger sequencing were then compared with deep sequencing results. The second part of the study focused on HIV-1 IC DNA levels among HIV-1 infected drug naïve individuals harbouring resistant or susceptible strains of HIV and discusses how HIV-1 IC DNA correlates with known markers of HIV disease progression. Most studies that have assessed HIV-1 IC DNA load in terms of its correlation with known markers of disease progression have been carried out in non-African populations and the few studies carried out on drug naïve African populations are mainly based on DRM surveillance. This study therefore looked jointly at HIV-1 DRMs and the role of HIV-1 IC DNA levels in relation to the traditional markers of disease progression such as viral load and CD4 T-cell counts in predicting disease among HIV-1 infected African patients.

## 4.1 Patient demographics and clinical characteristics

A total of 86 HIV infected patients were recruited for this study. There were 61 females and 24 males making 61% and 28% of the population respectively. In areas where

heterosexual transmission of HIV predominates, like in sub-Saharan Africa, it is known that there are more females living with HIV infection than males with females making up to 58% of all infected cases in the population (Garcia-Calleja et al., 2006; UNAIDS, 2014a). The male to female HIV infection ratio in this study (1:2.5) is in concordance with the general sex ratio of HIV in sub-Saharan Africa. However, Burkina Faso represents one of the few sub-Saharan African countries with a slight male preponderance in infections (Garcia-Calleja et al., 2006; UNAIDS, 2014a). In that regard, the sex ratio observed in this study may represent a local area with some characteristics different from the general population in Burkina Faso. The factors that may have resulted in a reversal of sex ratio in HIV infections in Nouna is a topic that may need to be looked into again in a population based study of infected persons, where the sample size is large. HIV infects people of all age groups, however majority of cases occur in adults over 15 years of age (UNAIDS, 2014a). In this study, the median age of infected persons was 35 years. The median age of infected females (34 years) was lower compared to males (42 years) and this is in conformity with regional trends reported by UNAIDS, where the median age of infected adult females are generally lower as compared to males UNAIDS (2014a) .

The median $\log_{10}$ transformed plasma viral load observed was 5.757 copies/ml of plasma (IQR 4.993 - 6.1640) and the median CD4 T-cell count was 199,986 copies/ ml of blood (IQR 120,504- 282,488). These values were higher than those observed among newly infected patients recruited for the Europe HIV resistance network study (Demetriou et al., 2010) where the median $\log_{10}$ transformed plasma viral load was 4.43 copies/ ml of plasma (IQR 3.87–5.14) and the median $CD4^+$ T-cell count was 423 copies/ ml of blood (IQR 255.3–578.5). The age ranges of infected persons in our study were comparable to that of the European study but the European study subjects were predominantly males whilst our study was predominated by females.

The CDC has categorized HIV infection into A, B and C based on $CD4^+$ T-cell counts and some clinical parameters (Schneider et al., 2008). Almost all the patients in this study were in the category B except one patient who was in the A category. It has been demonstrated that, levels of $CD4^+$ T-cell counts within which therapy is initiated correlates with the duration of $CD4^+$ T-cell recovery during treatment with ART (Okulicz et al., 2014). Furthermore, individuals who initiate treatment when $CD4^+$ T-cell count is <350 cells/µl do not reach levels

>500 cells/µl even after several years of treatment, although the viral load may reach below detection limit (Callegaro et al., 2014). An HIV-CAUSAL study consisting of 8,392 ART-naive patients who started ART when $CD4^+$ T-cell count had fallen to <350 cells/µl, demonstrated that patients in this category present with a greater risk of acquiring AIDS-defining illness or death than those patients initiating ART when $CD4^+$ T-cell counts are between 350 cells/µl and 500 cells/µl (Cain et al., 2011). In contrast to the aforementioned cohort, seventy-six patients (88%) in this study had $CD4^+$ T-cell counts of < 350 cells/µl of blood and ten patients (12%) had $CD4^+$ T-cell counts of ≥ 350 cells/µl of blood and no significant difference was observed between plasma viral levels of these two $CD4^+$ T-cell count divisions. This observation buttresses the point that in resource poor setting it is still a challenge to start ARTs when $CD4^+$ T-cell counts are between 350 cells/µl to 500 cells/µl since ARTs are most often inadequate and individuals start ARTs rather late as shown by the $CD4^+$ T-cell counts of the drug naïve patients in this study. This scenario is in contrast to the situation in developed countries where ARTs are readily available and started early to give the patient the maximum benefit from therapy (Menon, 2010).

Secondary bacterial or parasitic infections were recorded among 57% of the patients. The infections were mainly vulvovaginal candidiasis, one of the most common fungal infections noted to frequently recur in HIV-infected women (Ray et al., 2011). In a few instances, oral candidiasis and salmonellosis were also recorded. Studies have shown that oral candidiasis when left untreated in HIV infected patients, may often lead to morbidity. When these lesions are treated, the quality of life of the infected patient is maintained (Pienaar et al., 2010).

HIV co-infection with the hepatitis B virus (HBV) is endemic in sub-Saharan Africans (Mphahlele, 2008). There is a 17.4%- 21.3% seroprevalence rate of active HBV infection status observed in Burkina Faso (Nacro et al., 2001, Mulders et al., 2004). In Nouna, rural Burkina Faso where our study subjects came from and Ouagadougou urban Burkina Faso, the seroprevalence of both recent and past HBV infections are high (69.9% and 76.4% respectively) as seen in serological tests for HBV core antibodies while the reported seroprevalence for active HBV infections as indicated by HBV surface antigen tests are 14.3% and 17.3% respectively. The study reporting the seroprevalences of HBV also found an association between HIV and HBV core antibody status of infected subjects (Collenberg et al., 2006). In this study, among the 61 patients whose HBV sero-status were recorded, 6(10%) of the patients were reported to have

co-infection with HBV. Our data is however comparable to the general observation that up to 15% of HIV infected patients within sub-Saharan are estimated to be living with HBV (Hamers et al., 2013) but reports a lower HBV infection status than those previously observed in the general population in Burkina Faso (Nacro et al., 2001, Mulders et al., 2004) but again comparable with active HBV sero-status reported in Nouna (Collenberg et al., 2006). The report also agrees with the nationwide retrospective observational cohort study on HIV, HBV and HCV co-infection status among infected patients from the China national free ART program, where they observed an HBV seroprevalence of 8.7% among infected patients (Zhang et al., 2014).

Anemia is a common feature of HIV infection and has been correlated with disease progression (Sullivan, 2002). Indeed a prompt treatment of anemia in HIV patients has been shown to improve prognosis (CDC, 1992). Anemia is defined by the WHO as an Hb of less than 13.0 g/dl in men and 11.0 g/dl in non-pregnant women. In men, Hb in mild anemia ranges between 11-12.9 g/dl and moderate anemia ranges between 8-10.9 g/dl. In women, Hb in mild anemia ranges between 10.0-10.9 g/dl while moderate anemia lies between 7-9.9 g/dl (WHO, 2011). A study in South Africa reports up to 25% prevalence of anemia among HIV patients initiating HAART (Takuva et al., 2013). Meidani et al., 2012, reported a prevalence of as high as 67% anemia in a cohort of HIV patients (Meidani et al., 2012). In Burkina Faso, the prevalence of anemia among women of reproductive age is 58.6% (Meda et al., 1996). This prevalence is comparable to WHO's reported prevalence of 48.40% and 57.80% among non-pregnant and pregnant women respectively in 2011 (Stevens et al., 2013). In this study, the median Hb level was low 10.6 g/dl (IQR 9.7-11.8) indicating a high prevalence of anemia. Nearly all the patients fell in the range of mild to moderate anemia except one male who had a normal Hb. The prevalence of anemia reported in our study subjects was higher than trends reported in Burkina Faso. Such high prevalence of anemia may generally be a reflection of poor prognosis and is probably an indication of late initiation of treatment with HAART.

The mean body mass index (BMI) of women in rural Burkina Faso has been reported as 20.5 $kg/m^2$ (Savy et al., 2007). HIV infection has been associated with weight loss (WHO, 2007). The BMI is therefore an important factor in HIV infected patients. Indeed low BMI at the time of diagnosis has been shown to be a strong independent predictor of poor prognosis in HIV

infected patients prior to treatment (Maas et al., 1998; van der Sande et al., 2004). The median BMI of all subjects in this study was 17.72 kg/m$^2$ reflecting a possible poor prognosis.

## 4.2 Transmitted drug resistance mutations in HIV infected patients in Nouna, Burkina Faso

The expanding use of ARVs in HIV treatment globally parallels the increase in DRMs and TDRMs. A recent pooled prevalence of HIV drug resistance mutations in Africa was 10.6% with wide regional differences (Ssemwanga et al., 2014). Burkina Faso like most African countries has seen an increase in HAART in recent times and as expected, drug resistance and TDRMs have been reported (Tebit et al., 2006; Tebit et al., 2008). A previous study reported prevalence rates of DRMs as high as 40% for PI, 76% for NNRTIs and 85% for NRTIs among patients failing therapy in Ouagadougou (Tebit et al.). Prevalence of primary mutations in the general population of Burkina Faso was however 12.5% (Tebit et al.). The HAART drugs prescribed in Burkina Faso are two NRTIs and one NNRT1. In this study, the entire protease and two-thirds of reverse transcriptase in the pol region of HIV-1 were genotyped and sequenced because the drugs used for HIV management target these HIV enzymes. These HIV enzymes are also the targets of most antiretroviral drugs (Kantor and Katzenstein, 2003).

In assessing drug resistant mutations, plasma viral RNA or proviral DNA can be used. Given that the infective form of the virus (the viral RNA), is found in plasma, it is widely used for resistance mutation assessment (Saracino et al., 2008). However the use of proviral DNA for resistance mutation assessment may be relatively easier to carry out in resource limited countries, though the sequence profile may differ between RNA and DNA. Furthermore, profiles for interpreting HIV-1 drug resistance are generally formulated for HIV-1 subtype B, since this is the HIV subtype predominantly found in resource rich countries. However, evidence suggests that a similar set of resistance mutations are seen in non-B HIV-1 subtypes (Grossman et al., 2004; Kantor et al., 2005; Soares et al., 2007). Also alterations in amino acid sequences in the PR and RT regions of HIV-1 that result from genetic diversity in the viral genome affect drug resistance levels (Sanches et al., 2007; Velazquez-Campoy et al., 2001). In HIV-1 subtype B viruses, some minor or accessory mutations selected for under treatment with HAART are found as natural polymorphisms in HIV-1 non-B subtype viruses. (Holguin et al., 2002; Johnson et al.,

2011; Kantor and Katzenstein, 2003; Pieniazek et al., 2000). An example is the K20I resistance mutation that is a minor or accessory mutation selected for in HIV-1 subtype B viruses but occur as a natural polymorphism in HIV-1 non-B subtypes. This mutation cause reduced viral susceptibility and may also compensate for defects in viral fitness in subtype B HIV-1 viruses. In our study, the K20I polymorphism was found in almost all (98%) viral genomes sequenced. The K20I mutation in subtype B viruses have been identified as causing resistance to the protease inhibitor nelfinavir (Rhee et al., 2003). Evidence exist that natural polymorphisms to PIs in viruses of non-B subtype origin result in hypersusceptibility to PIs and increased viral fitness. (Santos et al., 2012).

In West Africa, the predominant HIV subtype has been reported to be CRF02_AG (Hemelaar et al., 2011; Nii-Trebi et al., 2013; Tebit et al., 2006). In our study, we identified 70% of the patients as being infected with CRF02_AG, followed by CRF06_cpx -like viruses, CRF01_AE-like, G-like and 02_A1-like subtypes in agreement with earlier work done in Burkina Faso and the West African sub region (Nii-Trebi et al., 2013; Tebit et al., 2006).

Sanger sequencing was used to assess the mutations in viral RNA and DNA among our patients. This platform detects mutations that occur in $\geq$ 20% of the total viral population of virus genomes while less frequent mutations < 20% are not detected with this method (Adje et al., 2001; Chabria et al., 2014). All patient samples were sequenced for DRMs using Sanger sequencing and targeting the pol region of HIV-1. Due to high sequence heterogeneity in 3 samples, analysis was only reported for 83 patients. The resistance mutations were classified as major or minor. Major DRMs on their own confer resistance to an ARV while minor DRMs do not (Wensing et al., 2014). A high rate of TDRM exists in drug naïve HIV-1 infected patients in Nouna, Burkina Faso

Among HIV-1 infected drug naïve patients in this study, TDRMs known to affect drug susceptibility prior to HAART was found in 9% of the patients. Among patients harbouring TDRMs only, most (87.5%) of the mutations affect susceptibility to RTIs and 12.5% to PIs. Furthermore, TDRMs affecting the action of NNRTIs formed the bulk (86%) of mutations affecting susceptibility to RTIs. Mutations affecting drug susceptibility were thus highest for NNRTIs followed by both NRTIs and PIs. Also, many minor mutations were detected that did not confer resistance to any antiretroviral drug in particular but could in combination with major

DRMs cause resistance. Generally, the frequency of occurrence of TDRMs to specific HIV drugs among HIV-1 infected drug naïve sub-Saharan Africans vary significantly between geographical regions (Price et al., 2011a; Bartolo et al., 2014; Tshabalala et al., 2011). The WHO reports an increasing trend for resistance against NNRTIs (WHO, 2012). A study done in Burkina Faso in 2006 also showed more resistance mutations affecting NNRTIs compared to PIs and NRTIs in drug naïve patients (Vergne et al., 2006) while a Malian study had a higher prevalence of mutations conferring resistance to PIs (Derache et al., 2008). In our study, there were more resistance mutations to NNRTIs compared to NRTs and PIs which reflects the trends observed by the WHO (WHO, 2012).

The WHO reports an average prevalence of 6.6% (CI 5.1-8.3) for TDRMs. However, various prevalence rates have been reported around Africa. In East and Southern Africa, an average prevalence of 5% was reported (Price et al., 2011b). Also recent studies in drug naïve pregnant women in Zimbabwe reported less than 5% prevalence in resistance mutation (Tshabalala et al., 2011). A similar trend of TDRMs was observed in Angola (Bartolo et al., 2014). Our observation of 9% prevalence is relatively high although a rate of 11.5% has been reported in Mali (Derache et al., 2008). Given the fact that the sample size in this study was small, trends of TDRMs have to be studied in a larger group of drug naïve patients. It should also be noted drug naïve patients in this study were assumed not to have taken antiretroviral drugs but this is not proven and therefore one should also be open to the possibility of some DRMs found in this group, being mutations that might have been selected for under prior drug use. Given the high prevalence of DRMs in our study, it is worthwhile to recommend that a patient's HIV resistance mutation profile should be known before the initiation of therapy. Resistance monitoring should also be incorporated into clinical practice in developing countries, for effective clinical management of HIV.

### 4.2.1 Patterns of the emergence of drug resistance mutations and its persistence or disappearance among HIV-1 infected patients on HAART

The current stage in the HIV pandemic poses the dual problem of development of DRMs during treatment. We further assessed sequences of follow-up patients who maintained their viral load levels at baseline or those who had higher viral load while under HAART. This group of patients were assessed under the assumption that they might have developed new or additional DRMs and thus escaped therapy. Therapy failure was determined using longitudinal plasma RNA load levels of infected patients from baseline to 24 weeks into treatment. We selected eight follow-up patients based on the aforementioned assessment and the availability of plasma samples with viral load above 1,000 copies/ml of plasma since this enabled us to successfully genotype and sequence patients RNA templates. All selected follow-up patients were assessed for DRMs development over time.

Highly active antiretroviral therapy exerts selection pressure on HIV quasispecies in an individual and that selects for HIV resistant strains during treatment (Metzner et al., 2009). If active HIV replication ensues in the presence of ARDs at a certain threshold, viral populations may evolve that select for critical DRMs to antiretroviral regimes administered and may render antiretroviral drugs ineffective. This process is termed the genetic barrier to resistance. Genetic barrier to resistance could also be defined as the threshold beyond which mutations that reduce viral susceptibility to ARDs develop. Factors that influence the genetic barrier to resistance are numerous among which are the number of critical mutations needed to cause resistance (it should be noted that a single mutation may be enough), the number of preexisting mutations and their replicative fitness. Some mutations influence drug susceptibility more than others. There exist ARTs with either high genetic barrier to resistance (ARDs of thymidine based origin and boosted PI drugs) or low genetic barrier to resistance (triple HAART regimes of NNRTIs and NRTIs origin as used by our study subjects). For drugs with high genetic barrier to resistance, a high number of DRMs are needed to render ARDs ineffective but drugs with low genetic barrier to resistance only require a few mutations to make ARDs ineffective (Eron, 1996; Gallant et al., 2003; Kempf et al., 2004; Kessler, 2005).

The major DRMs detected among the study subjects were those that conferred resistance to NNRTIs and NRTIs, reflecting that HAART regimes administered to patients in rural Burkina Faso have low genetic barrier to resistance and may only require a single or a few mutations to cause resistance to antiretroviral regimes being used. Major NNRTI mutations detected in ART naïve patients included K103N and Y181C mutations. These made up a small proportion (5%) of our patients. Tebit et al., 2009, reported a similar percentage (6.1%) of NNRTI mutations in drug naïve patients in Ouagadougou (Tebit et al., 2009). The neverapine related resistance mutations K103N and Y181C have been widely reported in sub-Saharan Africans undergoing treatment with NNRTIs because of its wide use in the treatment of mother-child-transmissions (Tebit et al., 2008) as well as in drug naïve patients starting therapy (Akinsete et al., 2004; Tebit et al., 2009). The Y181C and K103N mutations have also been reported in drug naïve (Magiorkinis et al., 2008) and treated individuals (Tebit et al., 2009). The NNRTI based DRMs found in our study subjects confer broad based resistance to all known NNRTIs and since they have a low genetic barrier to resistance, are selected more frequently.

Of the eight follow-up patients for whom DRMs were analyzed, the K103N was the most common major mutation detected after 24 weeks of HAART in 3 (38%) of patients, 48 weeks in 4 (50%) patients including a patient who developed the mutation at 24 weeks under HAART and another who had the mutation already at the drug naive period. In each case, the mutation then remained till 48 or 72 weeks depending on the time of the last analysis. The predominance of K103N mutation in the patients followed over time is in consonance with work done by Tebit et al., 2008, they showed that 44% of major resistance mutations found in HAART treated patients in Ouagadougou, Burkina Faso are K103N mutations (Tebit et al., 2008). Moreover, the dominance of the K103N mutation seen among drug naïve and follow-up patients might be explained by not only the low genetic barrier of resistance in the viral population due to the use of NNRTIs but also the added advantage that the mutation does not confer a reduced fitness cost and for that matter behaves like the wildtype virus. This advantage helps in facilitating the replication and emergence of new viral genomes harbouring the K103N mutation both in the presence and in the absence of ARDs. These mutated viral genomes may then be present in the viral quasispecies in equal proportions as wildtype viruses. Another possibility for the dominance of the K103N mutation among our patients might be the selection of minority

viral populations within the viral quasispecies that harbour this mutation. These mutant viruses may gain the advantage of replicating to levels proportional to wildtype viruses within the growing viral population in the patient (Mackie, 2006). The other major DRMs to NNRTIs observed in the follow up patients were the E138Q, G109A, Y181C and M230L mutations. Most of these mutations, in contrast to K103N, were infrequent. These infrequent mutations have reduced fitness thereby being unable to compete with the K103N mutation in the viral population. (Mackie, 2006). Given that the K103N mutation is by far the most observed, efforts to determine its presence before and during treatment is recommended.

### 4.2.2 Multiple drug resistance mutations may lead to virologic failure in follow-up patients on HAART

Virologic failure rates under HAART among Africans and Asians have been described as varied, ranging from a prevalence of 2.9% to 20.6% (Aghokeng et al., 2014). Other studies in certain African countries have seen rates as high as 35% to 50% (Barennes et al., 2014). In urban Burkina Faso, the prevalence of DRMs in patients with virologic failure have been studied with high prevalences reported per ARV use as follows; NRTIs 85%, NNRTIs 76%, and PIs 40% (Tebit et al., 2008). In our study 6 (75%) of the eight follow-up patients on HAART from rural Burkina Faso had multiple DRMs to NRTIs and NNRTIs, which could explain virologic failure. Several studies have also identified multiple DRMs as a common cause to therapy failure among patients on ART (Daar and Richman, 2005; Jiamsakul et al., 2014). The trend observed in our study, was similar to trends reported in previous studies conducted among 209 patients with virologic failure under HAART in urban Burkina Faso (Tebit et al., 2008). However, the sample size was small and therefore the observations need to be interpreted with caution. The two remaining follow-up patients 2(25%) with high viral loads (greater than 1,000 copies/ml of plasma) suggestive of therapy failure did not have DRMs known to significantly reduce drug susceptibility. In one of the two follow-up patients, the reason for virologic failure was the interruption of treatment for 14 days. The reason for therapy failure in the other patient was not known. It should therefore be noted that virologic failure under treatment could initially be due to reasons other than resistance mutations but may lead to future development of DRMs that render ARTs ineffective at a later time. Although the patients were indicated to have been on

HAART, it is possible that additional factors such as inadequate therapy or non-compliance to HAART regimes, drug rationing, frequent treatment interruptions and not only DRMs led to persistently high plasma viral RNA levels. The development of DRMs therefore makes it necessary that adequate monitoring of patients is done both for adherence to therapy and determination of emergence of resistance mutations.

Disease progression in HIV patients is dependent on many factors including the development of DRMs during treatment. Major TDRMs among drug naïve patients may have led to the emergence of multiple drug resistance mutations at latter time points. Major TDRMs are pre-existing DRMs that are found in HIV infected drug naïve patients. These mutations reduce the susceptibility of ARDs substantially on their own while minor mutations reduce drug susceptibility minimally or substantially in combination with major DRMs (Wensing et al., 2014). In our eight follow-up patients who showed signs of therapy failure indicated by high viral loads, two patients had major TDRMs to NNRTIs at baseline while the remaining six patients either had minor resistance mutations to RTIs or no mutations at baseline. Of the two patients with major DRMs to NNRTIs at baseline, one had the K103N mutation and failed HAART within 24 weeks while the other had the Y181C mutation and failed treatment within 12 weeks. These mutations are known to be selected for among patients under HAART in sub-Saharan Africa due to the NNRTI based HAART regimes given (Tebit et al., 2008). These two patients also developed multi- DRMs at 12 weeks and 24 weeks under HAART. It has been shown that the number of TDRMs at baseline may increase the risk of virologic failure (Daar and Richman, 2005; Jiamsakul et al., 2014). The two patients with major TDRMs at baseline failed therapy, giving an indication that such patients indeed run the risk of virologic failure although our sample size was small. The other six follow-up patients who experienced virologic failure but only had minor or no baseline mutations also failed therapy within the same duration of 12 to 24 weeks as in the case of those patients who had TDRMs. Two of these patients developed multi drug resistance mutations within 24 weeks into therapy. This observation indicates that patients could experience virologic failure with or without pre-existing baseline mutations and that other factors not limited to TDRMs may increase the risk of virologic failure as well, therefore proper patient monitoring and education are paramount to disease management.

Evidence of major multi- DRMs can be highlighted in four (50%) of the eight follow-up patients. They had the major NRTI resistance mutation M184V emerging together with other major NNRTI DRMs like the K103N, Y181C, E138Q, G109A, and M230L. In two of the four patients the M184V mutation appeared at 24 weeks into therapy and persisted with multiple NNRTI mutations. In the remaining two patients, one had the mutation at 12 weeks and then it disappeared at 64 weeks with all the other multiple mutations while the other had the mutation at 36 weeks and lost it at 48 weeks while a new mutation to NNRTIs emerged with pre-existing NNRTI mutations still present. The M184V mutation confers high-level resistance to 3TC and FTC and low level resistance to ABC and ddI (Turner et al., 2003) and has also been associated with reduced viral fitness (Turner et al., 2003). It is not clear why the mutation would disappear in the patient, however rapid disappearance of M184V mutations has been associated cessation of therapy (Mackie 2006; http://www.ncbi.nlm.nih.gov/books/NBK2249/). Stopping therapy will remove the pressure selecting for the M184V mutation, which has a fitness disadvantage in the absence of the drugs (Mackie 2006; http://www.ncbi.nlm.nih.gov/books/NBK2249/). Data on the patient does not indicate a termination of treatment, however it cannot be ruled out. The two other patients had minor mutations and within 24 and 48 weeks developed major resistance mutations to the action of NVP and EVF, till 72 weeks under HAART. Again although the sample size was small, the evidence leads to the fact that DRMs to NNRTI and NRTI based HAART are on the rise and HIV treatment plans for sub-Saharan Africa need alternatives drugs in combination with RTIs to reduce resistance.

### 4.2.3 A high concordance exists between HIV-1 RNA and proviral DNA in drug naïve patients in Nouna Burkina Faso

HIV provirus integrated into the genome serves as viral reservoirs and ensures viral persistence even with undetectable viral RNA load during therapy. While a lot of focus has been put on assessing viral RNA mutations for clinical decision-making, it is still unclear what the benefits of assessing proviral DNA mutations could be (Chew et al., 2005). Potential benefits of assessing mutations in viral DNA however exist. First, if the determination of viral DNA mutations can predict effectiveness of HAART to a similar extent in a given population as viral RNA mutations, the use of DNA can be implemented in less resourced regions due to ease of

amplification and the stability of the template (Banks et al., 2012). Also, where RNA viral loads are very low, proviral DNA may give some information on potential resistance in the future (Banks et al., 2012; Chew et al., 2005).

Assessing viral DNA as a template for resistance testing is an evolving concept and existing data on the topic is quite limited, especially data comparing viral RNA and DNA templates of HIV infected African patients in terms of DRMs. However one study has been done on in Zimbabwe which indicates the potential usefulness of viral DNA (Banks et al., 2012). Studies involving 253 newly-diagnosed HIV infected patients from the Europe HIV resistance network compared paired viral RNA and DNA derived PR/RT sequences from each of their patients for HIV-1 DRMs that confers resistance to inhibitors of PR and RT and found a high concordance between the two templates in regions assessed for DRMs (Demetriou et al., 2010). We therefore assessed the level of concordance or discordance between paired RNA and DNA of drug naïve patient derived HIV-1 PR/RT sequences for DRMs at regions that conferred resistance to PR/RT based ART in each of our patients. On comparing mutations in viral RNA to DNA we observed a high level of concordance (94%) and a low level of discordance (6%) among drug naïve patients. There were more resistance mutations seen in RNA compared to DNA in the discordant group. Regarding the disparities seen in paired viral RNA and DNA sequences, studies have shown diverse results. Some studies report a higher number of mutations for RNA compared to DNA while others report the opposite (Bon et al., 2007; Saracino et al., 2008). Again, existing studies on paired HIV-1 RNA and DNA sequence comparisons reveal a wide range of results (Bon et al., 2007; Saracino et al., 2008). While some studies show no significant differences between the two sequences and suggest DNA as an alternative to RNA (Derache et al., 2015; Vicenti et al., 2007), others have demonstrated wide disparities that might discourage the use of DNA as an alternative template (Chew et al., 2005; Smith et al., 1993). One study reported as high as 87 % discordance in RNA and DNA templates of the same patient, in ARV treated individuals (Saracino et al., 2008). Bon et al., 2007, reported a discordant rate of 26% in drug naïve patients (Bon et al., 2007). There are also further disagreements on whether detected DNA mutations can be used in clinical assessments instead of RNA or whether DNA should complement RNA in clinical decision-making (Banks et al., 2012; Chew et al., 2005).

However, discordant sequence pairs of RNA and DNA in terms of DRMs found in our study was low.

There are several reasons why discordance can occur between viral DNA and RNA. While the viral DNA is stably incorporated into the genome and is rarely altered, DNA dependent RNA polymerase II that transcribes the viral RNA can induce errors during transcription (Abram et al., 2010) and as a result, many mutations can be incorporated in viral RNA that would not be present in the viral DNA. This could explain why more mutations may be seen in the viral RNA than DNA. On the other hand, while resistant mutations can appear in RNA and disappear over time due to the changing environment of the infected patient, the mutations could be archived in the form of proviral DNA which can be detected at a later time when the RNA population has changed (Banks et al., 2012). Archiving of various resistant mutations in DNA over time may result in the viral DNA having more mutations than RNA. Finally, the high error-prone reverse transcriptase enzyme can cause mutations in the reverse transcribed viral RNA that would be archived in proviral DNA. These mutations may not be present in the RNA population depending on the transcription status of the DNA. Therefore, depending on the unique circumstance of the patient, some degree of discordance may be seen in mutations when comparing RNA and DNA. Some degree of discordance between RNA and DNA mutations is therefore expected in HIV infected patients.

DRMs in viral DNA are a reflection of previous and current resistant mutation states of viral RNA. It has been shown that detection of DRMs in viral RNA precedes DNA and as a consequence, mutations are detected in DNA latter in relation to RNA, however established mutations will eventually be reflected in DNA (Banks et al., 2012). Detection of DRMs in DNA should therefore be interpreted with the view that not all the DRMs detected at a given time point will be clinically relevant for immediate management decisions as some DRMs in DNA may indicate previous resistance states. However DNA DRMs may help in better understanding of evolution of the viral quasispecies and could also be used to predict potential resistance in the future. Also, DNA DRMs could provide useful clinical information in the absence of facilities for viral RNA studies but should be interpreted with caution (Derache et al., 2015).

In our patients, the high concordance rate between paired RNA and DNA samples in terms of HIV-1 DRMs suggests that DNA could be substituted for RNA in making clinical

decisions on drug resistance. A larger study would be necessary to determine the general likelihood of concordance between DNA and RNA and protocols could be developed to optimize interpretation of DNA data to closely reflect the prevailing RNA resistance mutations in a patient.

### 4.2.4 In-depth analysis of RNA DNA discordance using deep sequencing

Determination of DRMs using bulk sequencing is limited by the fact that only mutations that occur ≥ 20% of the viral quasi species are detected (Chabria et al.). Bulk sequencing therefore excludes a significant proportion of the viral population, which can impact disease progression. In order to obtain a more in depth assessment of viral populations in patients who had discordances in mutations between RNA and DNA or had ≥ 10 nucleotide differences in paired RNA and DNA sequences in Sanger sequencing, we resorted to use deep sequencing since it detects both mutations that occur in ≥ 20% of the viral population (comparable to Sanger sequencing) and mutations in less than 20% of the viral population referred to as minority viral populations which Sanger Sequencing does not detect. Deep sequencing therefore allowed analysis of viral RNA and DNA sequences at greater depth and therefore could be used to detect whether discordance possibly represented minority mutations.

Our deep sequencing method detected mutations occurring in as low as 1% of the viral population. We assessed for discordance in mutations between RNA and DNA in the range above 20% frequency. We also compared mutations in RNA and DNA in the range below 20% (minority mutations), which is below the range of bulk sequencing detection. These comparisons were done in order to determine whether discordances in RNA and DNA mutations vary or not and also determine whether mutations found in deep sequencing above the 20% mutation frequency reflects that seen in bulk or Sanger sequencing.

There was a high concordance (95%) for resistance mutations between RNA and DNA with deep sequencing in populations ≥ 20%, similar to what was observed in bulk sequencing. However there was a high level of discordance (74%) between RNA and DNA in the minority population. It is known that mutations are detected earlier in HIV viral RNA while mutations in proviral DNA are detected at a later time (Banks et al., 2012; Palmisano et al., 2009). Our data

suggests that mutations seen in RNA are also seen in DNA within the majority or dominant viral population and this strengthens the argument that DNA mutations picked up by Sanger sequencing mostly reflect the RNA mutant population hence either DNA or RNA templates could be used in genotyping and sequencing for the determination of DRMs routinely.

### 4.2.5 Non Synonymous mutations predominate in the viral populations of HIV infected patients in Nouna Burkina Faso

The nature of mutations in HIV impacts the behavior of the virus. Nucleotide changes that do not result in changes in amino acid are referred to as synonymous mutations while those that result in a change in the amino acid are non-synonymous mutations (Hunt et al., 2014). Non-synonymous mutations lead to changes in the primary structure of proteins and therefore could impact its function. In HIV, non-synonymous mutations provide mechanisms for escape of the virus by altering epitopes leading to evasion of immune cells and neutralizing antibodies (Zanini and Neher, 2013). Traditionally, synonymous mutations however have been considered inert, but this view has changed in recent times (Hunt et al., 2014). Synonymous mutations can alter transcriptional and translational processes and affect the stability of RNA. In HIV, it has been shown that synonymous mutations in the 3 prime region of the pol gene can significantly alter viral protein translation and viral replication rates (Nomaguchi et al., 2014).

In our study, paired sequences of RNA and DNA templates from HIV-1 infected drug naive patients were assessed for the presence of synonymous and non-synonymous nucleotide differences first by bulk or Sanger sequencing and then changes ≥ 10 were resolved with deep sequencing. We observed high levels of synonymous mutation rates as compared to non-synonymous mutations among patient sequences using the bulk sequencing method. The high synonymous mutation rates observed can be explained by the fact that synonymous mutations usually have a high fitness cost and will generally be selected for in the viral population (Zanini and Neher, 2013). Samples that were confirmed with deep sequencing showed a high non-synonymous mutation rate as compared to synonymous mutations both in bulk and all deep sequencing frequency cut-offs of 1, 2, 5, 10 and 20%. The high non-synonymous mutation rate implies that the viruses may be adapting to changing environments in patients in the face of drug

therapy and host immune effects. As stated earlier, synonymous mutations are generally selected for relative to non-synonymous mutations because synonymous mutations usually have a high fitness cost as compared to non- synonymous (Zanini and Neher, 2013). The high non-synonymous mutation rate among the study population is an indication of viral effort to overcome bottlenecks like drug treatment and host immune response and therefore strengthens the justification for the introduction of resistance monitoring.

## 4.3 The relationship between HIV-1 intracellular DNA and other markers of disease progression among HIV-1 infected Burkinabés

In the current clinical management of HIV, plasma viral load and $CD4^+$ T-cell counts are used for disease monitoring and they form the basis for decision-making on drug therapy and alteration of treatment regimes (Langford et al., 2007). Although viral load and $CD4^+$ T-cell counts have been useful clinically, there is still room for development of other disease monitoring tools in the face of increasing complexity of drug treatment and resistance mutations. Also, the development of new, easily available and cheaper disease monitoring tools can complement management of HIV in resource-limited settings. In recent times HIV-1 IC DNA, in contrast to plasma RNA of HIV, is becoming of interest to clinicians and researchers as a disease monitoring tool in clinical management of HIV. The HIV-1 IC viral DNA also serves as reservoir for the virus (Demetriou et al., 2010). Recent studies have shown intracellular DNA as an independent determinant of disease progression (Demetriou et al., 2010) and further assessment of this finding is necessary among sub Saharan Africans who may benefit from the use of HIV-1 IC DNA in disease monitoring.

In this study, we assessed HIV-1 IC DNA levels in HIV patients from Nouna, Burkina Faso and determined how they correlated with plasma viral load and $CD4^+$ T-cell counts as well as other surrogate markers of disease progression. We also assessed HIV-1 IC DNA loads in terms of comparing levels in patients with drug resistant strains and non-resistant strains and determined how they compared with established standards of disease monitoring like $CD4^+$ T-cell counts and plasma viral load. We also analyzed HIV-1 IC DNA loads in terms of patients

who experienced virologic failure after 24 weeks of HAART and those who experienced virologic suppression within the same time frame.

### 4.3.1 HIV-1 intracellular DNA levels ranged between 10 and 10,000 copies per one million cells among drug naïve patients.

HIV-1 IC DNA was measured by a molecular-beacon-based qPCR technique using CCR5 gene copies as a reference to calculate the number of cells in buffy coat. The majority of the patients (58%) had HIV-1 IC DNA loads between 100 and 1,000 copies/$10^6$ cells, 26% had HIV-1 IC DNA between 1,000 and 10,000 copies/$10^6$ cells, 14% were between 10 and 100 copies/$10^6$ cells while 2% were between 10,000 and 100,000 copies/$10^6$ cells and none fell in the range of 100,000 to 1,000,000 copies/$10^6$ cells.

HIV-1 IC DNA viral load has been shown to be relatively stable as compared to the plasma viral load especially during treatment (Saitoh et al., 2002). There is a paucity of studies that have analyzed the level of HIV-1 IC DNA among Africans. In studies done in developed countries, Kostrikis et al., 2002, demonstrated in an European cohort a median HIV-1 IC DNA level of 2.646 copies/$10^6$ PBMCs at baseline (Kostrikis et al., 2002a). Our median HIV-1 IC DNA level was in a range 2.645 (IQR 2.260 – 3.055), almost identical to that observed by Kostrikis et al., 2002.

### 4.3.2 HIV-1 intracellular DNA does not correlate with primary markers of disease progression in drug naïve HIV-1 infected Burkinabés.

In order to assess the utility of HIV-1 IC DNA as a potential substitute to established markers of disease progression like $CD4^+$ T-cell counts and plasma viral load, we assessed correlations between HIV-1 IC DNA on one hand with plasma viral load and $CD4^+$ T-cell counts on the other hand. Our results showed no correlation between cellular viral DNA and $CD4^+$ T-cell counts or plasma viral load using all patient samples (n=83). The correlation between HIV-1 IC DNA levels and $CD4^+$ T-cell counts of CDC clinical stage B stratified populations of drug naïve patients was also assessed but again no association was observed.

Reports on the association between HIV-1 IC DNA loads, plasma viral load and $CD4^+$ T-cell counts vary widely. While some studies have reported no significant correlations in drug naïve patients, others have reported significant associations especially in treated patients (Buzon et al., 2014). Poizot-Martin et al., 2013, reported no correlation between HIV-1 IC DNA and plasma viral load as well as HIV-1 IC DNA and $CD4^+$ T-cell counts, in a cohort of HIV-1 infected patients (Poizot-Martin et al., 2013). This was in agreement with our results. Furthermore, in a study conducted by Kostrikis et al., 2002, it was shown that HIV-1 IC DNA levels moderately correlated with plasma viral load and weakly correlated with $CD4^+$ T-cell counts (Kostrikis et al., 2002a). Watanabe et al., 2011, however reported a negative correlation between absolute HIV-1 IC DNA levels and $CD4^+$ T-cell counts but the correlation was not observed when relative HIV IC DNA quantifications were used (Watanabe et al., 2011). In this study, they defined the assessment of HIV-1 IC DNA load relative to the number of PBMCs as the 'relative amounts' in contrast to 'absolute amounts' when HIV-1 IC DNA was expressed as per milliliter of blood (Watanabe et al., 2011).

Our cross sectional data may also reflect treatment outcomes in our patients since quantifications of HIV-1 IC DNA levels in treated patients have been shown not to change significantly over time (Cone et al., 1998; Kostrikis et al., 2002b) but may rather be concordant with the time of treatment initiation and not the duration of treatment (Watanabe et al., 2011).

### 4.3.3 HIV-1 DNA levels are significantly higher in patients who experience virologic failure than those who do not.

Our studies demonstrated a significantly higher ($p=0.045$) median HIV-1 IC DNA level at baseline (median 3.16 $\log_{10}$ copies/$10^6$ cells, IQR 2.75-3.62) in patients who experienced virologic failure after 24 weeks of HAART as compared to their counterparts in whom virologic suppression was attained (the median baseline HIV-1 IC DNA was 2.63 $\log_{10}$copies/$10^6$ cells, corresponding to the range IQR, 2.12-3.04 $\log_{10}$copies/$10^6$ cells). This observation is in agreement with studies conducted by Saitoh et al., 2007, who demonstrated that HIV-1 IC DNA is lower in HAART treated children with undetectable RNA viral loads compared to those with high viral loads (Saitoh et al., 2002) also signifying that HIV-1C DNA load could be used to

predict disease progression at baseline before therapy commencement. A study undertaken by Parisi et al., 2012, demonstrated that low baseline levels of HIV-1 IC DNA load leads to undetectable levels of the cellular DNA load and plasma viral load levels during therapy (Parisi et al., 2012). Williams et al., 2014, stated from their studies that, cellular HIV-1 DNA is a better predictor of disease outcome than plasma viral load and could even help to identify individuals who can safely interrupt therapy (Williams et al., 2014). Our findings need to be reconfirmed by a larger sample size.

Plasma viral levels at 24 weeks post HAART was higher in patients who experienced virologic failure than those who did not. There was approximately $2.5\log_{10}$ difference ($p<0.001$) in plasma viral load levels between patients who experienced virologic failure and those who experienced virologic suppression at 24 weeks of treatment but there was no difference in their $CD4^+$ T-cell count levels ($p=0.15$). Also both their plasma viral load and $CD4^+$ T-cell counts at baseline were not significantly different. Our findings suggest that plasma viral load remains a strong predictor for disease prognosis and not $CD4^+$ T-cell counts (Mellors et al., 1996).

### 4.3.4 HIV-1 DNA levels are unaffected by drug resistance mutations in drug naïve HIV-1 infected patients

Not much has been studied on the factors that determine the level of HIV-1 IC DNA especially among African populations. Transmitted drug resistance mutations alter the kinetics of viral infection and may affect levels in HIV-1 IC DNA in HIV infected patients. In a large European cohort however, it was determined that TDRMs do not affect HIV-1 IC DNA (Demetriou et al., 2010). Little or no information currently exists on the effect of TDRMs on HIV-1IC DNA load in Africa. We compared the levels of HIV-1 IC DNA load in patients who had TDRMs and those who had no DRMs affecting drug susceptibility. Our data showed no significant difference between the median HIV-1 IC DNA in the drug sensitive and resistant groups of drug naïve patients assessed. Our results was similar to the observations made in the large European cohort studies (Demetriou et al., 2010).

### 4.3.5 Transmitted drug resistance mutations does not alter the levels of plasma viral load and CD4$^+$ T-cell counts among drug resistant and susceptible HIV infected Burkinabés

Our data indicated that cellular HIV-1 DNA levels were not affected by TDRMs. However we wanted to determine whether a similar pattern existed for the clinically used HIV-1 infection monitoring markers like CD4$^+$ T-cell counts and HIV plasma viral load. Again there was no difference in the levels of CD4$^+$ T-cell counts and plasma viral load in patients with and without TDRMs. Nevertheless we observed a trend towards higher CD4$^+$ T-cell counts among patients with TDRMs compared to those without (p=0.007). When we further segregated the cohort by CDC clinical stages and focused only on group B patients, we observed significantly higher CD4$^+$ T-cell counts among clinical stage B patients with TDRMs as compared to those without TDRMs. Studies have generally shown that mutant viruses tend to have reduced fitness and may be less virulent than wild type viruses and this may affect the degree of CD4$^+$ T-cell count reduction in patients (Hirsch et al., 2008). It has also been reported that individuals infected with drug resistant viruses have higher initial CD4$^+$ T-cell counts than those infected with wild type viruses although the high CD4$^+$ T-cell counts in TDRM groups are usually followed by a faster CD4$^+$ T-cell count decline compared to wild type (Hirsch et al., 2008; Pillay et al., 2006). Our data indicated a higher level of CD4$^+$ T-cell count in the TDRM group, however, analysis of later time point CD4$^+$ T-cell count values would give a clear assessment of CD4$^+$ T-cell count decline over time among the two groups of patients.

## 4.4 The effect of other indicators of disease progression on HIV-1 intracellular DNA, Plasma viral load and CD4$^+$ T-cell count in HIV-I infected patients

There are several factors that have been shown to influence HIV infection and disease progression among Africans. In this study we assessed the role of hepatitis (HBV) confection, predominant HIV subtypes, number of mutations and age on established markers of disease monitoring like CD4$^+$ T-cell counts, plasma viral load and HIV-1 IC DNA. Co-infection with

Hepatitis B did not alter the levels of HIV-1 intracellular DNA, plasma viral load or $CD4^+$ T-cell count among drug naïve patients. Our observation agrees with a meta-analysis done by Nikolopoulus et al., 2009, where they assessed the impact of HBV co-infection with HIV-1 on HIV disease progression and did not find any significant impact (Jain, 2009; Nikolopoulos et al., 2009). Also no correlation existed between HIV-1 intracellular DNA and age, HIV subtypes and number of mutations among infected patients, and this is in concordance with studies conducted by Kostrikis et al., 2002, who also reported the lack of association between HIV-1 IC DNA and age, HIV-1 subtypes and number of DRMs in their European cohort (Kostrikis et al., 2002b). Finally, principal component analysis of all the variables used to assess markers of disease progression in this study also showed that 25% of variations in all the variables examined explain HIV-1 disease progression in the study participants.

# 5 Conclusion

The current phase of the HIV pandemic brings with it hope for infected persons due to the effectiveness of HAART. However, challenges still lie ahead that require concerted efforts to overcome. The most important challenge involves the development of drug resistance mutations and their transmission that may threaten the effectiveness of HAART in the future. Also the need for improved clinical tools for monitoring disease states and progression and determination of drug resistance mutations is increasing, especially in low-income areas.

In this study we have shown that drug resistance mutations are quite common in drug naïve HIV-1 infected patients in Nouna Burkina Faso, indicating increasing transmission of resistant mutations in the African population. There is also a high rate of mutation development with treatment on HAART. The high similarity between DNA and RNA mutation rates found in this study indicates that the use of DNA can be adopted in drug resistance monitoring since it is very stable and would not easily be degraded with changes in temperature as compared to RNA which is easily degradable and needs specific temperature conditions to keep it viable for DRM testing. Deep sequencing also shows minority DRM variants that may become dominant over time or provoke the emergence of DRMs that reduce drug susceptibility. Drug resistance mutations in the minority quasispecies could also lower susceptibility to ARTs leading to virologic failure. Deep sequencing has on the other hand, proved to be a good tool that could be used to resolve disparities in DRMs and suspicious nucleotide differences. Finally, HIV-1 IC DNA in our study did not strongly correlate with key markers of disease progression but could be used to detect virologic failure at 24 weeks of HAART, signifying that HIV-1 IC DNA or HIV-1 STS DNA levels may predict disease progression independently and can be used either alone or together with plasma viral load and $CD4^+$ T-cell count to predict disease progression. Since our sample size was small, it is recommended that more studies be carried out to assess the utility of HIV-1 IC DNA in clinical practice.

# 6 Outlook

This study analyzed an alternative template (DNA) that is more stable and easy to handle, for HIV-1 drug resistance genotyping. It also shed light on the importance of detecting TDRMs before the initiation of ART, since this may play an important role in clinical decision-making and serve as a key to preventing the emergence of multiple drug resistance mutations at later time points in the progression of HIV-1 disease. A large-scale study involving deep sequencing of treatment naïve and follow-up patients harbouring DRMs should be evaluated to assess whether DRMs that occur over time were already present in the minority viral quasispecies. Such a study could be expanded to involve mutational linkages and the relevance of minority variants in clinical decision-making.

# APPENDIX

## Appendix I: Ethical clearance from Nouna Burkina Faso

MINISTERE DE LA SANTE
SECRETARIAT GENERAL
CENTRE DE RECHERCHE EN SANTE DE NOUNA
COMITE LOCAL D'ETHIQUE

BURKINA FASO
Unité – Progrès - Justice

Nouna, le 07 DEC 2007

**DECISION N°2007-.031./CLE/CRSN**

Le Comité Local d'Ethique du Centre de Recherche en Santé de Nouna, en sa séance du 08 novembre 2007, a reçu, examiné et délibéré sur le projet de recherche intitulé « ***Détermination du virus et des facteurs de risque associés à la transmission verticale du VIH à Nouna au Burkina Faso*** » du Prof Hans-Georg Kräusslich et du Dr Thomas Böhler, de l'Université de Heidelberg en Allemagne.

Tous les aspects éthiques relatifs à ce projet ont été examinés et le Comité approuve le projet tel que décrit dans le protocole.

Le Président

Boubakary TRAORE

# Appendix II: Ethical clearance from Heidelberg Germany

Medizinische Fakultät Heidelberg

Herrn PD Dr. med. habil. Th. Böhler
Hygieneinstitut der Universität
Abt. Virologie
Im Neuenheimer Feld 324
69120 Heidelberg

Datum: 21.01.2008
Seite

Votum

*Antrags-Nr.:* S-435/2007 (Bitte stets angeben!)
*Titel:* Monitoring der lymphozytären Immunrekonstitution von Erwachsenen mit HIV-1-Infektion unter antiretroviraler Kombinationstherapie (ART) in Burkina Faso
*Geprüfte Unterlagen:* Studienprotokoll vom 14.01.2008, Version 1.1 einschließlich Patienteninformation und Einverständniserklärung
CD-ROM
Schreiben vom 14.01.2008

Sehr geehrter Herr Dr. Böhler,

mit den Änderungen bzw. Ergänzungen im Studienprotokoll und in der Patienteninformation und Einverständniserklärung wurden die im Votum vom 12.12.2007 genannten Empfehlungen der Ethikkommission berücksichtigt. Damit liegt die Zustimmung der Kommission zu der o.g. Untersuchung vor.

Änderungen in Organisation und Ablauf der Studie sind der Kommission, zusammen mit einer Bewertung der Nutzen-Risiko-Relation, umgehend mitzuteilen. Sowohl die **Antragsnummer** als auch die **geänderten Passagen** sollten in den betreffenden Unterlagen **deutlich gekennzeichnet** sein, da anderenfalls keine zügige Bearbeitung möglich ist.

Die Ethikkommission der Medizinischen Fakultät Heidelberg arbeitet gemäß den nationalen gesetzlichen Bestimmungen und den ICH-GCP-Richtlinien. Ihren Beratungen liegt gemäß der gültigen Berufsordnung die maßgebende Deklaration des Weltärztebundes von Helsinki in der jeweils aktuellen Fassung zugrunde.

Unabhängig vom Beratungsergebnis macht die Ethikkommission Sie darauf aufmerksam, dass die ethische und rechtliche Verantwortung für die Durchführung einer Studie beim Leiter der Studie und bei allen beteiligten Ärzten liegt.

Mit freundlichen Grüßen

Prof. Dr. med. Thomas Sitzmann
Vorsitzender der Ethikkommission

Ethikkommission

Vorsitz:
Prof. Dr. med. Thomas Sitzmann

Stellv. Vorsitz:
Prof. Dr. med. Johannes Schröder
Prof. Dr. med. Rainer Mattern

Geschäftsleitung:
Dr. med. Verena Pfeilschifter

ANSPRECHPARTNER:

AMG-/MPG-Studien federführend:

AMG-Studien Soziologin:
Stefanie Schneider, Dipl. rer. pol.

Sonstige Studien:
Martina Jerman, DV Kauffrau

Rechnungswesen:
Katharina Schäfer, Industriekauffrau

Empfang:

Qualitätsmanagement:
Dr. med. Verena Pfeilschifter

Martina Jerman, DV Kauffrau

Bankverbindung
Baden-Württembergische Bank Stuttgart
Konto-Nr.: 2431 500 409
BLZ: 600 501 01
SWIFT/BIC Code: SOLADEST

# Appendix III: Reverse transcriptase and nested PCR reagents and conditions

## PCR reagents for cDNA and first round DNA synthesis

| Component | Volume |
|---|---|
| 2X Reaction Mix | 25 µl |
| Template RNA (33- 108ng/ul) | 5, 10 or 15µl |
| 10 µM/µl RT1 forward primer | 1 µl |
| 10 µM/µl RT-Rev-Stu1 reverse primer | 1 µl |
| SuperScript™ III RT/ Platinum® Taq High Fidelity Enzyme Mix* (5U/ul) | 1 µl |
| Nuclease free water | to 50 µl |

## PCR conditions for cDNA and first round DNA synthesis

| A: cDNA synthesis and pre-denaturation | B: PCR amplification | C: Final extension |
|---|---|---|
| Perform 1 cycle of: | Perform 40 cycles of: | 1 cycle of 68ºC for 5 minutes |
| 55°C or 60°C for 30 minutes | Denaturation, 94°C for 15 seconds | |
| 94°C for 2 minutes | Annealling, 50°C and increments of 1°C upto 60°C for 30 seconds at each trial | |
| | Extension, 68°C for 1 minute/kb | |

## PCR reagents for second round DNA synthesis

| Component | Volume |
|---|---|
| 5x G-C rich PCR Buffer | 5µl |
| 25 mM dNTPs | 0.5µl |
| 10 µM/µl RT3 forward primer | 1.0µl |
| 10 µM/µl RT4 reverse primer | 1.0µl |
| Verbatim High Fidelity DNA | 0.5 |
| Template DNA | 5.0µl |
| Nuclease free water | to 50 µl |

## Second round PCR amplification conditions

| PCR Step | Temperature | Duration | Number of Cycles |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 secs | 1 cycle |
| Denaturation | 98°C | 10 secs | 35 cycles |
| Annealing | 50°C upto with increments of 1°C upto 60°C | 30 secs | |
| Extension | 72°C | 30 secs | |
| Final Extension | 72°C | 10 mins | 1 cycle |
| Holding | 4°C | Infinity | |

## Primer groupings for PCR optimizations

| Group | Primer combinations for cDNA and first | Primer combinations for second round |
|---|---|---|
| A | F1 +R1 | F2 +R2 |
| B | F1 +R2 | F2 +R2 |
| C | F2+R1 | F2 +R2 |
| D | F2 +R2 | F2 +R2 |

F1: Pol1 forward primer; R1: RT-rev-stu1 reverse primer; F2: Prot-for-EcoR forward primer and R2: RT4 reverse primer

# Appendix IV: HIV-1 IC DNA and CCR5 reagents and PCR amplification conditions

## HIV-1 IC DNA PCR amplification reagents

| Component | Volume |
|---|---|
| 10X standard *Taq* reaction buffer | 5 µl |
| 25 mM dNTPs | 0.5 µl |
| 10 µM/µl 623 forward primer | 1 µl |
| 10 µM 788 reverse primer | 1 µl |
| Taq DNA polymerase | 0.25 µl |
| Nuclease free water | 37.25 µl |
| | 5µl |
| Nuclease free water | to 50 µl |

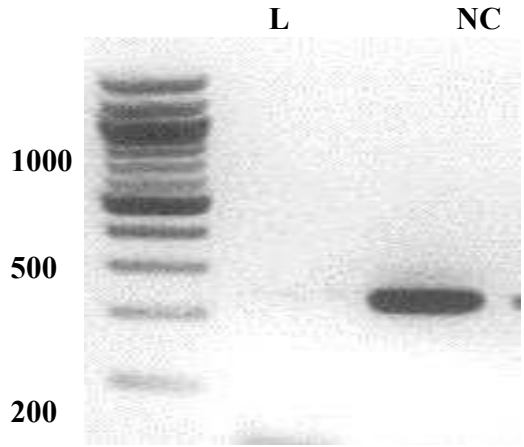
## CCR5 DNA PCR amplification reagents

| Component | Volume |
|---|---|
| 10X standard *Taq r*eaction buffer | 5 µl |
| 25 mM dNTPs | 0.5 µl |
| 10 µM LK46 forward primer | 1 µl |
| 10 µM LK47 reverse primer | 1 µl |
| Taq DNA polymerase | 0.25 µl |
| Template DNA | 5µl |
| Nuclease free water | to 50 µl |

## PCR cycling conditions for HIV-1 and CCR5 DNA

| PCR cycling conditions | | |
|---|---|---|
| Initial Denaturation | 1 cycle | |
| | 95 °C | 10 mins |
| | 50 cycles | |
| Denaturation | 95 °C | 30 secs |
| Annealing | 55 °C | **30 secs** |
| Extension | 72 °C | 30 secs |
| Final extention | 72 °C | 10 mins |
| Hold | 4 °C | Infinity |

# Appendix V: CCR5 gel amplification photograph




| Label on gel | Interpretation |
|---|---|
| L | 100 bp DNA ladder |
| NC | Negative control |
| SRIN48 | Patient derived CCR5 DNA |
| Target band size | 237 bp |

# Appendix VI: TOPO TA cloning reaction set-up

| Component | Volume |
|---|---|
| Fresh PCR product | 0.5-4 µl |
| Salt solution | 1 µl |
| pCR 4-TOPO plasmid vector | 1 µl |
| Nuclease free water | add to a total volume of 5µl |

# Appendix VII: Drug resistance profile of HIV-1 infected drug naïve patients from Nouna Burkina Faso

| Lab ID | Viral load Copies/ml | CD4 Tcell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN14 | 939988 | 202 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | | | | |
| SRIN15 | 971579 | 7 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | | | |
| SRIN16 | 707181 | 10 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN25 | 54837 | 152 | RNA | CRF06_cpx - Like | *K20I* | *V106I* | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | *V106I* | | | |
| SRIN30 | 1017643 | 163 | RNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| SRIN31 | 172739 | 271 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN32 | 430515 | 177 | RNA | CRF02_AG - Like | *K20I* | | *V118IV* | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | *V118I* | | |
| SRIN33 | 3399058 | 180 | RNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| SRIN34 | 2801732 | 3 | RNA | CRF02_AG - Like | *K20I* *L10V* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* *L10V* | no mutation | | | |
| SRIN38 | 223957 | 248 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN39 | 115615 | 262 | RNA | G - Like | *V82I* *K20I* | no mutation | | | |
| | | | DNA | G - Like | *V82I* *K20I* | no mutation | | | |
| SRIN41 | 2314931 | 144 | RNA | CRF02_AG - Like | *V11I* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *V11I* *K20I* | no mutation | | | |
| SRIN47 | 743697 | 66 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN48 | 2416935 | 147 | RNA | CRF06_cpx - Like | *K43T* *K20I* | **Y181C** | | | EFV,ETR,NVP,RPV |
| | | | DNA | CRF06_cpx - Like | *K43T* *K20I* | **Y181C** | *K219N* | | EFV,ETR,NVP,RPV |
| SRIN51 | 749061 | 284 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN52 | 194086 | 153 | RNA | CRF02_AG - Like | *L10I* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10I* *K20I* | no mutation | | | |

**DR** represents drug resistance; **LR** represents level resistance Type of template: RNA; **Region analyzed**: PR 1-99 codons and RT (2/3) 1-335 codons. Primary resistance mutations are **bolded,** primary resistance mutations in other expert list are **bolded and underlined** and minor and accessory resistance mutations ***are italicized***.

## Appendix VII continued

| Lab ID | Viral load Copies/ml | CD4 Tcell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN55 | 93865 | 227 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN56 | 1593082 | 81 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | | | *L10V* | | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | | | no mutation | | | | |
| SRIN57 | 653455 | 198 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | | | |
| SRIN58 | 256724 | 231 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | | | *L10V* | | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | | | *L10V* | | | | |
| SRIN60 | 121581 | 226 | RNA | CRF02_AG - Like | *K20IV* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN61 | 749061 | 284 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| SRIN62 | 2005003 | 180 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN63 | 86731 | 316 | RNA | 02A1-Like | *K20I* | | | | |
| | | | | | | *V108IV* | | NVP | |
| | | | | | | *V90I* | | | |
| | | | DNA | 02A1-Like | *K20I* | | | | |
| | | | | | | *V108IV* | | NVP | |
| | | | | | | *V90I* | | | |
| SRIN65 | 145511 | 32 | RNA | CRF01_AE- like | *L10I* | | | | |
| | | | | | | *V179I* | | | |
| | | | DNA | CRF01_AE- like | *L10I* | | | | |
| | | | | | | *V179I* | | | |
| SRIN66 | 227199 | 140 | RNA | CRF01_AE- like | *L10V* | | | | |
| | | | | | | *E138A* | | RPV | |
| | | | | | | *V179I* | | | |
| | | | DNA | CRF01_AE- like | *L10V* | | | | |
| | | | | | | *E138A* | | RPV | |
| | | | | | | *V179I* | | | |
| SRIN67 | 1195067 | 293 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN68 | 2201354 | 411 | RNA | G - Like | *K20I* | no mutation | | | |
| | | | DNA | G - Like | *K20I* | no mutation | | | |
| SRIN69 | 515487 | 428 | RNA | CRF02_AG - Like | *L10LV* | no mutation | | | |
| | | | | | *K20I* | | | | |
| | | | | | | | L210LW | AZT, D4T | |
| | | | DNA | CRF02_AG - Like | *L10V* | no mutation | | | |
| | | | | | *K20I* | | | | |
| | | | | | | | L210LW | AZT, D4T | |
| SRIN70 | 46747 | 13 | RNA | G - Like | *A71AT* | | | | |
| | | | | | *K20I* | | | | |
| | | | | | | V90I | | | |
| | | | DNA | G - Like | *A71T* | | | | |
| | | | | | *K20I* | | | | |
| | | | | | | V90I | | | |
| SRIN71 | 325435 | 256 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN72 | 95911 | 333 | RNA | G - Like | *K20I* | | | | |
| | | | | | | *V179I* | | | |
| | | | DNA | G - Like | *K20I* | | | | |
| | | | | | | *V179I* | | | |

## Appendix VII continued

| Lab ID | Viral load Copies/ml | CD4 Tcell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN73 | 1818940 | 157 | RNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10V* *K20I* | no mutation | | | |
| SRIN74 | 40625 | 7 | RNA | CRF02_AG - Like | *K20I* | *V90I* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | *V90I* | | | |
| SRIN75 | 13291 | 400 | RNA | CRF02_AG - Like | *L10LV* *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10V* *K20I* | | | | |
| SRIN76 | 51695 | 177 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN77 | 45750 | 123 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN78 | 24482 | 269 | RNA | D- Like | no mutation | | | | |
| | | | DNA | D- Like | no mutation | | | | |
| SRIN79 | 11678 | 256 | RNA | CRF02_AG - Like | *K20I* | *V179I* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | *V179I* | | | |
| SRIN80 | 23449 | 152 | RNA | CRF01_AE- like | no mutation | *V179I* | | | |
| | | | DNA | CRF01_AE- like | no mutation | *V179I* | | | |
| SRIN81 | 590908 | 323 | RNA | CRF02_AG - Like | *K20I, T74S* | *K238KR* | | NFV | |
| | | | DNA | CRF02_AG - Like | *K20I, T74S* | *K238R* | | NFV | |
| SRIN82 | 569328 | 140 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN83 | 43140 | 279 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN84 | 15793 | 156 | RNA | CRF02_AG - Like | *L10I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *L10I* | no mutation | | | |
| SRIN85 | 1434896 | 215 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN86 | 2233224 | 146 | | CRF02_AG - Like | *L10I* *K20I* | no mutation | | | |
| | | | | CRF02_AG - Like | *L10I* *K20I* | no mutation | | | |
| SRIN87 | 4040062 | 65 | RNA | CRF02_AG - Like | *L33F* *K20R* *I93M* | | | | |
| | | | DNA | CRF02_AG - Like | *L33F* *K20R* *I93M* | no mutation | | | |
| SRIN88 | 574163 | 45 | RNA | CRF02_AG - Like | *L10V* *K20I* | *V108I* | no mutation | | |
| | | | DNA | CRF02_AG - Like | *L10V* *K20I* | *V108I* | no mutation | | |
| SRIN89 | 677363 | 54 | RNA | 02A1-Like | *K20I* | *V179I* | no mutation | | |
| | | | DNA | 02A1-Like | *K20I* | *V179I* | no mutation | | |
| SRIN90 | 1379797 | 236 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| SRIN91 | 977237 | 66 | RNA | CRF06_cpx - Like | *L10V* *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *L10V* *K20I* | no mutation | | | |

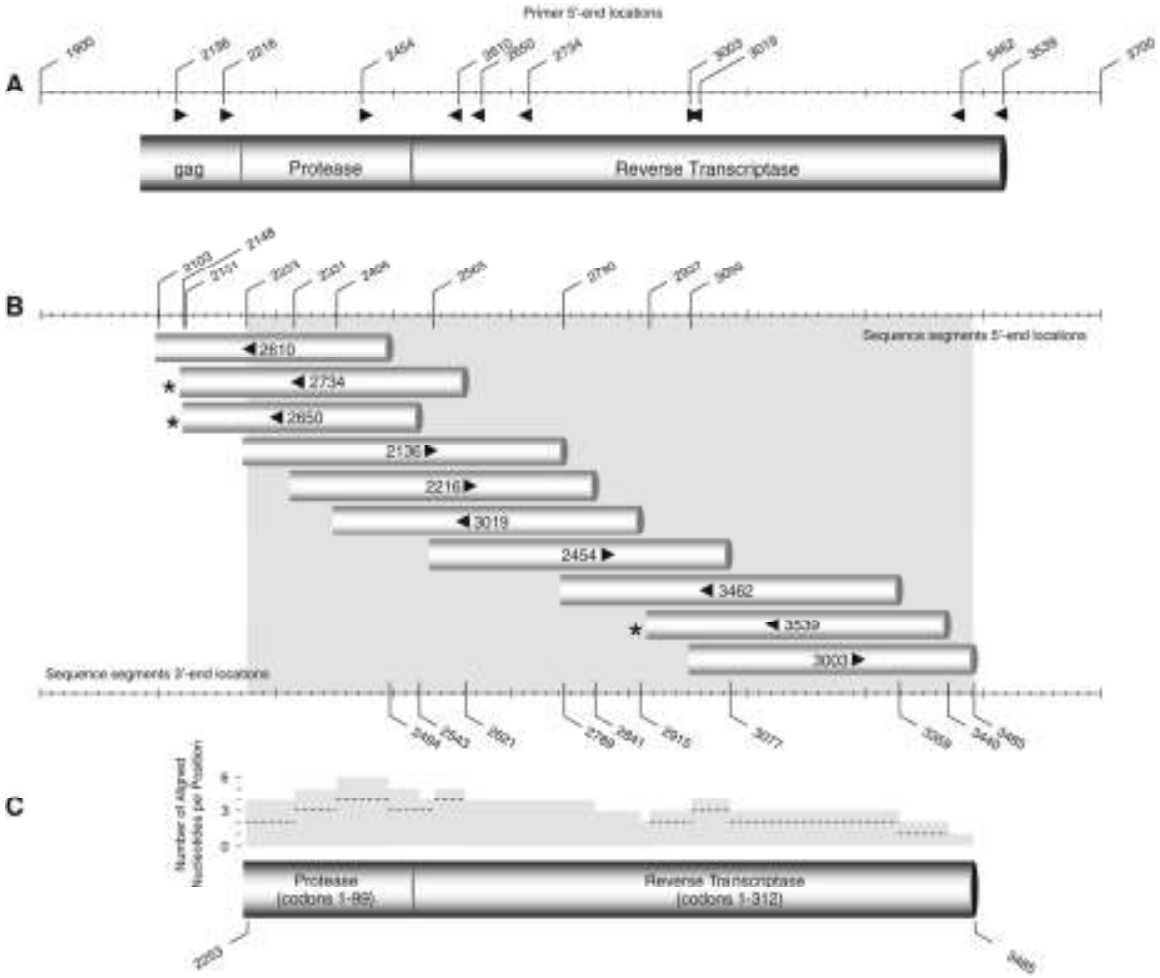
## Appendix VII continued.

| Lab ID | Viral load Copies/ml | CD4 Tcell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN92 | 561916 | 282 | RNA | CRF06_cpx - Like | *L10I* *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *L10V* *K20I* | no mutation | | | |
| SRIN93 | 10000000 | 643 | RNA | CRF02_AG - Like | *K20I* *L10V* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* no mutation | no mutation | | | |
| SRIN94 | 10000000 | 116 | RNA | CRF02_AG - Like | *K20I* | *V90I* | | | |
| | | | DNA | CRF02_AG - Like | no mutation | no mutation | | | |
| SRIN95 | 10000000 | 7 | RNA | CRF02_AG - Like | *K20R* *L10I* | *V90I* | | | |
| | | | DNA | | *K20I* *L10I* | no mutation | | | |
| SRIN96 | 1892979 | 90 | RNA | CRF02_AG - Like | *K20I* *L89I* | no mutation | | | |
| | | | DNA | | *K20I* *L89I* | no mutation | | | |
| SRIN97 | 1581674 | 106 | RNA | CRF06_cpx - Like | *K20I* | *no mutation* | *T69N* | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | *no mutation* | *T69N* | | |
| SRIN98 | 10000000 | 63 | RNA | CRF02_AG - Like | *K20I* *L10V* | *no mutation* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* *L10V* | *no mutation* | | | |
| SRIN99 | 10000000 | 322 | RNA | CRF02_AG - Like | *K20I* *L10V* | *no mutation* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* *L10V* | *no mutation* | | | |
| SRIN100 | 71380 | 292 | RNA | CRF02_AG - Like | *K20I* | *no mutation* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | *no mutation* | | | |
| SRIN101 | 339531 | 408 | RNA | CRF06_cpx - Like | *K20I* | *no mutation* | *E44D* | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | *no mutation* | *E44D* | | |
| SRIN102 | 3790705 | 88 | RNA | CRF02_AG - Like | *K20I* | *no mutation* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | *no mutation* | | | |
| SRIN103 | 182608 | 397 | RNA | CRF01_AE- like | *L10V* | *V179I* | no mutation | | |
| | | | DNA | CRF01_AE- like | *L10V* | *V179I* | no mutation | | |
| SRIN105 | 98211 | 218 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN106 | 3589550 | 176 | RNA | CRF02_AG - Like | *K20I* | *V108I* *V90I* | no mutation | NVP | |
| | | | DNA | CRF02_AG - Like | *K20I* | *V108I* *V90I* | no mutation | NVP | |
| SRIN107 | 290269 | 406 | RNA | CRF02_AG - Like | *K20I* *L89I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* *L89I* | no mutation | | | |

## Appendix VII continued.

| Lab ID | Viral load Copies/ml | CD4 Tcell counts (cells/ul) | Type of Template | Subtype PR-RT | PI Mutations | NNRTI Mutations | NRTI Mutations | Drug Resistance Interpretations Per Resistance Scores | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Low-LR | High/ Intermediate -LR |
| SRIN108 | 27833 | 232 | RNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | *V179I* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | *V179I* | | | |
| SRIN109 | 405359 | 313 | RNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | *V179I* | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | *V179I* | | | |
| SRIN110 | 1420627 | 44 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| SRIN111 | 236592 | 212 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| SRIN112 | 1373030 | 213 | RNA | CRF02_AG Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG Like | *K20I* | no mutation | | | |
| SRIN113 | 1803350 | 11 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN114 | 21777 | 349 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN116 | 262060 | 508 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN117 | 90498 | 341 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN118 | 2579 | 279 | RNA | CRF01_AE- like | *L10V* | | no mutation | | |
| | | | | | | *V179I* | | | |
| | | | DNA | CRF01_AE- like | *L10V* | | no mutation | | |
| | | | | | | *V179I* | | | |
| SRIN119 | 272998 | 364 | RNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | | | *V11I* | | | | |
| | | | DNA | CRF06_cpx - Like | *K20I* | no mutation | | | |
| | | | | | *V11I* | | | | |
| SRIN120 | 1400866 | 504 | RNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | *V11I* | | | | |
| | | | | | *L10I* | | | | |
| | | | | | | **K103N** | | | NVP, EFV |
| | | | DNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | *V11I* | | | | |
| | | | | | *L10I* | | | | |
| | | | | | | **K103KN** | | | NVP, EFV |
| SRIN121 | 224037 | 281 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN122 | 1531228 | 170 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN123 | 976517 | 301 | RNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | E138A | | RVP | |
| | | | DNA | CRF02_AG - Like | *K20I* | | no mutation | | |
| | | | | | | E138A | | RVP | |
| SRIN124 | 225569 | 467 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| SRIN125 | 1126782 | 33 | RNA | CRF02_AG - Like | *K20I* | no mutation | | | |
| | | | DNA | CRF02_AG - Like | *K20I* | no mutation | | | |

# Appendix VIII: Schematic diagram for the sequencing procedure of the PR/RT region of HIV-1 pol




a. Reproduced from (Kousiappa et al., 2009)

b. Diagram A represents the final amplicon containing gag ( p6) and pol ( prot and p51 RT) regions.

c. The black triangular arrows represents the 5'-end binding positions of the sequencing primers.

d. The gray cylindrical bars in diagram B represenst the sequence segments obtained from each sequencing primer

e. The dark asterisks represents alternative DNA segments. This signifies that primer numbered 2734and 2650 can be used instead of 2610 and 3539 instead of 3462.

f. Finally the gray cylindrical bar in diagram C represents the expected final sequenced derieved product after aligning the short sequenced fragments as shown above.

# List of peer reviewed abstracts for poster and oral presentations

**Affram Y** and Kraeusslich HG. Intracellular Viral Levels of Drug Resistant and Drug Sensitive HIV-1 Infection and its Correlation to Disease Progression in a Population of Africans. Summer School of Infection Research. Dresden, Germany. 22nd -27th June 2014. **Poster Presentation**.

**Affram Y** and Kraeusslich HG. Correlating Intracellular Viral Levels of Drug Resistant and Drug Sensitive HIV-1 Infection to Disease Progression Among Drug Naïve Africans. Young Researchers in Life Sciences Conference. Pasteur Institute Paris, France, 26th -28th March 2014. **Oral presentation.**

**Affram Y** and Kraeusslich HG. Correlating Intracellular Viral Levels of Drug Resistant and Sensitive HIV-1 Infection to Disease Progression Among Drug Naïve Africans 2nd European Seminar in Virology (EuSeV) on Vaccines and Antivirals. University of Bologna Residential Center in Bertinoro Italy, 13th -15th June 2014. **Oral and Poster Presentation.**

**Affram Y,** Pollakis G, Sagoe KWC, Barnor J, Mingle JAA, Ampofo W, Paxton W. Correlates of Non-transmission of Human Immunodeficiency Virus Type 1, In Serologically Discordant Couples In Accra, Ghana. 25th Annual Meeting of the German Society of Tropical Medicine and Parasitology, Universitaetsplatz, University of Heidelberg, Germany. 14th to 17th March 2012. Abstract Number 058. **Oral Presentation**.

# Manuscripts in preparation

1. **Affram Y**, Schnitzler P, Kräusslich HG. HIV-1 drug resistance mutations among drug naive and follow-up patients experiencing virological failure in rural Burkina Faso using population based Sanger sequencing.

2. **Affram Y**, Schnitzler P, Kräusslich HG. Relating HIV-1 Intracellular DNA levels to primary markers of disease progression among HIV infected drug naive Burkinabés.

3. **Affram Y**, Schnitzler P, Kräusslich HG. Resolving discrepancies in bulk sequencing of paired patient samples of RNA and DNA using ultra deep sequencing to reveal drug resistance mutations and amino acid changes.

# References

Abecasis, A.B., Wensing, A.M., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D.A., Albert, J., Asjo, B., Balotta, C., Beshkov, D., Camacho, R.J., Clotet, B., De Gascun, C., Griskevicius, A., Grossman, Z., Hamouda, O., Horban, A., Kolupajeva, T., Korn, K., Kostrikis, L.G., Kucherer, C., Liitsola, K., Linka, M., Nielsen, C., Otelea, D., Paredes, R., Poljak, M., Puchhammer-Stockl, E., Schmit, J.C., Sonnerborg, A., Stanekova, D., Stanojevic, M., Struck, D., Boucher, C.A., Vandamme, A.M., 2013. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. Retrovirology 10, 7.

Abram, M.E., Ferris, A.L., Shao, W., Alvord, W.G., Hughes, S.H., 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. Journal of virology 84, 9864-9878.

Adje, C., Cheingsong, R., Roels, T.H., Maurice, C., Djomand, G., Verbiest, W., Hertogs, K., Larder, B., Monga, B., Peeters, M., Eholie, S., Bissagene, E., Coulibaly, M., Respess, R., Wiktor, S.Z., Chorba, T., Nkengasong, J.N., 2001. High prevalence of genotypic and phenotypic HIV-1 drug-resistant strains among patients receiving antiretroviral therapy in Abidjan, Cote d'Ivoire. Journal of acquired immune deficiency syndromes (1999) 26, 501-506.

Aghokeng, A.F., Monleau, M., Eymard-Duvernay, S., Dagnra, A., Kania, D., Ngo-Giang-Huong, N., Toni, T.D., Toure-Kane, C., Truong, L.X., Delaporte, E., Chaix, M.L., Peeters, M., Ayouba, A., 2014. Extraordinary heterogeneity of virological outcomes in patients receiving highly antiretroviral therapy and monitored with the World Health Organization public health approach in sub-saharan Africa and southeast Asia. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 58, 99-109.

Akinsete, O., Hirigoyen, D., Cartwright, C., Schut, R., Kantor, R., Henry, K., 2004. K103N mutation in antiretroviral therapy-naive African patients infected with HIV type 1. Clinical

infectious diseases : an official publication of the Infectious Diseases Society of America 39, 575-578.

Albanese, A., Arosio, D., Terreni, M., Cereseto, A., 2008. HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery. PloS one 3, e2413.

Ananworanich, J., Dube, K., Chomont, N., 2015. How does the timing of antiretroviral therapy initiation in acute infection affect HIV reservoirs? Current opinion in HIV and AIDS 10, 18-28.

Arion, D., Parniak, M.A., 1999. HIV resistance to zidovudine: the role of pyrophosphorolysis. Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy 2, 91-95.

Arnott, A., Jardine, D., Wilson, K., Gorry, P.R., Merlin, K., Grey, P., Law, M.G., Dax, E.M., Kelleher, A.D., Smith, D.E., McPhee, D.A., 2010. High viral fitness during acute HIV-1 infection. PloS one 5.

Arts, E.J., Hazuda, D.J., 2012. HIV-1 antiretroviral drug therapy. Cold Spring Harbor perspectives in medicine 2, a007161.

Avila-Rios, S., Garcia-Morales, C., Garrido-Rodriguez, D., Tapia-Trejo, D., Giron-Callejas, A.C., Mendizabal-Burastero, R., Escobar-Urias, I.Y., Garcia-Gonzalez, B.L., Navas-Castillo, S., Pinzon-Meza, R., Mejia-Villatoro, C.R., Reyes-Teran, G., 2014. HIV-1 Drug Resistance Surveillance in Antiretroviral Treatment-Naive Individuals from a Reference Hospital in Guatemala, 2010-2013. AIDS Res Hum Retroviruses.

Baca, M., Kent, S.B., 1993. Catalytic contribution of flap-substrate hydrogen bonds in "HIV-1 protease" explored by chemical synthesis. Proceedings of the National Academy of Sciences of the United States of America 90, 11638-11642.

Banks, L., Gholamin, S., White, E., Zijenah, L., Katzenstein, D.A., 2012. Comparing Peripheral Blood Mononuclear Cell DNA and Circulating Plasma viral RNA pol Genotypes of Subtype C HIV-1. Journal of AIDS & clinical research 3, 141-147.

Barennes, H., Guillet, S., Limsreng, S., Him, S., Nouhin, J., Hak, C., Srun, C., Viretto, G., Ouk, V., Delfraissy, J.F., Segeral, O., 2014. Virological failure and HIV-1 drug resistance mutations among naive and antiretroviral pre-treated patients entering the ESTHER program of Calmette Hospital in Cambodia. PloS one 9, e105736.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (New York, N.Y.) 220, 868-871.

Bartolo, I., Zakovic, S., Martin, F., Palladino, C., Carvalho, P., Camacho, R., Thamm, S., Clemente, S., Taveira, N., 2014. HIV-1 Diversity, Transmission Dynamics and Primary Drug Resistance in Angola. PloS one 9, e113626.

Beaudrot, M.E., DeFranco, E.A., Elchert, J.A., 2015. Influence of Gestational Weight Gain and Body Mass Index on Cesarean Delivery Risk in Adolescent Pregnancies [6]. Obstetrics and gynecology 125 Suppl 1, 2S.

Blick, G., Kagan, R.M., Coakley, E., Petropoulos, C., Maroldo, L., Greiger-Zanlungo, P., Gretz, S., Garton, T., 2007. The probable source of both the primary multidrug-resistant (MDR) HIV-1 strain found in a patient with rapid progression to AIDS and a second recombinant MDR strain found in a chronically HIV-1-infected patient. The Journal of infectious diseases 195, 1250-1259.

Bon, I., Alessandrini, F., Borderi, M., Gorini, R., Re, M.C., 2007. Analysis of HIV-1 drug-resistant variants in plasma and peripheral blood mononuclear cells from untreated individuals: implications for clinical management. New Microbiol 30, 313-317.

Broder, S., Gallo, R.C., 1984. A pathogenic retrovirus (HTLV-III) linked to AIDS. The New England journal of medicine 311, 1292-1297.

Buckheit, R.W., Jr., 2004. Understanding HIV resistance, fitness, replication capacity and compensation: targeting viral fitness as a therapeutic strategy. Expert opinion on investigational drugs 13, 933-958.

Buonaguro, L., Tornesello, M.L., Buonaguro, F.M., 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. Journal of virology 81, 10209-10219.

Butler, S.L., Hansen, M.S., Bushman, F.D., 2001. A quantitative assay for HIV DNA integration in vivo. Nature medicine 7, 631-634.

Buzon, M.J., Martin-Gayo, E., Pereyra, F., Ouyang, Z., Sun, H., Li, J.Z., Piovoso, M., Shaw, A., Dalmau, J., Zangger, N., Martinez-Picado, J., Zurakowski, R., Yu, X.G., Telenti, A., Walker, B.D., Rosenberg, E.S., Lichterfeld, M., 2014. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. Journal of virology 88, 10056-10065.

Cain, L.E., Logan, R., Robins, J.M., Sterne, J.A., Sabin, C., Bansi, L., Justice, A., Goulet, J., van Sighem, A., de Wolf, F., Bucher, H.C., von Wyl, V., Esteve, A., Casabona, J., del Amo, J., Moreno, S., Seng, R., Meyer, L., Perez-Hoyos, S., Muga, R., Lodi, S., Lanoy, E., Costagliola, D., Hernan, M.A., 2011. When to initiate combined antiretroviral therapy to reduce mortality and

AIDS-defining illness in HIV-infected persons in developed countries: an observational study. Ann Intern Med 154, 509-515.

Callegaro, A., Di Filippo, E., Astuti, N., Ortega, P.A., Rizzi, M., Farina, C., Valenti, D., Maggiolo, F., 2014. Early clinical response and presence of viral resistant minority variants: a proof of concept study. Journal of the International AIDS Society 17, 19759.

CDC, 1992. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control 41, 1-19.

Chabria, S.B., Gupta, S., Kozal, M.J., 2014. Deep sequencing of HIV: clinical and research applications. Annual review of genomics and human genetics 15, 295-325.

Chew, C.B., Potter, S.J., Wang, B., Wang, Y.M., Shaw, C.O., Dwyer, D.E., Saksena, N.K., 2005. Assessment of drug resistance mutations in plasma and peripheral blood mononuclear cells at different plasma viral loads in patients receiving HAART. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 33, 206-216.

Clavel, F., Hance, A.J., 2004. HIV drug resistance. The New England journal of medicine 350, 1023-1035.

Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., et al., 1986. What to call the AIDS virus? Nature 321, 10.

Cohen, E.A., Lu, Y., Gottlinger, H., Dehni, G., Jalinoos, Y., Sodroski, J.G., Haseltine, W.A., 1990. The T open reading frame of human immunodeficiency virus type 1. Journal of acquired immune deficiency syndromes 3, 601-608.

Coleman, C.M., Wu, L., 2009. HIV interactions with monocytes and dendritic cells: viral latency and reservoirs. Retrovirology 6, 51.

Collenberg, E., Ouedraogo, T., Ganame, J., Fickenscher, H., Kynast-Wolf, G., Becher, H., Kouyate, B., Krausslich, H.G., Sangare, L., Tebit, D.M., 2006. Seroprevalence of six different viruses among pregnant women and blood donors in rural and urban Burkina Faso: A comparative analysis. J Med Virol 78, 683-692.

Colson, P., Ravaux, I., Tamalet, C., Glazunova, O., Baptiste, E., Chabriere, E., Wiedemann, A., Lacabaratz, C., Chefrour, M., Picard, C., Stein, A., Levy, Y., Raoult, D., 2014. HIV infection en route to endogenization: two cases. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases.

Cone, R.W., Gowland, P., Opravil, M., Grob, P., Ledergerber, B., 1998. Levels of HIV-infected peripheral blood cells remain stable throughout the natural history of HIV-1 infection. Swiss HIV Cohort Study. Aids 12, 2253-2260.

Craigie, R., Bushman, F.D., 2012. HIV DNA integration. Cold Spring Harbor perspectives in medicine 2, a006890.

Daar, E.S., Richman, D.D., 2005. Confronting the emergence of drug-resistant HIV type 1: impact of antiretroviral therapy on individual and population resistance. AIDS research and human retroviruses 21, 343-357.

Das, K., Arnold, E., 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. Current opinion in virology 3, 111-118.

Dawson, L., Yu, X.F., 1998. The role of nucleocapsid of HIV-1 in virus assembly. Virology 251, 141-157.

de Mendoza, C., Barreiro, P., Benitez, L., Soriano, V., 2014. Gene therapy for HIV infection. Expert opinion on biological therapy, 1-9.

Delaugerre, C., Braun, J., Charreau, I., Delarue, S., Nere, M.L., de Castro, N., May, T., Marchou, B., Simon, F., Molina, J.M., Aboulker, J.P., group, A.E.s., 2012. Comparison of resistance mutation patterns in historical plasma HIV RNA genotypes with those in current proviral HIV DNA genotypes among extensively treated patients with suppressed replication. HIV medicine 13, 517-525.

Delviks-Frankenberry, K., Galli, A., Nikolaitchik, O., Mens, H., Pathak, V.K., Hu, W.S., 2011. Mechanisms and factors that influence high frequency retroviral recombination. Viruses 3, 1650-1680.

Demetriou, V.L., van de Vijver, D.A., Kousiappa, I., Balotta, C., Clotet, B., Grossman, Z., Jorgensen, L.B., Lepej, S.Z., Levy, I., Nielsen, C., Paraskevis, D., Poljak, M., Roman, F., Ruiz, L., Schmidt, J.C., Vandamme, A.M., Van Laethem, K., Vercauteren, J., Kostrikis, L.G., 2010. Cellular HIV-1 DNA levels in drug sensitive strains are equivalent to those in drug resistant strains in newly-diagnosed patients in Europe. PLoS One 5, e10976.

Derache, A., Maiga, A.I., Traore, O., Akonde, A., Cisse, M., Jarrousse, B., Koita, V., Diarra, B., Carcelain, G., Barin, F., Pizzocolo, C., Pizarro, L., Katlama, C., Calvez, V., Marcelin, A.G., 2008. Evolution of genetic diversity and drug resistance mutations in HIV-1 among untreated patients from Mali between 2005 and 2006. The Journal of antimicrobial chemotherapy 62, 456-463.

Derache, A., Shin, H.S., Balamane, M., White, E., Israelski, D., Klausner, J.D., Freeman, A.H., Katzenstein, D., 2015. HIV drug resistance mutations in proviral DNA from a community treatment program. PloS one 10, e0117430.

Dionisio, D., Vivarelli, A., Zazzi, M., Esperti, F., Uberti, M., Polidori, M., 2001. Extent of human immunodeficiency virus type 1 drug resistance as a predictor of virological failure after genotype-guided treatment switch. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 33, 706-709.

Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. Microbiology and molecular biology reviews : MMBR 76, 159-216.

Dudek, R.W., 2007. High-yield cell & molecular biology, 2nd ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.

Dybul, M., Fauci, A.S., Bartlett, J.G., Kaplan, J.E., Pau, A.K., 2002. Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. Annals of internal medicine 137, 381-433.

Engelman, A., 2009. Isolation and analysis of HIV-1 preintegration complexes. Methods Mol Biol 485, 135-149.

Eron, J.J., Jr., 1996. The treatment of antiretroviral-naive subjects with the 3TC/zidovudine combination: a review of North American (NUCA 3001) and European (NUCB 3001) trials. Aids 10 Suppl 5, S11-19.

Essex, M., McLane, M.F., Kanki, P., Allan, J., Kitchen, L., Lee, T.H., 1985. Retroviruses associated with leukemia and ablative syndromes in animals and in human beings. Cancer research 45, 4534s-4538s.

Felsenstein, J., 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap Evolution 39, 783-791.

Ferrer, P., Montecinos, L., Tello, M., Tordecilla, R., Rodriguez, C., Ferres, M., Perez, C.M., Beltran, C., Guzman, M.A., Afani, A., 2013. HIV-1 tropism: a comparison between RNA and proviral DNA in routine clinical samples from Chilean patients. Virol J 10, 318.

Fidler, S., Olson, A., Fox, J., Phillips, A., Morrison, C., Thornhill, J., Bucher, H., Muga, R., Porter, K., 2014. The importance of viral blips and duration of therapy initiated in primary infection in maintaining viral control after stopping cART. Journal of the International AIDS Society 17, 19820.

Fox, E.J., Reid-Bayliss, K.S., Emond, M.J., Loeb, L.A., 2014. Accuracy of Next Generation Sequencing Platforms. Next generation, sequencing & applications 1.

Frankel, A.D., Young, J.A., 1998. HIV-1: fifteen proteins and an RNA. Annual review of biochemistry 67, 1-25.

Fu, J., Sha, B.E., Thomas, L.L., 2011. HIV-1-infected peripheral blood mononuclear cells enhance neutrophil survival and HLA-DR expression via increased production of GM-CSF: implications for HIV-1 infection. Journal of acquired immune deficiency syndromes 56, 16-25.

Fun, A., Wensing, A.M., Verheyen, J., Nijhuis, M., 2012. Human Immunodeficiency Virus Gag and protease: partners in resistance. Retrovirology 9, 63.

Gallant, J.E., Gerondelis, P.Z., Wainberg, M.A., Shulman, N.S., Haubrich, R.H., St Clair, M., Lanier, E.R., Hellmann, N.S., Richman, D.D., 2003. Nucleoside and nucleotide analogue reverse transcriptase inhibitors: a clinical review of antiretroviral resistance. Antiviral therapy 8, 489-506.

Garcia-Calleja, J.M., Gouws, E., Ghys, P.D., 2006. National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. Sexually transmitted infections 82 Suppl 3, iii64-70.

Garg, H., Blumenthal, R., 2008. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. Cellular and molecular life sciences : CMLS 65, 3134-3144.

Geretti, A.M., Conibear, T., Hill, A., Johnson, J.A., Tambuyzer, L., Thys, K., Vingerhoets, J., Van Delft, Y., 2014. Sensitive testing of plasma HIV-1 RNA and Sanger sequencing of cellular HIV-1 DNA for the detection of drug resistance prior to starting first-line antiretroviral therapy with etravirine or efavirenz. The Journal of antimicrobial chemotherapy 69, 1090-1097.

Gibson, R.M., Weber, J., Winner, D., Miller, M.D., Quinones-Mateu, M.E., 2014. Contribution of human immunodeficiency virus type 1 minority variants to reduced drug susceptibility in patients on an integrase strand transfer inhibitor-based therapy. PloS one 9, e104512.

Gonzalez-Serna, A., Min, J.E., Woods, C., Chan, D., Lima, V.D., Montaner, J.S., Harrigan, P.R., Swenson, L.C., 2014. Performance of HIV-1 drug resistance testing at low-level viremia and its ability to predict future virologic outcomes and viral evolution in treatment-naive individuals. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 58, 1165-1173.

Grant, R.M., Kuritzkes, D.R., Johnson, V.A., Mellors, J.W., Sullivan, J.L., Swanstrom, R., D'Aquila, R.T., Van Gorder, M., Holodniy, M., Lloyd Jr, R.M., Jr., Reid, C., Morgan, G.F., Winslow, D.L., 2003. Accuracy of the TRUGENE HIV-1 genotyping kit. Journal of clinical microbiology 41, 1586-1593.

Grossman, Z., Paxinos, E.E., Averbuch, D., Maayan, S., Parkin, N.T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C.J., Schapiro, J.M., 2004.

Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. Antimicrobial agents and chemotherapy 48, 2159-2165.

Hamers, R.L., Zaaijer, H.L., Wallis, C.L., Siwale, M., Ive, P., Botes, M.E., Sigaloff, K.C., Hoepelman, A.I., Stevens, W.S., Rinke de Wit, T.F., PharmAccess African Studies to Evaluate, R., 2013. HIV-HBV coinfection in Southern Africa and the effect of lamivudine- versus tenofovir-containing cART on HBV outcomes. Journal of acquired immune deficiency syndromes 64, 174-182.

Hemelaar, J., 2012. The origin and diversity of the HIV-1 pandemic. Trends in molecular medicine 18, 182-192.

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

Hill, C.P., Worthylake, D., Bancroft, D.P., Christensen, A.M., Sundquist, W.I., 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proceedings of the National Academy of Sciences of the United States of America 93, 3099-3104.

Hirnschall, G., Harries, A.D., Easterbrook, P.J., Doherty, M.C., Ball, A., 2013. The next generation of the World Health Organization's global antiretroviral guidance. Journal of the International AIDS Society 16, 18757.

Hirsch, M.S., Gunthard, H.F., Schapiro, J.M., Brun-Vezinet, F., Clotet, B., Hammer, S.M., Johnson, V.A., Kuritzkes, D.R., Mellors, J.W., Pillay, D., Yeni, P.G., Jacobsen, D.M., Richman, D.D., 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008

recommendations of an International AIDS Society-USA panel. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47, 266-285.

Hogan, C.M., Hammer, S.M., 2001. Host determinants in HIV infection and disease. Part 1: cellular and humoral immune responses. Annals of internal medicine 134, 761-776.

Holguin, A., Alvarez, A., Soriano, V., 2002. High prevalence of HIV-1 subtype G and natural polymorphisms at the protease gene among HIV-infected immigrants in Madrid. Aids 16, 1163-1170.

Hu, W.S., Hughes, S.H., 2012. HIV-1 reverse transcription. Cold Spring Harb Perspect Med 2.

Hu, Z., Giguel, F., Hatano, H., Reid, P., Lu, J., Kuritzkes, D.R., 2006. Fitness comparison of thymidine analog resistance pathways in human immunodeficiency virus type 1. Journal of virology 80, 7020-7027.

Hunt, R.C., Simhadri, V.L., Iandoli, M., Sauna, Z.E., Kimchi-Sarfaty, C., 2014. Exposing synonymous mutations. Trends in genetics : TIG 30, 308-321.

Ibrahim, K.Y., Recordon-Pinson, P., Malvy, D., Fleury, H., Segurado, A.C., 2012. Intermittent HIV-1 viremia (blips) and virological failure in a cohort of people living with HIV from Sao Paulo, Brazil. AIDS patient care and STDs 26, 512-515.

Ioannidis, J.P., Rosenberg, P.S., Goedert, J.J., Ashton, L.J., Benfield, T.L., Buchbinder, S.P., Coutinho, R.A., Eugen-Olsen, J., Gallart, T., Katzenstein, T.L., Kostrikis, L.G., Kuipers, H., Louie, L.G., Mallal, S.A., Margolick, J.B., Martinez, O.P., Meyer, L., Michael, N.L., Operskalski, E., Pantaleo, G., Rizzardi, G.P., Schuitemaker, H., Sheppard, H.W., Stewart, G.J., Theodorou, I.D., Ullum, H., Vicenzi, E., Vlahov, D., Wilkinson, D., Workman, C., Zagury, J.F., O'Brien, T.R., 2001a. Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A alleles on HIV-1

disease progression: An international meta-analysis of individual-patient data. Ann Intern Med 135, 782-795.

Ioannidis, J.P., Rosenberg, P.S., Goedert, J.J., Ashton, L.J., Benfield, T.L., Buchbinder, S.P., Coutinho, R.A., Eugen-Olsen, J., Gallart, T., Katzenstein, T.L., Kostrikis, L.G., Kuipers, H., Louie, L.G., Mallal, S.A., Margolick, J.B., Martinez, O.P., Meyer, L., Michael, N.L., Operskalski, E., Pantaleo, G., Rizzardi, G.P., Schuitemaker, H., Sheppard, H.W., Stewart, G.J., Theodorou, I.D., Ullum, H., Vicenzi, E., Vlahov, D., Wilkinson, D., Workman, C., Zagury, J.F., O'Brien, T.R., International Meta-Analysis of, H.I.V.H.G., 2001b. Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A alleles on HIV-1 disease progression: An international meta-analysis of individual-patient data. Annals of internal medicine 135, 782-795.

Jabara, C.B., Jones, C.D., Roach, J., Anderson, J.A., Swanstrom, R., 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proceedings of the National Academy of Sciences of the United States of America 108, 20166-20171.

Jain, M.K., 2009. Mortality in patients coinfected with hepatitis B virus and HIV: could antiretroviral therapy make a difference? Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 48, 1772-1774.

Jiamsakul, A., Sungkanuparph, S., Law, M., Kantor, R., Praparattanapan, J., Li, P.C., Phanuphak, P., Merati, T., Ratanasuwan, W., Lee, C.K., Ditangco, R., Mustafa, M., Singtoroj, T., Kiertiburanakul, S., Study, T.A.S.t.E.R.-M., 2014. HIV multi-drug resistance at first-line antiretroviral failure and subsequent virological response in Asia. Journal of the International AIDS Society 17, 19053.

Johnson, V.A., Brun-Vezinet, F., Clotet, B., Gunthard, H.F., Kuritzkes, D.R., Pillay, D., Schapiro, J.M., Richman, D.D., 2010. Update of the drug resistance mutations in HIV-1: December 2010. Topics in HIV medicine : a publication of the International AIDS Society, USA 18, 156-163.

Johnson, V.A., Calvez, V., Gunthard, H.F., Paredes, R., Pillay, D., Shafer, R., Wensing, A.M., Richman, D.D., 2011. 2011 update of the drug resistance mutations in HIV-1. Topics in antiviral medicine 19, 156-164.

Johnson.VA, Calvez. V, Günthard.HF, Paredes.R, Pillay, D., Shafer, R.W., Wensing, A., Richman, D., 2013. Update of the Drug Resistance Mutations in HIV-1: March 2013. IAS–USA Topics in Antiviral Medicine 21.

Jones, L.E., Perelson, A.S., 2007. Transient viremia, plasma viral load, and reservoir replenishment in HIV-infected patients on antiretroviral therapy. Journal of acquired immune deficiency syndromes 45, 483-493.

Kantor, R., Katzenstein, D., 2003. Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. AIDS reviews 5, 25-35.

Kantor, R., Katzenstein, D.A., Efron, B., Carvalho, A.P., Wynhoven, B., Cane, P., Clarke, J., Sirivichayakul, S., Soares, M.A., Snoeck, J., Pillay, C., Rudich, H., Rodrigues, R., Holguin, A., Ariyoshi, K., Bouzas, M.B., Cahn, P., Sugiura, W., Soriano, V., Brigido, L.F., Grossman, Z., Morris, L., Vandamme, A.M., Tanuri, A., Phanuphak, P., Weber, J.N., Pillay, D., Harrigan, P.R., Camacho, R., Schapiro, J.M., Shafer, R.W., 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. PLoS medicine 2, e112.

Kassa, E., Rinke de Wit, T.F., Hailu, E., Girma, M., Messele, T., Mariam, H.G., Yohannes, S., Jurriaans, S., Yeneneh, H., Coutinho, R.A., Fontanet, A.L., 1999. Evaluation of the World Health Organization staging system for HIV infection and disease in Ethiopia: association between clinical stages and laboratory markers. AIDS 13, 381-389.

Kempf, D.J., King, M.S., Bernstein, B., Cernohous, P., Bauer, E., Moseley, J., Gu, K., Hsu, A., Brun, S., Sun, E., 2004. Incidence of resistance in a double-blind study comparing lopinavir/ritonavir plus stavudine and lamivudine to nelfinavir plus stavudine and lamivudine. The Journal of infectious diseases 189, 51-60.

Kessler, H.A., 2005. Triple-nucleoside analog antiretroviral therapy: is there still a role in clinical practice? A review. MedGenMed : Medscape general medicine 7, 70.

Kostrikis, L.G., Touloumi, G., Karanicolas, R., Pantazis, N., Anastassopoulou, C., Karafoulidou, A., Goedert, J.J., Hatzakis, A., 2002a. Quantitation of human immunodeficiency virus type 1 DNA forms with the second template switch in peripheral blood cells predicts disease progression independently of plasma RNA load. Journal of virology 76, 10099-10108.

Kostrikis, L.G., Touloumi, G., Karanicolas, R., Pantazis, N., Anastassopoulou, C., Karafoulidou, A., Goedert, J.J., Hatzakis, A., Multicenter Hemophilia Cohort Study, G., 2002b. Quantitation of human immunodeficiency virus type 1 DNA forms with the second template switch in peripheral blood cells predicts disease progression independently of plasma RNA load. Journal of virology 76, 10099-10108.

Kousiappa, I., van de Vijver, D.A., Demetriades, I., Kostrikis, L.G., 2009. Genetic analysis of HIV type 1 strains from newly infected untreated patients in cyprus: high genetic diversity and low prevalence of drug resistance. AIDS research and human retroviruses 25, 23-35.

Kozal, M.J., Chiarella, J., St John, E.P., Moreno, E.A., Simen, B.B., Arnold, T.E., Lataillade, M., 2011. Prevalence of low-level HIV-1 variants with reverse transcriptase mutation K65R and the effect of antiretroviral drug exposure on variant levels. Antiviral therapy 16, 925-929.

Kwong, P.D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 2000. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. Structure 8, 1329-1339.

Lagarde, E., Congo, Z., Meda, N., Baya, B., Yaro, S., Sangli, G., Traore, Y., Van Renthergem, H., Carael, M., 2004. Epidemiology of HIV infection in urban Burkina Faso. International journal of STD & AIDS 15, 395-402.

Lamers, S.L., Salemi, M., McGrath, M.S., Fogel, G.B., 2008. Prediction of R5, X4, and R5X4 HIV-1 coreceptor usage with evolved neural networks. IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM 5, 291-300.

Langford, S.E., Ananworanich, J., Cooper, D.A., 2007. Predictors of disease progression in HIV infection: a review. AIDS research and therapy 4, 11.

Larder, B.A., Kemp, S.D., 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science (New York, N.Y.) 246, 1155-1158.

Lau, K.A., Wong, J.J., 2013. Current Trends of HIV Recombination Worldwide. Infectious disease reports 5, e4.

Laurent, C., Kouanfack, C., Laborde-Balen, G., Aghokeng, A.F., Mbougua, J.B., Boyer, S., Carrieri, M.P., Mben, J.M., Dontsop, M., Kaze, S., Molinari, N., Bourgeois, A., Mpoudi-Ngole, E., Spire, B., Koulla-Shiro, S., Delaporte, E., 2011. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. The Lancet. Infectious diseases 11, 825-833.

Ledergerber, B., Lundgren, J.D., Walker, A.S., Sabin, C., Justice, A., Reiss, P., Mussini, C., Wit, F., d'Arminio Monforte, A., Weber, R., Fusco, G., Staszewski, S., Law, M., Hogg, R., Lampe, F., Gill, M.J., Castelli, F., Phillips, A.N., Collaboration, P., 2004. Predictors of trend in CD4-positive T-cell count and mortality among HIV-1-infected individuals with virological failure to all three antiretroviral-drug classes. Lancet 364, 51-62.

Leitner.T, Korber. B, Daniels. M, Calef.C, B, F., 2005. HIV-1 subtype and circulating recombinant form (CRF) reference sequences (2005)

Li, Y., Hui, H., Burgess, C.J., Price, R.W., Sharp, P.M., Hahn, B.H., Shaw, G.M., 1992. Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation. Journal of virology 66, 6587-6600.

Lihana, R.W., Ssemwanga, D., Abimiku, A., Ndembi, N., 2012. Update on HIV-1 diversity in Africa: a decade in review. AIDS reviews 14, 83-100.

Lucas, G.M., 2005. Antiretroviral adherence, drug resistance, viral fitness and HIV disease progression: a tangled web is woven. The Journal of antimicrobial chemotherapy 55, 413-416.

Lucas, G.M., Gallant, J.E., Moore, R.D., 2004. Relationship between drug resistance and HIV-1 disease progression or death in patients undergoing resistance testing. AIDS 18, 1539-1548.

Lutzelberger, M., Reinert, L.S., Das, A.T., Berkhout, B., Kjems, J., 2006. A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. The Journal of biological chemistry 281, 18644-18651.

Maartens, G., Celum, C., Lewin, S.R., 2014. HIV infection: epidemiology, pathogenesis, treatment, and prevention. Lancet 384, 258-271.

Maas, J.J., Dukers, N., Krol, A., van Ameijden, E.J., van Leeuwen, R., Roos, M.T., de Wolf, F., Coutinho, R.A., Keet, I.P., 1998. Body mass index course in asymptomatic HIV-infected homosexual men and the predictive value of a decrease of body mass index for progression to AIDS. Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association 19, 254-259.

Mackie, N., 2006. Resistance to non-nucleoside reverse transcriptase inhibitors, in: Geretti, A.M. (Ed.), Antiretroviral Resistance in Clinical Practice, London.

Magiorkinis, E., Paraskevis, D., Sambatakou, H., Gargalianos, P., Haida, C., Vassilakis, A., Hatzakis, A., 2008. Emergence of an NNRTI resistance mutation Y181C in an HIV-infected NNRTI-naive patient. AIDS research and human retroviruses 24, 413-415.

Mayers, D., Bethel, J., Wainberg, M.A., Weislow, O., Schnittman, S., 1998. Human immunodeficiency virus proviral DNA from peripheral blood and lymph nodes demonstrates concordant resistance mutations to zidovudine (codon 215) and didanosine (codon 74). Division of AIDS Treatment Research Initiative 003 Study Group. The Journal of infectious diseases 177, 1730-1733.

Meda, N., Cousens, S., Kanki, B., 1996. Anaemia among women of reproductive age in Burkina Faso. World health forum 17, 369-372.

Meidani, M., Rezaei, F., Maracy, M.R., Avijgan, M., Tayeri, K., 2012. Prevalence, severity, and related factors of anemia in HIV/AIDS patients. Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences 17, 138-142.

Mellors, J.W., Kingsley, L.A., Rinaldo, C.R., Jr., Todd, J.A., Hoo, B.S., Kokka, R.P., Gupta, P., 1995. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Annals of internal medicine 122, 573-579.

Mellors, J.W., Munoz, A., Giorgi, J.V., Margolick, J.B., Tassoni, C.J., Gupta, P., Kingsley, L.A., Todd, J.A., Saah, A.J., Detels, R., Phair, J.P., Rinaldo, C.R., Jr., 1997a. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Annals of internal medicine 126, 946-954.

Mellors, J.W., Munoz, A., Giorgi, J.V., Margolick, J.B., Tassoni, C.J., Gupta, P., Kingsley, L.A., Todd, J.A., Saah, A.J., Detels, R., Phair, J.P., Rinaldo, C.R., Jr., 1997b. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Annals of internal medicine 126, 946-954.

Mellors, J.W., Rinaldo, C.R., Jr., Gupta, P., White, R.M., Todd, J.A., Kingsley, L.A., 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 272, 1167-1170.

Menon, S., 2010. Early initiation of antiretroviral therapy and universal HIV testing in sub-Saharan Africa: has WHO offered a milestone for HIV prevention? Journal of public health policy 31, 385-400.

Metzner, K., 2006. The significance of minority drug-resistant quasispecies, in: Geretti, A.M. (Ed.), Antiretroviral Resistance in Clinical Practice, London.

Metzner, K.J., Giulieri, S.G., Knoepfel, S.A., Rauch, P., Burgisser, P., Yerly, S., Gunthard, H.F., Cavassini, M., 2009. Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and -adherent patients. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 48, 239-247.

Miller, L.G., Golin, C.E., Liu, H., Hays, R.D., Hua, J., Wenger, N.S., Kaplan, A.H., 2004. No evidence of an association between transient HIV viremia ("Blips") and lower adherence to the antiretroviral medication regimen. The Journal of infectious diseases 189, 1487-1496.

Mphahlele, M., 2008. Impact of HIV co-infection on hepatitis B prevention and control: a view from sub-Saharan Africa
. South Afr J Infect Dis 23.

Mueller, N., van Bel, N., Berkhout, B., Das, A.T., 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. Virology 468-470C, 609-620.

Muula, A.S., 2008. HIV infection and AIDS among young women in South Africa. Croatian medical journal 49, 423-435.

Ndembi, N., Hamers, R.L., Sigaloff, K.C., Lyagoba, F., Magambo, B., Nanteza, B., Watera, C., Kaleebu, P., Rinke de Wit, T.F., 2011. Transmitted antiretroviral drug resistance among newly HIV-1 diagnosed young individuals in Kampala. AIDS 25, 905-910.

Neogi, U., Bontell, I., Shet, A., De Costa, A., Gupta, S., Diwan, V., Laishram, R.S., Wanchu, A., Ranga, U., Banerjea, A.C., Sonnerborg, A., 2012. Molecular epidemiology of HIV-1 subtypes in India: origin and evolutionary history of the predominant subtype C. PloS one 7, e39819.

Nettles, R.E., Kieffer, T.L., Kwon, P., Monie, D., Han, Y., Parsons, T., Cofrancesco, J., Jr., Gallant, J.E., Quinn, T.C., Jackson, B., Flexner, C., Carson, K., Ray, S., Persaud, D., Siliciano, R.F., 2005. Intermittent HIV-1 viremia (Blips) and drug resistance in patients receiving HAART. Jama 293, 817-829.

Nii-Trebi, N.I., Ibe, S., Barnor, J.S., Ishikawa, K., Brandful, J.A., Ofori, S.B., Yamaoka, S., Ampofo, W.K., Sugiura, W., 2013. HIV-1 drug-resistance surveillance among treatment-experienced and -naive patients after the implementation of antiretroviral therapy in Ghana. PloS one 8, e71972.

Nikolenko, G.N., Kotelkin, A.T., Oreshkova, S.F., Il'ichev, A.A., 2011. [Mechanisms of HIV-1 drug resistance to nucleoside and non-nucleoside reverse transcriptase inhibitors]. Molekuliarnaia biologiia 45, 108-126.

Nikolopoulos, G.K., Paraskevis, D., Hatzitheodorou, E., Moschidis, Z., Sypsa, V., Zavitsanos, X., Kalapothaki, V., Hatzakis, A., 2009. Impact of hepatitis B virus infection on the progression of AIDS and mortality in HIV-infected individuals: a cohort study and meta-analysis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 48, 1763-1771.

Nomaguchi, M., Miyake, A., Doi, N., Fujiwara, S., Miyazaki, Y., Tsunetsugu-Yokota, Y., Yokoyama, M., Sato, H., Masuda, T., Adachi, A., 2014. Natural single-nucleotide polymorphisms in the 3' region of the HIV-1 pol gene modulate viral replication ability. Journal of virology 88, 4145-4160.

Obermeier, M., Pironti, A., Berg, T., Braun, P., Daumer, M., Eberle, J., Ehret, R., Kaiser, R., Kleinkauf, N., Korn, K., Kucherer, C., Muller, H., Noah, C., Sturmer, M., Thielen, A., Wolf, E., Walter, H., 2012. HIV-GRADE: a publicly available, rules-based drug resistance interpretation algorithm integrating bioinformatic knowledge. Intervirology 55, 102-107.

Ojesina, A.I., Sankale, J.L., Odaibo, G., Langevin, S., Meloni, S.T., Sarr, A.D., Olaleye, D., Kanki, P.J., 2006. Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. AIDS research and human retroviruses 22, 770-779.

Okulicz, J.F., Le, T.D., Agan, B.K., Camargo, J.F., Landrum, M.L., Wright, E., Dolan, M.J., Ganesan, A., Ferguson, T.M., Smith, D.M., Richman, D.D., Little, S.J., Clark, R.A., He, W., Ahuja, S.K., 2014. Influence of the Timing of Antiretroviral Therapy on the Potential for Normalization of Immune Status in Human Immunodeficiency Virus 1-Infected Individuals. JAMA internal medicine.

Palmisano, L., Giuliano, M., Galluzzo, C.M., Amici, R., Andreotti, M., Weimer, L.E., Pirillo, M.F., Fragola, V., Bucciardini, R., Vella, S., 2009. The mutational archive in proviral DNA does not change during 24 months of continuous or intermittent highly active antiretroviral therapy. HIV medicine 10, 477-481.

Parisi, S.G., Andreis, S., Mengoli, C., Scaggiante, R., Ferretto, R., Manfrin, V., Cruciani, M., Giobbia, M., Boldrin, C., Basso, M., Andreoni, M., Palu, G., Sarmati, L., 2012. Baseline cellular HIV DNA load predicts HIV DNA decline and residual HIV plasma levels during effective antiretroviral therapy. Journal of clinical microbiology 50, 258-263.

Petropoulos, C., 1997. Retroviral Taxonomy, Protein Structures, Sequences, and Genetic Maps.

Phillips, A.N., Staszewski, S., Weber, R., Kirk, O., Francioli, P., Miller, V., Vernazza, P., Lundgren, J.D., Ledergerber, B., 2001. HIV viral load response to antiretroviral therapy according to the baseline CD4 cell count and viral load. Jama 286, 2560-2567.

Pienaar, E.D., Young, T., Holmes, H., 2010. Interventions for the prevention and management of oropharyngeal candidiasis associated with HIV infection in adults and children. The Cochrane database of systematic reviews, Cd003940.

Pieniazek, D., Rayfield, M., Hu, D.J., Nkengasong, J., Wiktor, S.Z., Downing, R., Biryahwaho, B., Mastro, T., Tanuri, A., Soriano, V., Lal, R., Dondero, T., 2000. Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group. Aids 14, 1489-1495.

Pillay, D., Bhaskaran, K., Jurriaans, S., Prins, M., Masquelier, B., Dabis, F., Gifford, R., Nielsen, C., Pedersen, C., Balotta, C., Rezza, G., Ortiz, M., de Mendoza, C., Kucherer, C., Poggensee, G.,

Gill, J., Porter, K., Collaboration, C.V., 2006. The impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy. AIDS 20, 21-28.

Piller, S.C., Caly, L., Jans, D.A., 2003. Nuclear import of the pre-integration complex (PIC): the Achilles heel of HIV? Current drug targets 4, 409-429.

Piot, P., Quinn, T.C., 2013. Response to the AIDS pandemic--a global health model. The New England journal of medicine 368, 2210-2218.

Poizot-Martin, I., Faucher, O., Obry-Roguet, V., Nicolino-Brunet, C., Ronot-Bregigeon, S., Dignat-George, F., Tamalet, C., 2013. Lack of correlation between the size of HIV proviral DNA reservoir and the level of immune activation in HIV-infected patients with a sustained undetectable HIV viral load for 10 years. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 57, 351-355.

Pornillos, O., Ganser-Pornillos, B.K., Kelly, B.N., Hua, Y., Whitby, F.G., Stout, C.D., Sundquist, W.I., Hill, C.P., Yeager, M., 2009. X-ray structures of the hexameric building block of the HIV capsid. Cell 137, 1282-1292.

Price, M.A., Wallis, C.L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., Sanders, E.J., Bekker, L.G., Twesigye, R., Hunter, E., Kaleebu, P., Kayitenkore, K., Allen, S., Ruzagira, E., Mwangome, M., Mutua, G., Amornkul, P.N., Stevens, G., Pond, S.L., Schaefer, M., Papathanasopoulos, M.A., Stevens, W., Gilmour, J., 2011a. Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. AIDS Res Hum Retroviruses 27, 5-12.

Price, M.A., Wallis, C.L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., Sanders, E.J., Bekker, L.G., Twesigye, R., Hunter, E., Kaleebu, P., Kayitenkore, K., Allen, S., Ruzagira, E., Mwangome, M., Mutua, G., Amornkul, P.N., Stevens, G., Pond, S.L., Schaefer, M.,

Papathanasopoulos, M.A., Stevens, W., Gilmour, J., Group, I.E.I.C.S., 2011b. Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. AIDS research and human retroviruses 27, 5-12.

Quinones-Mateu, M.E., Avila, S., Reyes-Teran, G., Martinez, M.A., 2014. Deep sequencing: becoming a critical tool in clinical virology. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 61, 9-19.

Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., et al., 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313, 277-284.

Ray, A., Ray, S., George, A.T., Swaminathan, N., 2011. Interventions for prevention and treatment of vulvovaginal candidiasis in women with HIV infection. The Cochrane database of systematic reviews, Cd008739.

Ray, A.S., Murakami, E., Basavapathruni, A., Vaccaro, J.A., Ulrich, D., Chu, C.K., Schinazi, R.F., Anderson, K.S., 2003. Probing the molecular mechanisms of AZT drug resistance mediated by HIV-1 reverse transcriptase using a transient kinetic analysis. Biochemistry 42, 8831-8841.

Recsky, M.A., Brumme, Z.L., Chan, K.J., Wynhoven, B., Yip, B., Dong, W.W., Heath, K.V., Montaner, J.S., Levy, A.R., Hogg, R.S., Harrigan, P.R., 2004. Antiretroviral resistance among HIV-infected persons who have died in British Columbia, in the era of modern antiretroviral therapy. The Journal of infectious diseases 190, 285-292.

Rhee, S.Y., Gonzales, M.J., Kantor, R., Betts, B.J., Ravela, J., Shafer, R.W., 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res 31, 298-303.

Richman, D.D., Morton, S.C., Wrin, T., Hellmann, N., Berry, S., Shapiro, M.F., Bozzette, S.A., 2004. The prevalence of antiretroviral drug resistance in the United States. AIDS 18, 1393-1401.

Russell, R.R., Bowmer, M.I., Nguyen, C., Grant, M.D., 2001. HIV-1 DNA burden in peripheral blood CD4+ cells influences disease progression, antiretroviral efficacy, and CD4+ T-cell restoration. Viral immunology 14, 379-389.

Saez-Cirion, A., Bacchus, C., Hocqueloux, L., Avettand-Fenoel, V., Girault, I., Lecuroux, C., Potard, V., Versmisse, P., Melard, A., Prazuck, T., Descours, B., Guergnon, J., Viard, J.P., Boufassa, F., Lambotte, O., Goujard, C., Meyer, L., Costagliola, D., Venet, A., Pancino, G., Autran, B., Rouzioux, C., 2013. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. PLoS pathogens 9, e1003211.

Sage, E.K., Noursadeghi, M., Evans, H.E., Parker, S.J., Copas, A.J., Edwards, S.G., Miller, R.F., 2010. Prognostic value of C-reactive protein in HIV-infected patients with Pneumocystis jirovecii pneumonia. International journal of STD & AIDS 21, 288-292.

Saitoh, A., Hsia, K., Fenton, T., Powell, C.A., Christopherson, C., Fletcher, C.V., Starr, S.E., Spector, S.A., 2002. Persistence of human immunodeficiency virus (HIV) type 1 DNA in peripheral blood despite prolonged suppression of plasma HIV-1 RNA in children. The Journal of infectious diseases 185, 1409-1416.

Samson, M., Labbe, O., Mollereau, C., Vassart, G., Parmentier, M., 1996. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. Biochemistry 35, 3362-3367.

Sanches, M., Krauchenco, S., Martins, N.H., Gustchina, A., Wlodawer, A., Polikarpov, I., 2007. Structural characterization of B and non-B subtypes of HIV-protease: insights into the natural susceptibility to drug resistance development. J Mol Biol 369, 1029-1040.

Santoro, M.M., Fabeni, L., Armenia, D., Alteri, C., Di Pinto, D., Forbici, F., Bertoli, A., Di Carlo, D., Gori, C., Carta, S., Fedele, V., D'Arrigo, R., Berno, G., Ammassari, A., Pinnetti, C., Nicastri, E., Latini, A., Tommasi, C., Boumis, E., Petrosillo, N., D'Offizi, G., Andreoni, M., Ceccherini-Silberstein, F., Antinori, A., Perno, C.F., 2014. Reliability and clinical relevance of the HIV-1 drug resistance test in patients with low viremia levels. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 58, 1156-1164.

Santos, A.F., Tebit, D.M., Lalonde, M.S., Abecasis, A.B., Ratcliff, A., Camacho, R.J., Diaz, R.S., Herchenroder, O., Soares, M.A., Arts, E.J., 2012. Effect of natural polymorphisms in the HIV-1 CRF02_AG protease on protease inhibitor hypersusceptibility. Antimicrobial agents and chemotherapy 56, 2719-2725.

Saracino, A., Gianotti, N., Marangi, M., Cibelli, D.C., Galli, A., Punzi, G., Monno, L., Lazzarin, A., Angarano, G., Mutations, S.S.G., 2008. Antiretroviral genotypic resistance in plasma RNA and whole blood DNA in HIV-1 infected patients failing HAART. Journal of medical virology 80, 1695-1706.

Savy, M., Martin-Prevel, Y., Traissac, P., Delpeuch, F., 2007. Measuring dietary diversity in rural Burkina Faso: comparison of a 1-day and a 3-day dietary recall. Public health nutrition 10, 71-78.

Schneider, E., Whitmore, S., Glynn, K.M., Dominguez, K., Mitsch, A., McKenna, M.T., 2008. Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged <18 months and for HIV infection and AIDS among children aged 18 months to <13 years--United States, 2008. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control 57, 1-12.

Schwartz, S., Felber, B.K., Benko, D.M., Fenyo, E.M., Pavlakis, G.N., 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. Journal of virology 64, 2519-2529.

Shafer, R.W., 2006. Rationale and uses of a public HIV drug-resistance database. The Journal of infectious diseases 194 Suppl 1, S51-58.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nature reviews. Genetics 15, 121-132.

Singh, K.K., Spector, S.A., 2009. Host genetic determinants of human immunodeficiency virus infection and disease progression in children. Pediatr Res 65, 55R-63R.

Sluis-Cremer, N., Arion, D., Parniak, M.A., 2000. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). Cellular and molecular life sciences : CMLS 57, 1408-1422.

Smith, M.S., Koerber, K.L., Pagano, J.S., 1993. Zidovudine-resistant human immunodeficiency virus type 1 genomes detected in plasma distinct from viral genomes in peripheral blood mononuclear cells. The Journal of infectious diseases 167, 445-448.

Soares, E.A., Santos, A.F., Sousa, T.M., Sprinz, E., Martinez, A.M., Silveira, J., Tanuri, A., Soares, M.A., 2007. Differential drug resistance acquisition in HIV-1 of subtypes B and C. PloS one 2, e730.

Solbak, S.M., Reksten, T.R., Hahn, F., Wray, V., Henklein, P., Henklein, P., Halskau, O., Schubert, U., Fossen, T., 2013. HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. Biochimica et biophysica acta 1828, 816-823.

Somda, A., Sangare, L., Soro, M., Yameogo, S., Bazie, B., Bigirimana, F., Bertagnolio, S., Peeters, M., Mouacha, F., Rivera, A.M., Jordan, M.R., Sanou, M.J., 2012. Surveillance of transmitted drug-resistant HIV among young pregnant women in Ouagadougou, Burkina Faso. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 54 Suppl 4, S317-319.

Ssemwanga, D., Lihana, R.W., Ugoji, C., Abimiku, A., Nkengasong, J., Dakum, P., Ndembi, N., 2014. Update on HIV-1 Acquired and Transmitted Drug Resistance in Africa. AIDS reviews 17.

Stevens, G.A., Finucane, M.M., De-Regil, L.M., Paciorek, C.J., Flaxman, S.R., Branca, F., Pena-Rosas, J.P., Bhutta, Z.A., Ezzati, M., Nutrition Impact Model Study, G., 2013. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data. Lancet Glob Health 1, e16-25.

Sullivan, P., 2002. Associations of anemia, treatments for anemia, and survival in patients with human immunodeficiency virus infection. The Journal of infectious diseases 185 Suppl 2, S138-142.

Takuva, S., Maskew, M., Brennan, A.T., Sanne, I., Macphail, A.P., Fox, M.P., 2013. Anemia among HIV-Infected Patients Initiating Antiretroviral Therapy in South Africa: Improvement in Hemoglobin regardless of Degree of Immunosuppression and the Initiating ART Regimen. Journal of tropical medicine 2013, 162950.

Tambuyzer, L., Azijn, H., Rimsky, L.T., Vingerhoets, J., Lecocq, P., Kraus, G., Picchio, G., de Bethune, M.P., 2009. Compilation and prevalence of mutations associated with resistance to non-nucleoside reverse transcriptase inhibitors. Antiviral therapy 14, 103-109.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24, 1596-1599.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30, 2725-2729.

Tebit, D.M., Ganame, J., Sathiandee, K., Nagabila, Y., Coulibaly, B., Krausslich, H.G., 2006. Diversity of HIV in rural Burkina Faso. J Acquir Immune Defic Syndr 43, 144-152.

Tebit, D.M., Sangare, L., Makamtse, A., Yameogo, S., Somlare, H., Bado, G., Kouldiaty, B.G., Sathiandee, K., Tiba, F., Sanou, I., Ouedraogo-Traore, R., Zoungrana, L., Diallo, I., Drabo, J.Y., Krausslich, H.G., 2008. HIV drug resistance pattern among HAART-exposed patients with suboptimal virological response in Ouagadougou, Burkina Faso. Journal of acquired immune deficiency syndromes (1999) 49, 17-25.

Tebit, D.M., Sangare, L., Tiba, F., Saydou, Y., Makamtse, A., Somlare, H., Bado, G., Kouldiaty, B.G., Zabsonre, I., Yameogo, S.L., Sathiandee, K., Drabo, J.Y., Krausslich, H.G., 2009. Analysis of the diversity of the HIV-1 pol gene and drug resistance associated changes among drug-naive patients in Burkina Faso. Journal of medical virology 81, 1691-1701.

Tshabalala, M., Manasa, J., Zijenah, L.S., Rusakaniko, S., Kadzirange, G., Mucheche, M., Kassaye, S., Johnston, E., Katzenstein, D., 2011. Surveillance of transmitted antiretroviral drug resistance among HIV-1 infected women attending antenatal clinics in Chitungwiza, Zimbabwe. PloS one 6, e21241.

Tuboi, S.H., Brinkhof, M.W., Egger, M., Stone, R.A., Braitstein, P., Nash, D., Sprinz, E., Dabis, F., Harrison, L.H., Schechter, M., 2007. Discordant responses to potent antiretroviral treatment in previously naive HIV-1-infected adults initiating treatment in resource-constrained countries:

the antiretroviral therapy in low-income countries (ART-LINC) collaboration. Journal of acquired immune deficiency syndromes 45, 52-59.

Turner, D., Brenner, B., Wainberg, M.A., 2003. Multiple effects of the M184V resistance mutation in the reverse transcriptase of human immunodeficiency virus type 1. Clinical and diagnostic laboratory immunology 10, 979-981.

Tyagi, S., Kramer, F.R., 1996. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14, 303-308.

UNAIDS, 2013a. ACCESS TO ANTIRETROVIRAL THERAPY IN AFRICA. 2-3.

UNAIDS, 2013b. Report on the global AIDS epidemic 2013. 4-6.

UNAIDS, 2014a. THE GAP REPORT. 26-37.

UNAIDS, 2014b. UNAIDS fact sheet.
http://www.unaids.org/en/resources/campaigns/2014/2014gapreport/factsheet.

van der Sande, M.A., Schim van der Loeff, M.F., Aveika, A.A., Sabally, S., Togun, T., Sarge-Njie, R., Alabi, A.S., Jaye, A., Corrah, T., Whittle, H.C., 2004. Body mass index at time of HIV diagnosis: a strong and independent predictor of survival. Journal of acquired immune deficiency syndromes 37, 1288-1294.

Vandegraaff, N., Kumar, R., Burrell, C.J., Li, P., 2001. Kinetics of human immunodeficiency virus type 1 (HIV) DNA integration in acutely infected cells as determined using a novel assay for detection of integrated HIV DNA. Journal of virology 75, 11253-11260.

Velazquez-Campoy, A., Todd, M.J., Vega, S., Freire, E., 2001. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proc Natl Acad Sci U S A 98, 6062-6067.

Venkataraman, K., 2013. A study of biological markers in HIV disease progression and management in the highly active antiretroviral therapy (HAART) era. American Journal of Bioscience and Bioengineering 1, 24-37.

Vergne, L., Diagbouga, S., Kouanfack, C., Aghokeng, A., Butel, C., Laurent, C., Noumssi, N., Tardy, M., Sawadogo, A., Drabo, J., Hien, H., Zekeng, L., Delaporte, E., Peeters, M., 2006. HIV-1 drug-resistance mutations among newly diagnosed patients before scaling-up programmes in Burkina Faso and Cameroon. Antiviral therapy 11, 575-579.

Verhofstede, C., Reniers, S., Van Wanzeele, F., Plum, J., 1994. Evaluation of proviral copy number and plasma RNA level as early indicators of progression in HIV-1 infection: correlation with virological and immunological markers of disease. AIDS 8, 1421-1427.

Vicenti, I., Razzolini, F., Saladini, F., Romano, L., Zazzi, M., 2007. Use of peripheral blood DNA for genotype antiretroviral resistance testing in drug-naive HIV-infected subjects. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 44, 1657-1661.

von Kleist, M., Metzner, P., Marquet, R., Schutte, C., 2012. HIV-1 polymerase inhibition by nucleoside analogs: cellular- and kinetic parameters of efficacy, susceptibility and resistance selection. PLoS computational biology 8, e1002359.

Vyas, S., Kumaranayake, L., 2006. Constructing socio-economic status indices: how to use principal components analysis. Health policy and planning 21, 459-468.

Wainberg, M.A., Zaharatos, G.J., Brenner, B.G., 2011. Development of antiretroviral drug resistance. The New England journal of medicine 365, 637-646.

Watanabe, D., Ibe, S., Uehira, T., Minami, R., Sasakawa, A., Yajima, K., Yonemoto, H., Bando, H., Ogawa, Y., Taniguchi, T., Kasai, D., Nishida, Y., Yamamoto, M., Kaneda, T., Shirasaka, T., 2011. Cellular HIV-1 DNA levels in patients receiving antiretroviral therapy strongly correlate with therapy initiation timing but not with therapy duration. BMC infectious diseases 11, 146.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711-716.

Weiss, R.A., 2013. Thirty years on: HIV receptor gymnastics and the prevention of infection. BMC biology 11, 57.

Wensing, A.M., Calvez, V., Gunthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., Shafer, R.W., Richman, D.D., 2014. 2014 Update of the drug resistance mutations in HIV-1. Topics in antiviral medicine 22, 642-650.

WHO, 1993. Proposed 'World Health Organization staging system for HIV infection and disease': preliminary testing by an international collaborative cross-sectional study. The WHO International Collaborating Group for the Study of the WHO Staging System. AIDS 7, 711-718.

WHO, 2005. BURKINA FASO. Estimated number of people needing antiretroviral therapy (0-49 years), 2005:Antiretroviral therapy target declared by count.

WHO, 2006. Surveillance of HIV Drug Resistance. Report of a Workshop, Pune, India, 10.

WHO, 2007. WHO Case Definitions of HIV for Surveillance and Revised Clinical Staging and Immunological Classification of HIV-related Disease in Adults and Children. 10-18.

WHO, 2011. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.Vitamin and Mineral Nutrition Information System. Geneva, World Health Organization, 2011.

WHO, 2012. WHO Drug Resistance Report. 2012. 5-8.

WHO, 2013a. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach, Geneva.

WHO, 2013b. HIV/AIDS Surveillance Report in Europe. 2012.

Williams, J.P., Hurst, J., Stohr, W., Robinson, N., Brown, H., Fisher, M., Kinloch, S., Cooper, D., Schechter, M., Tambussi, G., Fidler, S., Carrington, M., Babiker, A., Weber, J., Koelsch, K.K., Kelleher, A.D., Phillips, R.E., Frater, J., Investigators, S.P., 2014. HIV-1 DNA predicts disease progression and post-treatment virological control. eLife 3, e03821.

Wittkop, L., Gunthard, H.F., de Wolf, F., Dunn, D., Cozzi-Lepri, A., de Luca, A., Kucherer, C., Obel, N., von Wyl, V., Masquelier, B., Stephan, C., Torti, C., Antinori, A., Garcia, F., Judd, A., Porter, K., Thiebaut, R., Castro, H., van Sighem, A.I., Colin, C., Kjaer, J., Lundgren, J.D., Paredes, R., Pozniak, A., Clotet, B., Phillips, A., Pillay, D., Chene, G., 2011. Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. The Lancet. Infectious diseases 11, 363-371.

Zanini, F., Neher, R.A., 2013. Quantifying selection against synonymous mutations in HIV-1 env evolution. Journal of virology 87, 11843-11850.

Zhang, F., Zhu, H., Wu, Y., Dou, Z., Zhang, Y., Kleinman, N., Bulterys, M., Wu, Z., Ma, Y., Zhao, D., Liu, X., Fang, H., Liu, J., Cai, W.P., Shang, H., 2014. HIV, hepatitis B virus, and hepatitis C virus co-infection in patients in the China National Free Antiretroviral Treatment Program, 2010-12: a retrospective observational cohort study. The Lancet. Infectious diseases 14, 1065-1072.