

https://dspace.org/

Population dynamics in HIV-1 transmitted antiretroviral drug resistance

Harris, Dean Mark

2018
https://hdl.handle.net/10539/25353

Downloaded from a DSpace repository. https://dspace.org/

Population Dynamics in HIV-1 Transmitted Antiretroviral Drug Resistance

Dean Mark Harris



A dissertation submitted to Faculty of Health Sciences,

University of the Witwatersrand,

in fulfilment of the requirements for the degree of

Master of Science in Medicine

Johannesburg, June 2018

Declaration

I, Dean Mark Harris, hereby declare that this research report is my own, unaided work. It is being submitted in part for the Degree of Master of Science in Medicine at the University of the Witwatersrand Medical School, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

Dean Mark Harris

_____day of_____ 2018

Abstract

It is well known that antiretroviral (ARV) drug resistant variants of HIV-1 can be sexually transmitted. Several studies have shown that in resource-rich geographical locations as many as 15-20% of individuals are newly infected with HIV-1 containing at least one drug resistant mutation. In contract, resource limited geographical locations, such as Sub-Saharan Africa, have shown prevalences in the range of 5 to 10%. Since the ART rollout in these resource-limited locations are generally not well monitored with virological genotyping, the transmission of drug resistant HIV-1 is likely to increase, with significant clinical and public health consequences.

HIV-1 transmission is characterised by the transmission of a single founder virus, or narrow spectrum of founder viruses, that develop into the viral quasispecie. It is unlikely that drug resistant virus will coexist with wild type (wt) virus, in the case of non-drug resistance transmission. However, initiating in ARV treatment, drug non-adherence may select of ARV drug resistance mutations and may subsequent lead to treatment failure. Drug resistant virus may be transmitted to a new host, as drug resistant mutations do not appear to hamper transmission efficiency of the mutated virus. Several studies have shown that transmitted drug resistance mutations (TDRMs) persist either as the dominant species or as minority variants, or revert to wild type over time, in the absence of drug pressure. It is generally acknowledged that many drug resistance mutations decrease the replicative capacity of HIV-1, and thus reversion confers a potential survival advantage.

Because of the emergence of wild type variants from TDRM quasispecies requires evolution and back-mutation, the rate at which individual TDRMs become undetectable may vary substantially. Contradictory findings of persistence versus reversion of TDRMs have been reported, and may be attributed to the fact that minority variants are difficult to detect by conventional population based Sanger sequencing, and patient numbers studied are small. Consequently, individuals infected with HIV-1 harbouring TDRM have a higher chance of failing their first-line therapy. Understanding the population dynamics of transmitted drug resistant HIV-1 in the absence of drug pressure is essential for clinical management and public health strategies.

The individuals identified with TDRMs from the IAVI-Early Infections Cohort (Protocol C) provides a unique research opportunity to address the aforementioned issue. This study describes

the evolutionary mechanisms of ARV drug resistant HIV-1 after transmission to a new host to provide insight into persistence and/or rates of reversion to wild type.

TDRMs initially identified by Price et al. (2011) in the IAVI-Early Infections Cohort (Protocol C) using population-based Sanger sequencing (the current diagnostic gold standard), were confirmed in this study by newer ultra-deep next generation sequencing (NGS) technology on the Illumina Miseq platform. Longitudinal samples were made available for individuals in which transmitted drug resistance were identified, and we also sequenced using NGS on the Illumina Miseq platform. Additional minority variants (present at <20% of the sequenced viral population) were identified by NGS. This study found a large percentage of TDRMs to persist for a significant amount of time after transmission to a new, drug naïve host, in the longitudinal samples. The level of persistence, or rate of reversion of TDRMs, appear to be subject to the type of resistance (NRTI, NNRTI or PI), level of resistance the mutation confers, as well as the combination of mutations that are co-transmitted. Findings of this study highlight the importance of drug resistance screening prior to ART initiation, as well as the importance of the drug resistance screening assay sensitivity. As rates of transmitted drug resistance are increasing in developing countries of which the IAVI-Early Infections Cohort (Protocol C) are composed of, understanding the population dynamics of transmitted drug resistant HIV-1 in the absence of drug pressure is essential for clinical management, public health strategies and informing future vaccine design.

Acknowledgments

I would to thank the funding bodies for allowing this research to be undertaken. The work in this thesis is based on the research supported in part by the National Research Foundation of South Africa (Grant Numbers 86055 and 93462). Additionally, to the University of the Witwatersrand for providing me with a Postgraduate Merit Award, and the Poliomyelitis Research Foundation and National Research Foundation for bursary funding.

Thank you to the International Aids Vaccine Initiative (IAVI) and Dr Matt Price, for providing samples from the protocol C cohort to be studied and enabling us to perform this research.

To my supervisor, Prof Maria Papathanasopoulos, thank you for your guidance, mentorship, expertise, understanding and advice throughout this project. The life experience and opportunities that were made available to me during this journey has helped me grow enormously as a person, enriching me in all areas of my life and career. I am deeply appreciative and grateful for the experiences and perspective I have gained.

To my co-supervisor Prof Simon Travers, thank you for your guidance and assistance during this project. The brief time that I spent at SANBI was tremendously exciting and enriching, and stimulated new ways of thinking and problem solving. Your hospitality, insight, and assistance is greatly valued and appreciated.

To my second co-supervisor Dr Irene Ketseoglou, thank you for your advice, insight, training and introduction and guidance through the daunting field of next generation sequencing. It is a skillset that I greatly value and hope to further develop in the future.

To my friends and colleges at the HIV Pathogenesis Research Unit, both past and present, thank you for creating a positive environment that stimulates interesting thoughts and discussions. A special thank you to Mark for the advice, guidance, mentorship and willingness to help. Your perspective has been of great value to me. To Roberto, thank you for being a true friend, always engaging in interesting conversation, debate and banter.

To my wife Gabriella, thank you for your unconditional support and love. In difficult times you are always a shining light of advice and clarity. Always willing to go the extra mile and assist me where you can. Your companionship and love during this project is deeply appreciated.

Contents

Declaration I

Abstract II

Acknowledgments IV

List of Figures 1

List of Tables 5

List of Abbreviations 6

CHAPTER 1: Introduction 9

1.1 HIV Life Cycle 10

1.2 Natural Disease Progression of HIV-1 Infection 12

1.3 HIV-1 Treatment 14

1.4 Antiretroviral drug therapy (ART) 15

1.4.1 Entry Inhibitors 17

1.4.2 Nucleoside Reverse Transcriptase Inhibitors (NRTIs)/ Nucleotide Reverse Transcriptase Inhibitors (NtRTIs). 17

1.4.3 Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) 18

1.4.4 Integrase Inhibitors 19

1.4.5 Protease Inhibitors (PIs) 20

1.4.6 HIV-1 Viral Diversity and Quasispecies 20

1.5 HIV-1 Drug Resistance 21

1.5.1 Mechanisms of Drug Resistance 22

1.5.2 Types of Drug Resistance 25

1.5.3 Persistence and/or reversion of drug resistant mutations 29

1.6 HIV-1 Viral Genotyping and Drug Resistance Detection 31

1.7 Rationale, Aims and Objectives of the study 43

CHAPTER 2: Materials and Methods 44

2.1 International Aids Vaccine Initiative (IAVI) Protocol C Cohort 44

2.2 Viral RNA Extraction, Reverse Transcription and Nested Polymerase Chain Reaction (RT-PCR) Amplification 44

2.2.1 Viral RNA Extraction 44

2.2.2 Reverse Transcription PCR (RT-PCR) 45

2.2.3 Nested PCR of RT-PCR Products 45

2.3 Next Generation Sequencing on the Illumina Miseq Platform 46

2.3.1 Amplicon Quantification 46

2.3.2 Sample/Library preparation 47

2.3.2.1 Input DNA Fragmentation (Tagmentation) 47

2.3.2.2 Sample Indexing 47

2.3.2.3 Fragment Size Selection 49

2.3.2.5 Library Normalisation 52

2.3.3 Library pooling and denaturing 53

2.3.4 Clustering and Sequencing 54

2.4 Data Analysis 56

CHAPTER 3: Results 59

3.1 IAVI Protocol C Cohort Demographic and Clinical Data 59

3.2 Viral RNA Extraction and RT-PCR Amplification 59

3.3 Illumina Miseq Nextera XT Library preparation 60

3.3.1 Amplicon Quantification 60

3.3.2 DNA Fragmentation and Size Selection 61

3.4 Illumina Miseq Next Generation Sequencing 64

3.5 Exatype Results 65

3.6 Phylogenetic Analysis 69

3.7 Antiretroviral Drug Resistance Mutation Results 71

3.7.1 Conformation of Sanger Data by NGS, and Persistence and Reversion of TDRMs identified at baseline 71

3.7.2 Longitudinal Persistence and Reversion Analysis of TDRMs 75

CHAPTER 4: Discussion and Conclusion 83

4.1 Discussion 83

4.2 Conclusion 106

5. References 109

6. Appendices 121

Appendix A: Ethics Clearance 121

Appendix B: DNA Work 122

1. GeneJet DNA (Thermo Fisher Scientific, Massachusetts, USA) purification kit protocol ……..122

2. Agarose Gel Preparation 123

3. Qubit 2.0 Fluorometer High Sensitivity Assay 123

Appendix C: Nextera XT library preparation guide 126

Appendix D: IAVI Protocol C Demographics and DRM Data 134

Appendix E: Turnitin Report 146

List of Figures

Figure 1.1: A schematic representation of the replicative life cycle of HIV-1, highlighting the main stages of the life cycle. The various stages of the HIV-1 life cycle (A to G) correspond to descriptions in the text below (Copied from Laskey & Siliciano, 2014)…….10

Figure 1.2: Changes in plasma viral load (RNA copies/ml) and CD4+ T cells (cells/mm3) post HIV-1 transmission. An initial spike in viral load and sudden decline in CD4+ T cells is observed during the first several weeks of acute HIV-1 infection. During the asymptomatic clinical latency phase, a slow decline of CD4+ T cells occurs as HIV-1 viral load gradually increases. Without treatment, a state of immunodeficiency is achieved. (Adapted from Fauci & Pantaleo, 1996)……..12

Figure 1.3: A schematic representation of the HIV-1 lifecycle and the target sites of ARVs. Steps A to D correspond to stages of the HIV-1 life cycle that are targeted by ARV drugs, and are described below…….....16

Figure 1.4: The molecule structures of the nucleotide adenine shown on the left, and the nucleotide analogue emtricitabine shown on the right. Black arrows indicate where, using Emtricitabine as an example, the hydroxyl group is removed for NRTIs that result in DNA polymerization termination. Red arrows indicate where the phosphate group is replaced with an OH group for nucleoside reverse transcriptase inhibitors (NRTIs) (Adapted from Esposito et al., 2012)……..17

Figure 1.5: A schematic representation of HIV-1 quasispecies that are each genetically linked to the initial founder virus sequence. The black dot represents the founder virus, which soon diversifies to establish an initial quasispecies, followed by further diversification and growth of the quasispecies. Each branch in the figure represents two variants connected by a point mutation, and the concentric circles each represent a replication cycle (copied from Lauring and Andino, 2010)……..20

Figure 1.6: An illustration of the process followed from instrument to a clinical result that reveals the drug susceptibility or resistance of virus sequenced from an HIV-1 infected individual……..31

Figure 1.7: A schematic representation of bridge amplification. Bridge amplification is the process in which fragments are isothermally amplified multiple times on the surface of the sequencing flow cell (Adapted from Illumina)……...36

Figure 1.8: The 5' to 3'direction of the first sequencing read. The first read sequencing primer binds to the primer binding site, and produces the first read by incorporating nucleotides in the sequencing cycles that follow…….37

Figure 1.9: The production of the second sequencing read. The second read is produced by reading the second index primer first, and then proceeding the sequencing reaction……..38

Figure 2.1: Schematic representation of Illumina paired end sequencing adapters added to both ends of the cDNA fragments. Forward reads are indicated by an orange arrow moving in a 5' to 3' direction. Reverse reads are indicated by a blue arrow moving in a 5' to 3'on the complementary strand (Adapted from http://www.gendx.com).....47

Figure 2.2: A range of solid phase reverse interphase (SPRI) bead ratios (0.4 to 2.5 times SPRI bead volume to DNA sample volume) were used to purify fragmented DNA samples, and the resulting purified fragments were resolved on a 2% agarose gel (Connolly et al., 2010)…….....48

Figure 2.3: The series of library dilutions, starting at 2 nM, performed to obtain the final concentration of 7 pM suitable for Miseq loading and subsequent sequencing……... 52

Figure 2.4: Bioinformatics pipeline utilised for processing raw HIV-1 pol sequence data to usable antiretroviral drug resistance mutation prevalence, and construction of phylogenetic trees for phylogenetic analysis…….55

Figure 3.1: RT-PCR amplification of the pol region (~1800 bp) from representative extracted HIV-1 viral RNA from the IAVI Protocol C cohort, electrophoresed on a 1% agarose gel under ultraviolet (UV) trans-illumination. Amplicon sizes were determined utilising the molecular weight marker (MWM, O'GeneRuler DNA Ladder Mix #SM1173, Thermo Fisher Scientific, Massachusetts, USA) basepairs (bp) depicted on the left lane. Participant number and visit dates are assigned to each respective lane…….59

Figure 3.2: Electropherogram showing library fragment sizes for sample five from participant one. The electropherogram graphically illustrates the spread of fragments selected for by a double size selection using a 1x and 0.5x SPRI to DNA bead ratio, respectively. The ratios yielded a peak fragment size of 252 bp and an average library fragment size of 280 bp……....60

Figure 3.3: Electropherogram showing library fragment sizes for sample 47 from participant seven. The electropherogram graphically illustrates the spread of fragments selected for by a single size selection using a 0.5x SPRI to DNA ratio. The single selection yielded a peak fragment size of 780 bp and an average library fragment size of 712 bp……...…….61

Figure 3.4: Miseq sequencing reads generated in runs 1 to 4, for samples 1 to 159. The four graphs illustrate the number of forward and reverse sequencing reads, as well as unaligned and low-quality sequencing reads generated for each sample, per run. Run one to four are shown in graphs A to D, respectively……..…65

Figure 3.5: The average number of read coverage for the region of pol that contains *prot* and *RT*, for all samples (1 to 159) sequenced in the 4 sequencing runs. Coverage in number of reads, per run, relative to the codon position of pol obtained for all samples is shown. The reads that fall within the *prot* and *RT* codons are coloured in red and blue, respectively……...67

Figure 3.6: Neighbour-joining tree (Kimura-80 model with 1000 bootstrap) of all 23 participants. Participant samples are shown in blue, Subtype A-K reference genomes are shown in red, and background noise HIV-1 samples are shown in black. Sub-type branches, with each participant's cluster are annotated on the tree……..68

Figure 3.7.1: Viral load against follow up period in days for participants in which complete reversion was observed. Participant seven and 21 are shown in blue (A) and red (B) respectively. A TDRM is written in red where it was detected below 1%. Reversion of TDRMs for both participant seven and 21 occurred within 300 days after the baseline sample……..74

Figure 3.7.2: Viral load against follow up period in days for participants in which both persistence and reversion of TDRMs was observed. Participant 15 (A) and participant three (B) are shown in red and blue, respectively. TDRMs are written in red where detected below 1%.....76

Figure 3.8: Viral load changes during the follow up period (in days) for participants in which no TDRM reversion was observed. The various participants are each shown in a different colour, and participants followed for less than 800 days are shown in A, and more than 800 days are shown in B…….78

List of Tables

Table 1.1: Five classes of United States Food and Drug Administration approved Antiretroviral drugs……...15

Table 1.2: Summary of commonly used next generation sequencing technologies (adapted from Gibson et al., 2014)…….34

Table 3.1: Demographic and clinical data of 23 participants from the IAVI Protocol C Cohort identified to harbour HIV-1 with TDRMs…….58

Table 3.2: Miseq sequencing metrics recorded for the sequencing runs one to four…….63

Table 3.3: Persistence or reversion of HIV-1 transmitted drug resistance mutations in 23 participants from the IAVI Protocol C Cohort, showing clinical parameters, subtype, ARV drug resistance mutations (% of amplified sequenced virus) detected by NGS, and mutations detected by Sanger Sequencing……70 – 72

Table 3.4: Summary table of TDRMs detected in participants by both NGS and Sanger, and the TDRM persistence/reversion outcome…….….79-80

Appendix D:

Table 1: IAVI Protocol C Clinical and Demographic Data with baseline mutations detected by Sanger sequencing……....131

Table 2: IAVI Protocol C Drug Resistant Mutation Results Detected for All Samples……132-142

List of Abbreviations

3TC	Lamivudine
ABC	Abacavir
ADR	Acquired Drug Resistance
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral
ATM	Amplicon Tagment Mix
ATV	Atazanavir
AZT/ZDV	Zidovudine
bp	Base Pairs
CCR5	C-C Chemokine Receptor Type 5
cDNA	Complementary DNA
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
CXCR4	C-X-C Chemokine Receptor Type 4
d4T	Stavudine
ddi	Didanosine
DRM	Drug Resistance Mutation
DTG	Dolutegravir
EFV	Efavirenz
EVG	Elvitegravir

ETR	Etravirine
FDA	Food and Drug Administration
FTC	Emtricitabine
GB	Gigabyte
IDV	Indinavir
int	Integrase
HIV-1	Human Immunodeficiency Virus Type 1
HIV-2	Human Immunodeficiency Virus Type 2
HAART	Highly Active Antiretroviral Therapy
INSTIs	Integrase Strand Transfer Inhibitors
pH	Potential of Hydrogen
pol	Polymerase
PrEP	Pre-exposure Prophylaxis
pro	Protease
RAL	Raltegravir
RAM	Resistance Associated Mutation
RT	Reverse Transcriptase
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RTV	Ritonavir
RVP	Rilpivirine
SMRT	Single Molecule Real Time

SPRI	Solid Phase Reverse Immobilisation
SQV	Saquinavir
LPV	Lopinavir
MTCT	Mother-to-Child Transmission
NFV	Nelfinavir
NGS	Next Generation Sequencing
NRTIs	Nucleoside Reverse Transcriptase Inhibitors
NtRTIs	Nucleotide Reverse Transcriptase Inhibitors
NVP	Nevirapine
NT	Neutralize Tagment
TD	Tagment DNA Buffer
TDF	Tenofovir
TDR	Transmitted Drug Resistance
TDRM	Transmitted Drug Resistance Mutation(s)
WHO	World Health Organization
WT	Wild Type

CHAPTER 1: Introduction

Infection with Human Immunodeficiency Virus (HIV) is one of the most significant, confounding and demanding medical challenges the world is currently facing. The virus is classified under the lentivirus genus, and is part of the retrovirus family (Fauci, 1988). Two types of HIV have been classified; HIV type 1 (HIV-1) and HIV type 2 (HIV-2), the former being the most prominent global circulating type with the highest prevalence, infectivity and virulence of the two virus types. The United Nations Programme on HIV/AIDS (UNAIDS) statistics, published at the end of 2015, reveal that a total of 36.7 million people are globally infected with HIV-1, with 2.1 million new infections and 1.1 million acquired immune deficiency syndrome (AIDS) related deaths in 2015 (UNAIDS, 2016b). The majority of global HIV-1 infections are located in sub-Saharan Africa, accounting for over 69.5% (25.5 million individuals) of global HIV-1 infection (UNAIDS, 2016b). Statistics provided by Statistics South Africa shows that the estimated HIV-1 prevalence rate in South Africa is 12.7%, amongst the highest in the world. The total number of individuals that are currently infected with HIV-1 in South Africa is 7.03 million individuals (Statistics-SA, 2014). Since the start of the HIV-1 epidemic, over 78 million individuals have been infected by the virus, and more than 39 million individuals have died in AIDS related deaths (UNAIDS, 2016b).

1.1 HIV Life Cycle

The HIV viral life cycle goes through several distinct stages, producing infectious HIV virions shown in **Figure 1.1.**

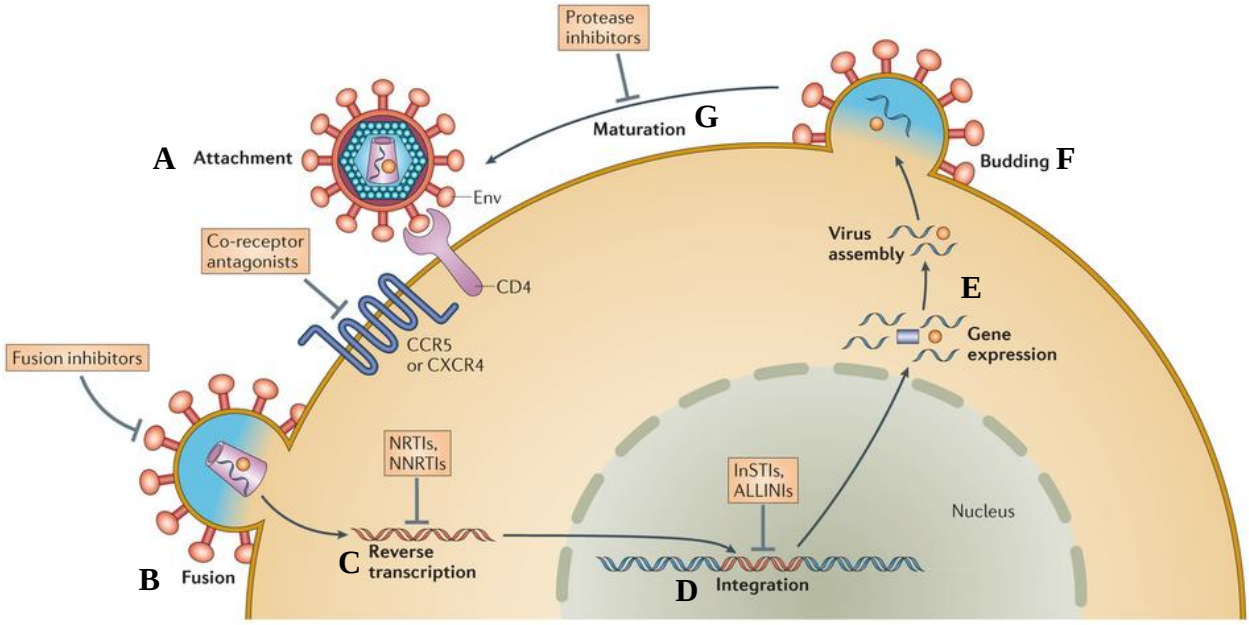


Figure 1.1: A schematic representation of the replicative life cycle of HIV-1, highlighting the main stages of the life cycle. The various stages of the HIV-1 life cycle (A to G) correspond to descriptions in the text below (Copied from Laskey & Siliciano, 2014).

The primary and preferential target cell of HIV-1 is the CD4+ T-cell. The CD4+ T-cell plays a crucial role in the immune protection, by regulating several vital immune functions. They regulate functions such as activating B cells to produce antibodies, inducing macrophage cells to develop and enhanced microbicidal activity. Furthermore CD4+ T-cells recruit a range of cells such as neutrophils, eosinophils, and basophils to sites of infection and inflammation. CD4+ T-cells additionally mediate the production of cytokines and chemokines (Mosmann, Cherwinski, Bond, Giedlin, & Coffman, 1986).

The viral surface protein gp120 binds to a glycoprotein called Cluster of Differentiation 4 (CD4) that is located on the of surface of the target CD4+ T-cells (shown as **A** in **Figure 1.1**). CD4 functions in conjunction with the T cell receptor (TCR) as a co-receptor, and furthermore assists

TCR to communicate with antigen-presenting cells. CD4 and TCR are constituent parts of the T cell receptor complex. The main natural target of CD4 is a distinct region of the antigen-presenting MHCII molecule. An extracellular domain of CD4, D1, binds to a region termed β2 on MHCII molecules. CD4 is close in proximity both intra- and extracellularly to TCR, which allows for the activation of motifs located in the cytoplasmic regions of the complex to be phosphorylated. Once phosphorylated, it allows the cytoplasmic motifs to amplify signals that are generated by the TCR complex. Depending on the antigen the TCR complex interacts with, different signals may be produced and thus amplified, which in turn may signal specific signal dependent T helper cells.

Once the HIV-1 gp120 interacts and binds with CD4, a conformational change in the gp120-CD4 complex is induced, which in turn allows for the gp120 viral protein to successfully make contact and interact with the co-receptor CCR5 (S. Miller, 2002). An additional conformational change in which leads to hydrophobic portions of gp41 to be exposed is caused by the gp120-CCR5 interaction. A six-helix bundle is additionally subsequently formed. The conformational changes described above brings the target cell and the viral surfaces closer to one another, and ultimately leads to fusion (shown as **B** in **Figure 1.1**) of HIV-1 with the target cell (S. Miller, 2002). Once fusion has taken place, gp120 is shed into the peripheral blood of the infected individual, which then acts as an immune decoy (Costin, 2007).

Following fusion, the core of the virus enclosed by p24 viral proteins, is released into the host cell cytoplasm where uncoating occurs. The uncoating event leads to the release of the HIV-1 RNA genome, viral proteins reverse transcriptase (RT) and integrase (IN) and several additional regulatory proteins. Shortly after uncoating, reverse transcription by RT occurs in the host cell's cytoplasm where the RNA viral genome is transcribed into viral complementary DNA (cDNA) (shown as **C** in **Figure 1.1**). A pre-integration complex is formed after reverse transcription, which is composed of the viral cDNA and a range of viral proteins (including Vpr and IN) and host proteins. The HIV-1 pre-integration complex is transported into the host cell nucleus by the viral assessor protein Vpr (Costin, 2007). Once in the nucleus, the complex is permanently integrated into the host cell's genome by the HIV-1 viral protein IN (shown as **D** in **Figure 1.1**). The Vpr protein first binds to the long terminal repeat (LTR) regions of the reverse transcribed viral DNA, whereby it then catalyses the reaction in which the 3' terminal of the viral DNA is cleaved and subsequently transferred via strand transfer into the host cell genome (Hazuda, 2009).

As the infected host cell assumes normal function and progresses through its cell growth cycle, the integrated proviral genome is actively transcribed (S. Miller, 2002). Products of proviral transcription, messenger RNA (mRNA), is spliced into several variants, after which the variants are transported and translated in the host cell cytoplasm (shown as **E** in **Figure 1.1**). An important viral enzyme that is transcribed and translated at this point is protease (PR). PR has a critically important function in the process in the maturation of new HIV-1 virions. PR additionally influences all constituent viral components that have been transcribed and translated to assemble, and to furthermore interact to the host cell plasma membrane. Shortly after assembly, budding of immature HIV-1 virions occur (shown as **F** in **Figure 1.1)**, which leads to the release of newly assembled HIV-1 virions**.** The PR enzyme continues the maturation of the HIV-1 virions post budding (shown as **G** in **Figure 1.1**) to ensure that the newly produced viruses will have the full capacity to infect additional target cells (Costin, 2007; S. Miller, 2002).

1.2 Natural Disease Progression of HIV-1 Infection

HIV-1 primarily and preferentially infects cells that express the cell surface glycoprotein CD4, such as $CD4^+$ T cells. $CD4^+$ T cells are a crucial component of an individual's immune system, and make up parts of both the innate and adaptive immunity (Douek et al., 2002; Zhu & Paul, 2008). HIV-1 infection causes target cell death and thus leads to the decline in the levels of $CD4^+$ T-cells (Fink & Cookson, 2005). This results in the immune system being unable to effectively contain pathogenic infections (Perelson, Kirschner, & De Boer, 1993). Advanced stages of HIV-1 infection leads to the progressive weakening of the immune system, primarily due to the low levels of $CD4^+$ T-cells, with the infected individual ultimately advancing to a state of immunodeficiency, as graphically represented below in **Figure 1.2.**

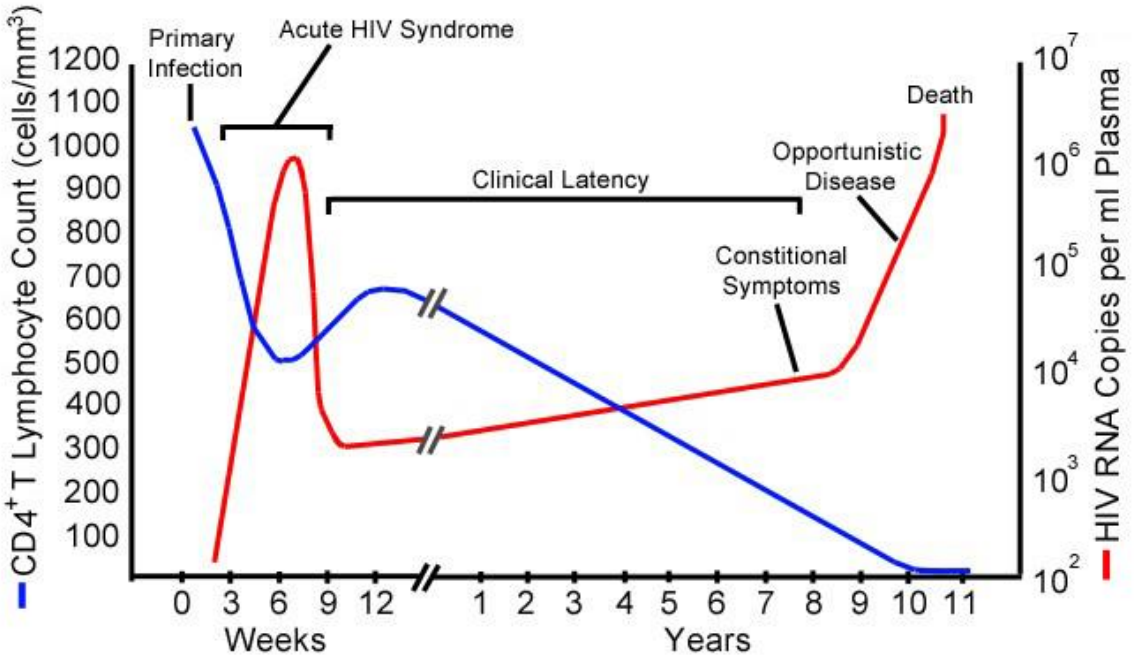


Figure 1.2: Changes in plasma viral load (RNA copies/ml) and CD4$^+$ T cells (cells/mm3) post HIV-1 transmission. An initial spike in viral load and sudden decline in CD4$^+$ T cells is observed during the first several weeks of acute HIV-1 infection. During the asymptomatic clinical latency phase, a slow decline of CD4$^+$ T cells occurs as HIV-1 viral load gradually increases. Without treatment, a state of immunodeficiency is achieved (Adapted from Fauci & Pantaleo, 1996).

The initial spike in viral load and decline of CD4$^+$ T cells observed within the first weeks of infection is followed by a partial recovery of CD4$^+$ T cells as the infected individual's immune system produces anti-HIV-1 cellular responses to partly contain the infection. Cellular responses involve activity of CD4$^+$ and CD8$^+$ T (otherwise known as cytotoxic T lymphocytes) cells. The CD8$^+$ T cells act against HIV-1 during acute HIV-1 infection by eliminating HIV-1 infected cells and producing anti-HIV-1 chemokines (Walker, Moody, Stites, & Levy, 1986). The humoral response to HIV-1 infection involves the production of anti-HIV-1 antibodies 4 to 8 weeks post infection, defined as seroconversion, when an individual is acutely infected.

Viral diversification however allows for the selection of viruses that have the capacity to evade and escape responses produced by the infected individual's immune system. During the HIV-1 life cycle viral RNA is reverse transcribed to DNA prior to host genome integration, by an error prone RT that lacks a proof-reading capability and has a mutation rate of ~3 x 10^{-5} per nucleotide per replication cycle (Steinhauer, Domingo, & Holland, 1992). The errors introduced by the RT leads to the vast and rapid diversification of HIV-1, and hence the selection of viruses to which the both humoral and cellular immune responses are ineffective. Over time the selection of viruses which

escape immune responses perpetuate, which leads to viral load increases and $CD4^+$ T cell levels to decrease.

The advanced stage of HIV-1 infection, in which the immune system does not function properly and is unable to contain pathogenic infection, regardless of CD4+ T cell count, is defined as AIDS (Blattner, Gallo, & Temin, 1988). AIDS is characterised by the appearance/emergence of secondary opportunistic infections due to the impaired function of the immune system, which ultimately leads to the death of the infected individual (Blattner et al., 1988).

1.3 HIV-1 Treatment

The primary treatment strategy that is implemented to treat HIV-1 infection is the use of antiretroviral (ARV) drugs. Several different ARV drugs are administered in combination in order to target multiple stages/sites of the HIV-1 life cycle, and are referred to as antiretroviral therapy (ART) or highly active antiretroviral therapy (HAART). This is currently the standard treatment for advanced HIV-1 infection (Hamers et al., 2011). This treatment aims to suppress and contain viral replication in preventing HIV-1 disease progression to AIDS. The advantage of implementing a potent HAART regimen is that the total burden of HIV-1 with regards to viral load is reduced in an individual, as well as disease progression to AIDS is suppressed (Oyugi et al., 2004). An individual's eligibility to receive ART/HAART is dependent on age, viral load and CD4+ T cell count. The World Health Organization's (WHO) 2013 guidelines recommended that individuals were eligible for treatment when having a CD+ T cell count of $\leq$ 500 cells/mm^2. Furthermore, the now outdated guidelines encouraged that individuals with CD4+ T cell counts of less than 350 cells/mm^2 should have been treated as a priority with ART (World Health Organization, 2013) . In July 2014, the guidelines were updated to a strong recommendation that ART should be initiated regardless of CD4+ T cell count (World Health Organization, 2014). The most recent consolidated guidelines recommend that as a priority that ART should be administered to all individuals over 19 years of age that are HIV positive regardless of WHO clinical stage or CD4+ T cell count (World Health Organization, 2016b). It is additionally recommended that infants and children that are HIV-1 positive should initiate ART without delay (World Health Organization, 2016b).

The WHO additionally recommends that ART should be initiated in individuals regardless of CD4 count should the individual be co-infected with active Tuberculosis (TB) disease; and/or co-infected with Hepatitis B virus (HBV) with severe chronic liver disease (World Health

Organization, 2016b). Pregnant and breastfeeding women infected with HIV-1, and HIV serodiscordant couples are immediately eligible for ART by WHO recommendations in order to reduce HIV-1 transmission (World Health Organization, 2016a). Children less than five years old are immediately eligible for ART. Children in the range of 5 to 10 years of age however may initiate treatment when their CD4 ≤ 500 cells/μl, or when they present symptomatic characteristics irrespective of CD4 count (Department of Health, 2015a). ART for adolescents aged 10 to 15 may be initiated when CD4 count ≤500 cells/μl (Department of Health, 2015a). In South Africa, HIV-1 infected individuals with a CD4 count of < 500 cells/mm^2 qualify for ART.

ARV drugs may, in addition, be utilised as a successful prevention strategy to protect individuals with a high risk of infection. Examples of this risk group include men who have sex with men (MSM) and sex workers. Current pre-exposure prophylaxis (PrEP) is a combination of two ARV drugs (tenofovir and emtricitabine) (Centers for Disease Control and Prevention, 2016). PrEP is thus a preventative intervention, significantly reducing infection among high risk individuals within the last decade (S. E. Cohen et al., 2015). In 2014, the WHO recommended that high risk individuals, should consider consuming preventative ARV drugs such PrEP to significantly reduce their risk of contracting HIV-1 (Grant et al., 2010).

1.4 Antiretroviral drug therapy (ART)

The availability of ART has increased significantly in developed countries over the past decade, and is rapidly increasing in developing sub-Saharan African countries (Abdool Karim, 2015). The use of combination HAART has the potential to significantly reduce the morbidity and mortality among patients infected with HIV-1, as well as manage HIV-1 as a chronic condition and attempt to prevent HIV-1 infection progressing and developing to AIDS (Palella Jr et al., 2006). Several ARV drugs are available to treat HIV-1 treatment by effectively supressing viral replication and cause viremia remission. ARVs target different sites and stages of the HIV-1 life cycle. The different classes of ARVs are presented below in **Table 1.1**, and discussed below.

Table 1.1: Five classes of United States Food and Drug Administration approved Antiretroviral drugs

Fusion and Entry inhibitors	NRTIs	NNRTI	Integrase inhibitors	PIs
Enfuvirtide (ENF)	Tenofovir (TDF)	Delavirdine (DLV)	Dolutegravir (DTG)	Amprenavir (APV)
Maraviroc (MVC)	Emtricitabine (FTC)	Efavirenz (EFV)	Elvitegravir (EVG)	Atazanavir (ATV)
	Abacavir (ABC)	Etravirine (ETR)	Raltegravir (RAL)	Darunavir (DRV)
	Didanosine (DDI)	Nevirapine (NVP)		Fosamprenavir (APV)
	Lamivudine (3TC)	Rilpivirine (RVP)		Indinavir (IDV)
	Stavudine (d4T)			Lopinavir (LPV)
	Zidovudine (AZT/ZDV)			Nelfinavir (NFV)
	Zalcitabine (ddC)			Ritonavir (RTV)
				Saquinavir (SQV)
				Tipranavir (TPV)

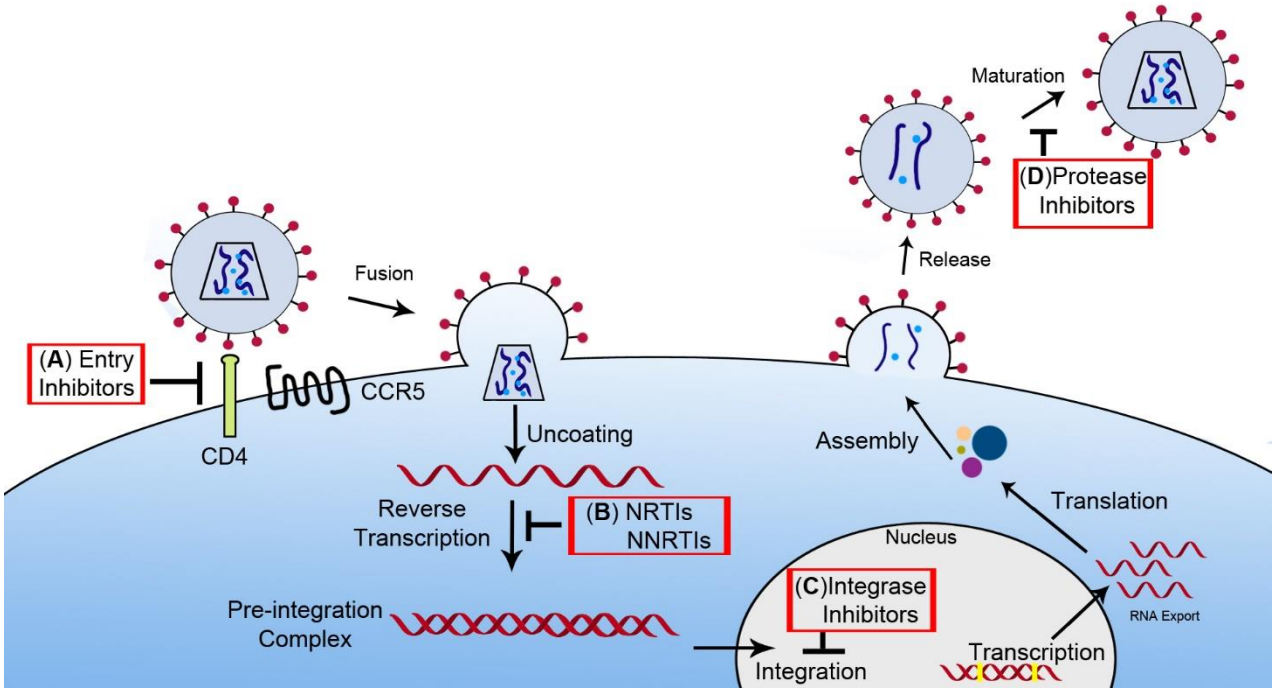


Figure 1.3: A schematic representation of the HIV-1 lifecycle and the target sites of ARVs. Steps A to D correspond to stages of the HIV-1 life cycle that are targeted by ARV drugs, and are described below.

1.4.1 Entry Inhibitors

The first stage of the HIV life cycle is the binding of the viral envelope glycoprotein (gp120) on the surface of HIV with the CD4 glycoprotein expressed on the surface of target cells, followed by subsequent fusion of the virus to the target cell. Entry inhibitors (shown **A** in **Figure 1.3**) are also termed fusion inhibitors and thus impede the virus from acquiring entry to the host cell cytoplasm. Entry inhibitors work mechanistically in one of two possible processes. The drug will interfere with either the gp120 envelope glycoprotein or CD4 complex and/or subsequent downstream conformational changes of the complex, or interfere with the CCR5/CXCR4 co-receptor binding and the gp120 envelope glycoprotein. The two drugs currently approved as ARV's by U.S. Food and Drug administration include Enfuvirtide and Maraviroc

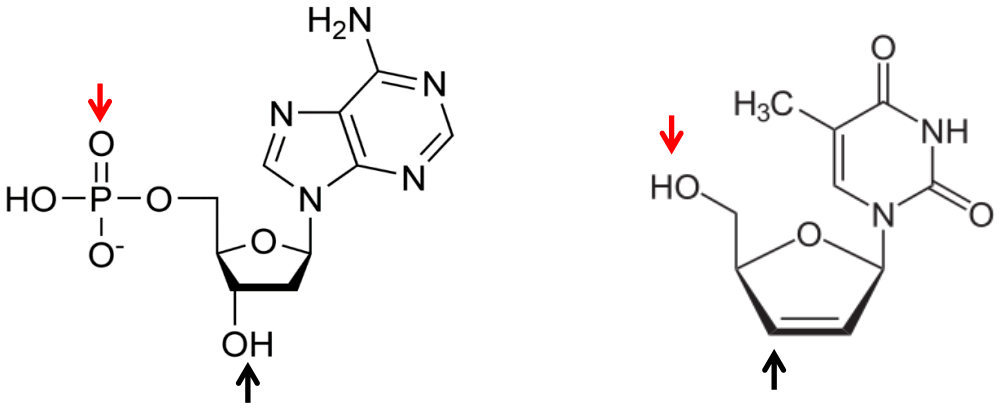
1.4.2 Nucleoside Reverse Transcriptase Inhibitors (NRTIs)/ Nucleotide Reverse Transcriptase Inhibitors (NtRTIs).

As HIV-1 is an RNA virus, possessing a genome comprised entirely of single stranded positive sense RNA, the viral genome must therefore first be converted into double stranded DNA before it can be successfully integrated into the target host cell genome. The subsequent viral gene replication and expression may then occur (Little et al., 2002; S. Miller, 2002). Therefore, a vital step in the HIV-1 lifecycle, post-fusion and cell entry, is the un-coating of the virus and the reverse transcription of the single stranded positive sense viral RNA to cDNA. This process is completed by a virally encoded RT enzyme that is transported along with the virus particle and released upon viral un-coating. It is important to note that this reverse transcription is critically important for the downstream success of viral replication and the production of new virions; to inhibit this imperative step in the HIV-1 life cycle would result in overall stagnation of new virus production (Cann & Karn, 1989). This class of ARV's specifically inhibit the viral reverse transcription stage.

The NRTIs/ NtRTIs (shown **B** in **Figure 1.3**) are nucleotide analogues, which mean that they are similar to nucleosides, but differ structurally in the fact that they lack a 3' hydroxyl (OH) group to which a succeeding nucleotide must bind, as seen in **Figure 1.4.**

The difference between NRTIs and NtRTIs arise in that NRTIs lack a phosphate group on the 5' end of the molecule, which is replaced with an OH group. It is important to note that NRTIs and NtRTIs block replication by the same mechanism – a lack of an OH group on the 3' end of the molecule, which leads to chain termination during DNA replication. NRTIs are first

phosphorylated by host cell kinases; NtRTIs avoid the initial phosphorylation step. NtRTIs are essentially "pre-activated" and subsequently require less metabolic processing to an active state whereby the compounds may be utilized in DNA polymerization (Jochmans, 2008).



Adenine Stavudine

Figure 1.4: The molecule structures of the nucleotide adenine shown on the left, and its nucleotide analogue Stavudine shown on the right. Black arrows indicate where the hydroxyl group is removed for NRTIs, and that result in DNA polymerization termination. Red arrows indicate where the phosphate group is replaced with an OH group for nucleoside reverse transcriptase inhibitors (NRTIs).

"NRTIs are essentially non-extendable nucleoside analogue monophosphate which attach to the 3' end of the growing pro-viral DNA chain" (Clavel & Hance, 2004). This prevents the subsequent nucleotide to be incorporated by RT, as the crucial hydrogen bond between the defective and successive nucleotide cannot be formed. This results in an incomplete and non-functional sequence of polynucleotides that cannot be integrated into the host genome. Thus NRTIs inhibit the reverse transcription process by causing chain termination of the cDNA polymerisation (Gu et al., 1999). Several NRTIs are available and used to suppress the viral replication of HIV-1, and are commonly used in combination as first line drug therapies for HIV-1 infected individuals.

1.4.3 Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

Non-Nucleoside reverse transcriptase inhibitors (NNRTIs) (shown **B** in **Figure 1.3**), likewise also inhibit reverse transcription, but are mechanistically different from NRTIs. NNRTIs are chemical compounds that bind to an allosteric site of the HIV-1 RT enzyme, which results in a non-

functional enzyme and thus inhibiting the reverse transcription process and blocking DNA polymerization (Frentz, Boucher, & Van De Vijver, 2012). The hydrophobic pocket that NNRTIs bind to is not part of the active site of the enzyme, but is located close to it (Clavel & Hance, 2004). The pocket to which NNRTIs bind to is however not present in the absence of NNRTIs. Once NNRTIs have bound, they associate with amino acids and create a pocket, which in turn results in a sterically blocking movement, preventing the appropriate function of the enzyme. HIV-2 replication cannot, however, be suppressed with NNRTI's, as it exhibits natural resistance against NNRTIs (Geretti, 2006).

In South Africa, an NRTI combination of tenofovir (TDF) and emtricitabine (FTC) with the NNRTI of efavirenz (EFV) is used as a fixed dose combination first line therapy (Department of Health, 2015b).

1.4.4 Integrase Inhibitors

The fourth type of ARV drug that is used to suppress HIV-1 viral replication, aims to inhibit the crucial integration step into the target host cell's genome. Integrase inhibitors (INIs) (shown as **C** in **Figure 1.3**), alternatively known as integrase strand transfer inhibitors (INSTIs) targets the HIV-1 integrase enzyme which is responsible for integrating the HIV-1 cDNA (the pre-integration complex) into the host genome. Integration is achieved through a several DNA cutting and joining reactions. The first reaction is termed 3'-end processing. Two nucleotides are removed from each 3'-end of the viral DNA which to be integrated (Temesgen & Siraj, 2008). The second step is termed DNA strand transfer. In this step, the processed viral DNA ends are inserted and joined into the host cell's DNA. The first two steps are catalysed by the viral IN enzyme. In the third and final step, host cellular enzymes repair the single gaps in the host DNA by the removal of the two unpaired nucleotides at the 5' ends of the viral DNA (Temesgen & Siraj, 2008).

INIs contain potent moieties that specifically inhibit the second strand transfer step that subsequently results in an incomplete integration process. The production of a provirus is can thus not occur and HIV-1 viral replication is suppressed as no new HIV-1 viral genomic material may be produced and subsequent products thereof (Pommier, Johnson, & Marchand, 2005; Temesgen & Siraj, 2008).

1.4.5 Protease Inhibitors (PIs)

The final ARV drug class that is utilised in order to contain and suppress HIV-1 replication is protease inhibitors (PIs) (shown as **D)** in **Figure 1.3**. During or directly after the budding of new HIV-1 viral particles, the HIV-1 protease enzyme cleaves the gag and gag-pol polypeptide precursors into functional peptides. HIV-1 protease cleaves viral polypeptides at several different cleavage sites to yield the HIV-1 structural proteins and viral enzymes such as RT and IN (Clavel & Hance, 2004). PIs aim to arrest maturation, and effectively block the infectivity of nascent HIV-1 virions (Clavel & Hance, 2004). PIs bind to the active site of the viral protease enzyme and actively inhibit the function of the enzyme. This is accomplished due to the fact that PIs mimic the natural polypeptides substrates that the enzyme aims to cleave.

1.4.6 HIV-1 Viral Diversity and Quasispecies

HIV-1 is transmitted through viral exposure to mucosal surfaces, or transmitted by means of percutaneous or intravenous inoculations (M. S. Cohen, 1998). In the majority of HIV-1 transmission, a single (founder virus) (represented as a black dot in **Figure 1.5**) or limited number of viral variants establish infection in a new host, resulting in a uniform initial viral population (Keele et al., 2008; Poss et al., 1995).

Shortly after infection, swift diversification of the founder virus occurs, and the initial quasispecies is established. HIV-1 has an exceptionally high evolutionary potential in the human host. The rapid diversification is due to the high replication rate of HIV-1, in conjunction with error prone RT which has a mutation rate of ~3×10^{-5} per nucleotide per replication cycle. This results in a viral turnover of 10^8 to 10^9 virions per day. This diversification occurs via the acquisition of mutations throughout the viral genome (Mansky & Temin, 1995; Menéndez-Arias, 2011; Steinhauer et al., 1992). The high level of recombination of different viral strains and immune pressure additionally drives the virus to diversify (Department of Health, 2015a; Steinhauer et al., 1992). It is important to note that it is most likely that the majority of mutations may have a negative effect on the virus, and would be eliminated by negative selection (Coffin, 1995). The continued diversification of the initial founder virus leads to an expanding range of mutant swarms that are genetically linked to the founder virus. HIV-1 has a remarkably plastic genome, with a very small number of nucleotide positions conserved across the entire genome (Coffin, 1995; Deacon et al., 1995; Maldarelli et al., 2013). The combination of complex HIV-1 mutant swarms may thus exist together as a

quasispecies (Coffin, 1995; Deacon et al., 1995). In the case of more than one founder virus, it is possible for one or more quasispecies to emerge that are each genetically linked to their respective founder virus (Lauring & Andino, 2010).

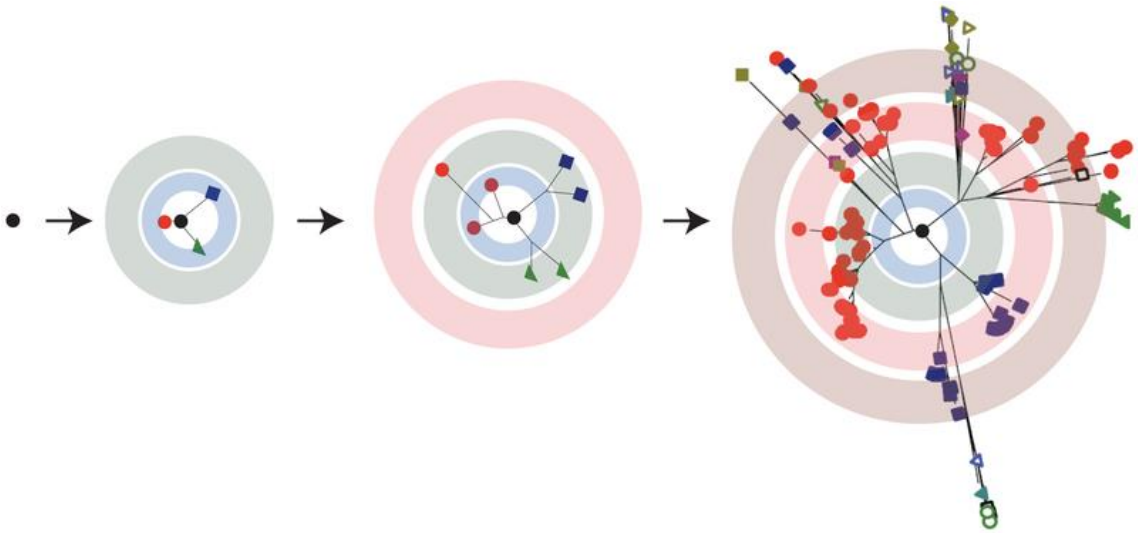


Figure 1.5: A schematic representation of HIV-1 quasispecies that are each genetically linked to the initial founder virus sequence. The black dot represents the founder virus, which soon diversifies to establish an initial quasispecies, followed by further diversification and growth of the quasispecies. Each branch in the figure represents two variants connected by a point mutation, and the concentric circles each represent a replication cycle (copied from Lauring and Andino, 2010).

The highly diverse viral quasispecies that emerges following long term infection is highly susceptible to selective pressures that facilitate resistance to ARV drugs and immune escape (Maldarelli et al., 2013). Certain populations contained within the viral quasispecies may contain sufficient mutations in key drug target sites, which will allow replication despite the drug being present. Susceptible viral populations will be suppressed, and resistant populations will continue to replicate and have the capacity to be transmitted to a new host.

1.5 HIV-1 Drug Resistance

HIV-1 drug resistance means that the virus has the ability to continue replicating itself despite the presence of ARV drugs. Drug resistance is caused by mutations in viral genes that encode for proteins used in viral replication. These proteins are drug targets, and if mutated are not fully affected by the drug.

Mutations in the viral genome are denoted by the original amino acid, the position according to the start of the viral protein, and the resulting amino acid change. An example of the aforementioned is G190A; in which a glycine mutates to an alanine at position 190 in the *RT* gene of the HIV-1 genome. HIV-1 drug susceptibility is detected by observing the viral genome on a nucleotide level by utilising sequencing techniques.

1.5.1 Mechanisms of Drug Resistance

The most common cause for treatment failure and subsequent virological failure in HIV-1 infected individuals is due to ARV drug resistance. There are several types of ARV drug resistance, that differ mechanistically from one another respective to the drugs selecting the resistance.

1.5.1.1 Resistance to Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

There are two types of resistance to NRTIs (Clavel & Hance, 2004). The first type involves mutations that lead to an increased drug discrimination at the drug target site. As mentioned before, NRITs are nucleotide analogues which bind the HIV-1 genome and effectively inhibits the reverse transcription reaction to cDNA by a blocking action. Mutations in the RT gene of HIV-1, results in an increased drug selection capacity. This essentially means that with this type of NRTI mutation/s and resistance, the HIV-1 RT enzyme will preferentially bind analogous natural deoxynucleoside triphosphate (dNTPs) and avoid binding of NRTIs (Clavel & Hance, 2004; Marcelin, 2006).

The second type of NRTI resistance is enhancing the phosphorolytic removal of the chain terminating NRTI from the 3’ end of the primer, post incorporation into the viral DNA. This is often referred to as primer unblocking. After the NTRI has been incorporated, it is effectively removed by a process referred to as pyrophosphorolysis. Key mutations significantly enhance pyrophosphorolysis, such as mutations selected for by Zidovudine and Stavudine. The mutations that lead to the manifestation of this type of resistance are referred to as thymidine analogue mutations (TAMS) (Clavel & Hance, 2004; Marcelin, 2006).

1.5.1.2 Resistance to Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

NNRTIs are designed to bind to amino acids that the HIV-1 RT enzyme is composed of (Clavel & Hance, 2004). Drug binding to a target sequence of amino acids subsequently lead to the formation of a hydrophobic pocket within the RT enzyme (the pocket is not present in the absence of the

drug, and thus induced by the NNRTI drug) (Clavel & Hance, 2004). The formation of the pocket causes conformational change of the RT enzyme, and sterically hinders various areas of the enzyme. Conformational change and steric hindrance thus restricts the movement of the enzyme and thus is unable to complete its intended function of transcribing viral RNA to DNA. It is important to note however that the drug target site and the subsequent hydrophobic pocket do not comprise part of the active site of the enzyme (Clavel & Hance, 2004). Mutations such as L100I, Y181C, G190A and M230L are amino acids that prevent the formation of the pocket, and thus results in a functional RT enzyme (Clavel & Hance, 2004; Jingshan Ren & Stammers, 2008). A common NNRTI mutation, K103N, has a different resistance mechanism. Position 103 of RT is not contained within the hydrophobic binding pocket, but is close to the entrance of the pocket (Clavel & Hance, 2004; Jingshan Ren & Stammers, 2008). The mutation in which the lysine at position 103 is altered to an asparagine creates a hydrogen bond in the un-ligated enzyme near the entrance to the hydrophobic pocket (Clavel & Hance, 2004; Jingshan Ren & Stammers, 2008). The hydrogen bond effectively blocks the NNRTI drugs from accessing the amino acids which create the pocket upon binding, thus resulting in a functional enzyme that can transcribe viral RNA to DNA.

1.5.1.3 Resistance to Integrase Inhibitors

The HIV-1 integrase enzyme is responsible for and mediates the irreversible integration of viral DNA to genomic host DNA (Chiu & Davies, 2004; Craigie, 2001). Efficient expression of viral proteins for generation of new viruses are a crucial step of the HIV-1 life cycle; integration via integrase allows for the maintenance of viral DNA in the host genome and thus expression of viral proteins (Hazuda, 2009).

Integrase inhibitors target strand transfer, the last step of the integration reaction (A. A. Johnson, Marchand, & Pommier, 2004; Pommier et al., 2005; Young, 2001), and thus are referred to as integrase strand transfer inhibitors (INSTIs) (see 1.2.4). Thus, integrase inhibitors bind only to the integrase-viral DNA complex, prior to strand transfer. The INSTI drugs contain a metal binding pharmacophore, and a hydrophobic group. The pharmacophore interacts with two essential magnesium metal ion co-factors that are located in the active site of the integrase enzyme (Espeseth et al., 2000). Upon binding, the drug sequesters the magnesium ions completely inhibiting the function of the enzyme. The hydrophobic group that constitutes part of the drug does not interact

with the magnesium ions, but however significantly increases the affinity for integrase to bind the drug (Hazuda, 2009).

Resistance occurs when amino acids residues that are localised within the active site of integrase, in proximity to residues that are required for coordinating the magnesium ion co-factors, are altered. This essentially results in a reduced susceptibility of INSTI drug, depending on the residue change (Hazuda, 2009).

1.5.1.4 Resistance to Protease Inhibitors (PIs)

The HIV-1 protease enzyme is a homodimer, which means that it is a complex of two identical proteins (Clavel & Hance, 2004). The active site of the enzyme is located within a pocket at the centre of the two parts of the dimer. The main function of HIV-1 protease is to process gag-pol (p160) and gag (p55) poly-protein products into viral enzymes that may be utilised by the virus for subsequent downstream steps. The process of viral maturation may occur instantaneously during or directly after viral budding/viral release (see **Figure 1.1**). In this process, HIV-1 protease cleaves the aforementioned poly-proteins at 9 distinct cleavage sites, resulting in the viral enzymes RT, integrase and protease. Other proteins that are yielded by the cleavage are p17, p24, p7 and p6 (Brik & Wong, 2003; Clavel & Hance, 2004).

The FDA approved PIs that are currently in use as part of ART are peptide sequences that bind to the active site located in a pocket, in the centre of the homodimer (Clavel & Hance, 2004). The PIs contain a synthetic analogue of the phenylalanine-proline amino acids at 167 and 168 of the gag-pol poly-protein, which when bound, inhibits the activity of the enzyme. This results in incomplete cleavage of the poly-protein precursors, and thus arresting maturation and barring the infectivity capacity of emerging virions (Brik & Wong, 2003; Clavel & Hance, 2004).

Many mutations that are selected to allow for the functioning of HIV-1 protease in the presence of PIs, are not located near the active site of the enzyme. These various mutations all mechanistically act in a similar fashion, causing slight conformational changes to the protease that leads to a subsequent widening of the active site pocket in the middle of the homodimer (Appadurai & Senapati, 2016; Clavel & Hance, 2004). The widened active site cavity results in the ineffective binding of PIs, and thus ineffective inhibition of enzymatic activity, which varies by mutation and respective PI. Additionally, the expanded cavity of the active site does not completely inhibit the

viral enzyme from binding its natural poly-protein substrates. The affinity for the natural substrates are altered and lost to some extent, affecting the viral fitness of the mutant strains. However, the loss of affinity is significantly more profound for the protease inhibitors drugs (Clavel & Hance, 2004). This results in a protease that has a slightly less fit, but ultimately may still perform its intended function.

1.5.2 Types of Drug Resistance

The mutations that lead to HIV-1 drug resistance may be acquired before or during treatment. Acquired drug resistance (ADR) occurs by diversification of the virus, and the selection of resistant variants by a selection pressure such as ART. This type of resistance occurs during treatment. Transmitted drug resistance (TDR) differs to ADR in that it occurs before treatment initiation. It occurs when drug resistant variant population of HIV-1 is transmitted to a new host. The two types of drug resistance will be further described below.

1.5.2.1 Acquired Antiretroviral Drug Resistance (ADR)/Secondary Drug Resistance

During the early stages of infection, latent reservoirs of memory CD4+ T cells with integrated provirus are established (Chun et al., 1997). A latent reservoir is defined as "a cell type or anatomical site in association with which a replication-competent form of the virus accumulates and persists with more stable kinetic properties than in the main pool of actively replicating virus" (Arion, Sluis-Cremer, & Parniak, 2000). In these reservoirs, viral strains may be harboured that contain mutations which may allow for a small survival advantage in the presence of an ARV; or several mutations in key sites which permits viral replication regardless of the presence of certain ARV drug classes. Additionally, the wild type (WT) predominantly circulating HIV-1 strains may be harboured in the reservoirs.

At the point where an individual would either be eligible for ART or be diagnosed with HIV infection and be placed on an ART regimen, the virus has diversified a considerable amount. Two major causes for selection of acquired drug resistance is i) incorrect and suboptimal drug treatment, and ii) patient medication non-adherence.

When treatment is initiated, a drug pressure is introduced, which results in the main circulating productively replicating wild-type virus replication to be supressed. Certain viral strains may have diversified to the extent in which the strains may harbour several sufficient key mutations, or

resistance-associated mutations (RAMs), which would allow the stains to replicate in the presence of certain ARV drug classes. If viral genotyping is not performed (to access drug susceptibility and resistance) and thus RAMs are not detected, the said resistant viral strains may emerge as dominant circulating strains. With the continued use of incorrect ART/HAART, susceptible virus replication will be hindered. However, viruses that harbour RAMs in key drug target locations would thus replicate unhindered, and would subsequently result in an increase in viral load.

Alternatively, in the case of drug non-adherence, given sufficient time post the initial date of infection for the establishing wild time virus to diversify, small sub-populations mutated may emerge. The said sub-populations may contain key mutations which would allow a small survival advantage in the presence of an ART/HAART drug pressure. Upon drug non-adherence, in which the drug pressure is reduced, the small population of viral strains with the survival advantage may re-emerge and replicate at an adequate level, which will allow the acquisition of additional key mutations, RAMs, and thus permit a greater survival advantage than before. If this process is permitted to perpetuate, with continued drug on and off ART use, a strain/s will inevitably emerge that may replicate in the presence of the drug pressure. Strains that can replicate, regardless of the specific drug classes used initially to suppress it, are said to be drug resistant to those specific drugs. Viruses that are not suppressed and replicate under drug pressure are at risk of developing additional RAMs that will lead to the development of multi-drug resistance and resistance to several drug classes (Barber et al., 2012). Drug non-adherence leads to virologic failure, and significantly reduces ART's long term clinical success. A significant minority of individuals receiving treatment do not achieve optimal adherence, and the average treatment adherence is estimated to be greater than ~70%, reported in an extensive review de Kok et al. (2018). Adherence is generally measured by adherence self-reporting and pharmacy dispensation records. The general recommendation is that long term viral suppression requires near perfect drug adherence to prevent the emergence of drug resistance and virologic failure (De Kok, Widdicombe, Pilnick, & Laurier, 2018; Pingen, Sarrami-Forooshani, et al., 2014; Rhee, Liu, Ravela, Gonzales, & Shafer, 2004). Literature reports however have shown that exceptionally high levels of adherence will not prevent population levels of drug resistance, and drug resistance will inevitably develop in certain individuals (Bangsberg et al., 2003; Rocheleau, Brumme, Shoveller, Lima, & Harrigan, 2017; Tabb et al., 2018).

Additionally, there is evidence that indicates on-going viral replication in patients receiving HAART/ART, even when viruses are suppressed to levels that are undetectable (Meyer, Matsuura, So, & Scott, 1998; V. Miller et al., 2000). Latent reservoirs of CD4+ T-cells with integrated provirus are believed to perpetuate this on-going viral replication (Chun et al., 1997; Condra et al., 1996). Some of the integrated provirus in these reservoirs may harbour drug resistance mutations, acquired in the early stages of diversification when the viral reservoirs were established; whereas other integrated provirus may be a form of the predominantly circulating "WT". The integrated provirus is competent to replicate, if and when the CD4+ T-cells undergo antigen-driven activation, and thus will drive viral replication in the presence of HAART/ART (Chun et al., 1997; Little et al., 2002). Drug resistant viral strains that are archived will emerge rapidly when activated, as the drug pressure is unable to suppress the viral replication.

1.5.2.2 Transmitted Antiretroviral Drug resistance (TDR)/ Primary Drug Resistance

An HIV-1 infected individual that harbours viral strains containing sufficient mutations in key locations, permitting viral replication in the presence of certain drug types, are considered to be drug resistant to the said drug types. TDR is defined as the transmission of virus that contains one or more RAM (Blower, Aschenbach, Gershengorn, & Kahn, 2001). Drug resistant strains may be transmitted from individuals that are drug naïve and unaware of their infection status, or from individuals that are treatment experienced that have acquired drug resistance and subsequent virologic failure (Baxter et al., 2015). It is generally acknowledged that ~80% of heterosexual transmission events results in the transmission of a single transmitted/founder virus (T/F virus) (Abrahams et al., 2009; Derdeyn et al., 2004; Haaland et al., 2009; Keele et al., 2008; Wolinsky & Wike, 1992). Because of this genetic bottleneck which allows for the transmission of a single or otherwise narrow spectrum of quasispecies that goes on to establish infection, WT virus is unlikely to coexist with drug resistant variants. Thus, due to this genetic bottleneck, it is likely that in a case of transmitted drug resistance, a single drug resistant quasispecies establishes infection.

TDR occurs from infection with an HIV-1 strain that contains one or more RAMs (Perelson et al., 1993). The transmission of HIV-1 containing at least one or more RAM is typically in the range from 5-15% in resource rich geographical areas such as USA and Europe (Steinhauer et al., 1992; Walker et al., 1986). Transmission of drug resistant HIV-1 in Africa is lower at, ranging from 5 - 10% (Hamers et al., 2011; Ssemwanga et al., 2014), however this percentage is expected to rise

significantly in the next decade. Increased availability and rollout of ART/HAART in Sub-Saharan Africa, in conjunction with limited virological monitoring techniques will drive the percentage of TDR (Hamers et al., 2011; Oyugi et al., 2004). With very limited laboratory monitoring available in Africa, there is a severe increase in the potential risk for the rise of drug resistant HIV-1 stains and transmission thereof; given that the prolonged time between virological failure and inevitable clinical consequences (Perelson et al., 1993).

Most common transmitted drug resistant mutations (TDRMs) that have been detected are first generation NRTIs and NNRTIs, with PIs having a lower prevalence (Perelson et al., 1993). NRTI mutations that are most common are M184V/I, K65R, K70E, L74V/I and Y115F (Stanford University, 2017). The M184V/I mutation, a high prevalence mutation, is selected by Lamivudine (3TC) and FTC and reduces the susceptibility to these drugs more than a 100 fold (Grant et al., 2010). The M184V/I mutation additionally reduces the transmission and replication capacity of the viral strain (Abdool Karim, 2015). NRTI mutations additionally include Thymidine Analog Mutations (TAMs) discussed in section 1.1.3.1. TAMs are mutations that are non-polymorphic selected by thymidine analogues Azidothymidine (AZT) and Stavudine (d4T). These TAMs lead to the reduction in NRTI susceptibility by facilitating primer unblocking and thus permitting polymerisation, and subsequent viral genome replication (Centers for Disease Control and Prevention, 2016; S. E. Cohen et al., 2015; World Health Organization, 2016a). The most common TAMs are M41L, D67N, K70R, L210W, T215Y/F and K219Q/E (World Health Organization, 2016b). Mutations M41L and T215Y, frequent TAMs, are commonly found in combination, and confers a high level of drug resistance to AZT and d4T. T215Y is known to have a significant impact on viral fitness, and is replaced with intermediate variants at a rapid rate (Perelson et al., 1993). The K70R mutation is additionally known to have a significantly higher fitness cost when present in combination with other mutations (Perelson et al., 1993). Commonly detected major NNRTIs mutations are L100I, K101E/P, K103N/S, V106A/M, E138A/G/K/Q, Y181C/I/V, Y188L/C/H, G190A/G/E and M230L. K101E/P and K103N/S are frequently detected NNRTIs, with K101E/P occurring in combinations with other NNRTI mutations and selected by Nevirapine (NVP) which reduces susceptibility 3 to 10 fold. K103N/S is similarly selected by NVP and EFV, and reduces the susceptibility of these drugs by 50 and 20 fold, respectively (Palella Jr et al., 2006). The less frequent, but still prominent, major PI mutations are D30N, V32I, L33F, M46I/L, I47V/A, G48V/M, I50L/V, I54V/T/A/L/M, L76V, V82A/F/T/S, I84V, N88D/S and L90M. The most

commonly detected PI mutation is M46I/L, and is selected for by Indinavir (IDV), Nelfinavir (NFV), Fosamprenavir (FPV), Atazanavir (ATV) and Lopinavir (LPV) (Cann & Karn, 1989; Little et al., 2002). M46I/L is commonly detected in combination with several other mutations such as I54V and V82A, and increases the protease catalytic efficiency of the HIV-1 protease (Henderson et al., 2012; Schock, Garsky, & Kuo, 1996).

1.5.3 Persistence and/or reversion of drug resistant mutations

The persistence and/or reversion of drug resistance mutations in both transmitted and acquired drug resistant HIV-1 strains is multifactorial and differ from one another.

Reversion of drug resistance mutations have often been observed in individuals with ADR. ART regimens that span for limited, short periods are often used to reduce the risk of transmission from HIV-1 infected pregnant mothers to their unborn children. Transmission of HIV-1 from an infected mother to an unborn child is referred to as HIV-1 mother-to-child transmission (MTCT). In resource limited settings, the strategy of utilising short-term ART to reduce MTCT is commonly used. The NNRTI Nevirapine (NVP) (see section 1.2.3), is administered in a single dose (also known as single dose nevirapine (SD-NVP)) to pregnant mothers. Infants are often additionally administered SD-NVP shortly after birth with the intention to prevent HIV-1 transmission from an HIV-1 positive mother (Guay et al., 1999). This method of prevention is simple, safe, inexpensive and effective in the context of HIV-1 MTCT prevention (Flys et al., 2005; Guay et al., 1999). The disadvantage of utilising this regimen to prevent transmission is the emergence of NVP drug resistance. In the HIV-1 NET 012 study, NVP drug resistance was detected in 25% of women and in 46% of infants 6 – 8 weeks after the single dose delivery (Eshleman et al., 2004; Eshleman et al., 2001). Another study investigating the prevalence of NVP drug resistance in infants found that in their study cohort of HIV-1 infected infants that were previously exposed to SD-NVP at birth, 62% of infants harboured NVP drug resistance mutations below the age of 6 months (Hunt et al., 2011). Interestingly they also found that the prevalence of NVP drug resistance would decrease as the children got older. NVR drug resistance mutations were detected in 39% of children between the ages of 6 – 12 months, in 22% aged 12 – 18 months, and 16% aged 18 – 24 months (Hunt et al., 2011).

Further investigation into the HIV-1 NET 012 participants detected NVP drug resistance in the form of the K103N mutation in 8 of 9 women and 4 of 5 infants 6 – 8 weeks after SD NVP

exposure. Drug resistance mutations were detected with the more sensitive LigAmp mutation detection method; a method which is more sensitive than population-based sequencing methods as it utilises mutation-specific ligation of 2 oligonucleotides to a DNA template, followed by detection of the ligated product by real-time polymerase chain reaction (PCR). The LigAmp method thus had a K103N mutation detection cut-off of 0.1%. The study detected NVP drug resistance mutation K103N in 3 of 9 women and in 1 of 5 infants 24 months after SD NVP. The less sensitive detection method, such as population-based sequencing did not detect any mutations. The results of the study highlight that the NVP drug resistance mutation K103N reverted to below 0.1% of the viral population within 24 months (Flys et al., 2005).

Overall, in the context of ADR, there is sufficient evidence that suggests that reversion of drug resistance mutations (DRMs) occurs in the absence of a selecting pressure such as ARVs. A broad range of viral strains are present prior to ART in ADR individuals. When ART is initiated, viral replication of drug susceptible virus is supressed, and conversely, strains that harbour mutations that allow for replication in the presence of treatment continues to replicate. Studies have however shown that persistent replication of drug susceptible virus occurs at levels below detectable limits (<50 RNA copies/ml) when an individual that is considered to be on fully suppressive therapy (Dinoso et al., 2009; Fletcher et al., 2014). The same low-level replication of drug susceptible virus also occurs in individual that harbour drug resistant virus. Subsequently, when the selective pressure of ART is stopped, fitter, WT virus replication is no longer suppressed and allowed to compete with drug resistant virus. Inevitably, the fitter WT re-emerges and out-competes the drug resistant virus and leads to the observed reversion of viral strains in ADR individuals.

The reversion of TDRMs however differ from ADR mutations. Due to the narrow genetic bottle neck that ensues upon HIV-1 transmission (Wolfs, Zwart, Bakker, & Goudsmit, 1992; Wolinsky & Wike, 1992), it is highly unlikely that a WT viral strain may be transmitted and co-exist with a drug resistant strain (Castro et al., 2013). Thus, when ART is initiated in an individual that harbours TDRMs, the transmitted strain will continue to replicate. In the cessation of treatment, there is no fitter susceptible virus present at low level in reservoir or in lymphatic tissues and thus reversion is not observed.

Reversion of TDRMs however may occur in the absence of treatment. The reversion of some drug resistant mutations to WT has been observed and reported in studies that have investigated TDR

(Barbour et al., 2004; Brenner et al., 2002; Gandhi et al., 2003; Little et al., 2008; Pao et al., 2004). Viral diversification from the TDR viral strain which established infection occurs due to HIV-1's error prone RT. Viral progeny that emerge as fitter viral strains, with improved viral replicative capacity that have subsequently back mutated to WT sequences or sequences more similar to WT, may continue to diversify and thus outcompete the less fit transmitted mutated virus. The rates at which TDRMs become undetectable are dependent on multiple factors, such as the number of required back mutations, the fitness of the mutated virus relative to the WT, the rate at which the mutated virus replicated and thus viral turnover, and the presence of possible compensatory mutations (Castro et al., 2013). The rate at which TDRMs may be lost varies significantly, and thus there is a paucity of data on viral diversity and evolution of TDR HIV-1.

It has been reported that the most significant factor in the reversion rate of TDRMs is the associated fitness cost; the higher the fitness cost of the mutation, the faster the reversion of the mutation (Perelson et al., 1993). Recent studies have quantified the rate at which TDRMs are lost, and have discovered that NNRTI and PI mutations revert more rapidly, and at a similar rate (Perelson et al., 1993). NRTIs and more specifically TAMs are lost at a slower rate, relative to PIs and NNRTIs (Perelson et al., 1993).

1.6 HIV-1 Viral Genotyping and Drug Resistance Detection

HIV-1 drug susceptibility is determined by observing the viral genome on a nucleotide level by utilising sequencing techniques, and determining whether nucleic acid changes will lead to functional amino acid alterations. Initially a blood sample is collected from an HIV-1 infected individual. Cell free HIV-1 virus is isolated from the plasma layer of blood (Katzenstein et al., 1992). Viral RNA is obtained by initial viral particle lysis and subsequent viral RNA isolation. Viral RNA genes encoding drug target regions are reverse transcribed to cDNA and subsequently amplified. The amplified cDNA amplicons are then sequenced using an appropriate sequencing technology. Conventional Sanger sequencing is the current gold standard for HIV-1 viral genotyping, allowing up to 20% sequenced viral population to be detected (Palmer et al., 2005). Newer technologies such as next generation sequencing (NGS) have the capacity to detect DRMs well below 1% of the sequenced viral population. Identifying the DRMs at high confidence in the analysis steps however presents a challenge. Thus, taking into consideration the error profiles of

NGS platforms, the accepted cut-off for high confidence DRM identification using NGS is set at 1% of the sequenced viral population (Fonager et al., 2015; Zhou, Jones, Mieczkowski, & Swanstrom, 2015). It is important to note however that the prevalence cut-off may differ, depending on the error correction models used in respective downstream bioinformatic tools. The general process of obtaining clinically relevant results from HIV-1 sequencing is illustrated in **Figure 1.6**, and described below.

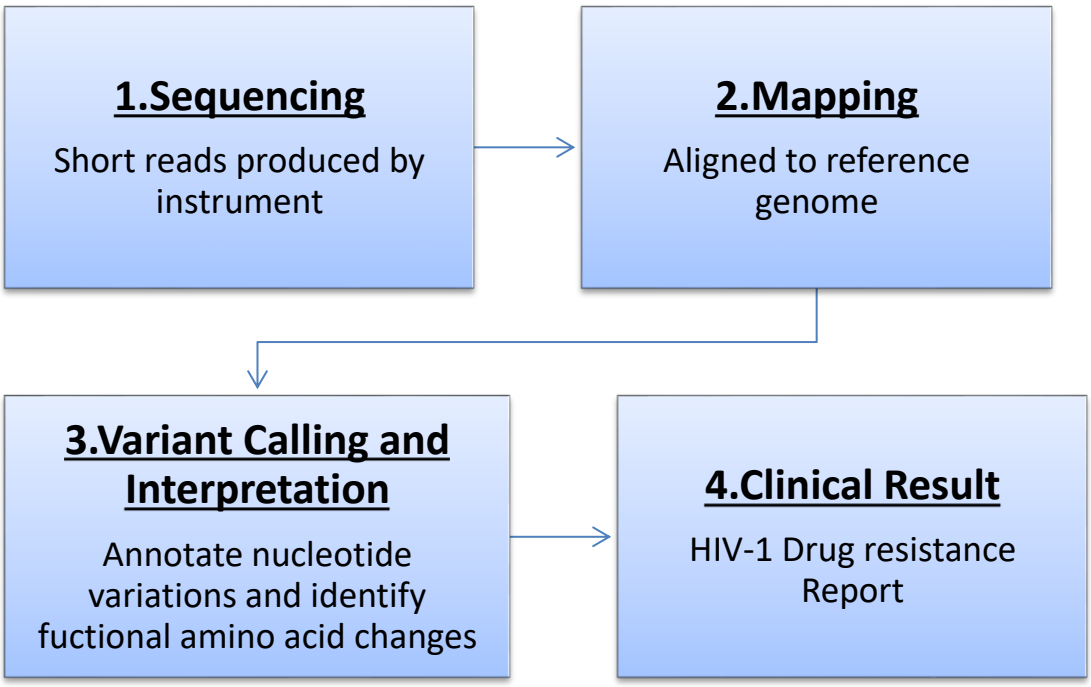


Figure 1.6: An illustration of the process followed from instrument to a clinical result that reveals the drug susceptibility or resistance of virus sequenced from an HIV-1 infected individual.

For conventional Sanger sequencing, the raw sequencing reads are assembled, aligning and a consensus sequence generated. Various software tools are available (RECall (pssm.cfenet.ubc.ca), QualTrace III (www.nucleics.com), ClinQC (www.sourceforge.net/projects/clinqc), and several others) to assemble and align reads, and furthermore may automatically perform analysis. The software will apply quality control measures in which indeterminate basecalls in the chromatogram will be processed in the software's internal algorithms, to assign the most probable correct basecall.

Once a consensus sequence is generated, it is used as an input for analysis by the Stanford University HIV drug resistance database (https://hivdb.stanford.ed). The Stanford algorithm aligns

the consensus sequence to the HIV-1 reference genome and identifies variants relative to the reference. The Stanford algorithm uses the subtype B consensus sequence that is derived from an alignment of subtype B sequences maintained at the Los Alamos HIV Sequence Database (hiv.lanl.gov). Once variants are identified, the algorithm furthermore interprets the significance of the variant in terms of drug resistance. A drug resistant variant is then scored on the level of resistance the mutation confers.

For NGS, reads produced by are short sequences that range in nucleotide length. The short reads are produced without any position information relative to the organism's genome and need to be aligned to a reference genome in a process termed mapping. Once a quality control step has removed any reads that are not of sufficient quality, the reads are aligned and compared to a reference genome (HXB2 in the case of HIV-1). Variants are identified relative to the reference, and based on the number of times the variant is identified in the sequencing reads, an overall prevalence percentage is assigned. Variants that are present at a defined cut-off of the sequenced viral population are identified, analysed and scored by the Stanford University HIV drug resistance database tool to determine the level of drug resistance a variant confers. Alternatively, a consensus sequence of the mapped reads may be generated, and thereafter analysed by the Stanford University HIV Drug Resistance Database tool. Other tools available for HIV drug resistance analysis are Agence Nacional de Recherche sur le SIDA (ANRS) (http://www.hivfrenchresistance.org/) and Rega Institute (Rega) (http://regaweb.med.kuleuven.be/software/rega_algorithm/) algorithms.

1.6.1 Conventional Sanger Sequencing

Fred Sanger and Alan R. Coulson developed and published two methodological papers describing methods to rapidly determine the sequence of DNA (Fred Sanger & Coulson, 1975; Frederick Sanger, Nicklen, & Coulson, 1977). Their chemistry method was based on the incorporation of dideoxynucleosides (ddNTPs) by DNA polymerase. The ddNTPs lack a 3' hydroxyl group that is required for the subsequent formation of a phosphodiester bond to a succeeding nucleotide, and thus the ddNTPs have chain terminating properties (Fred Sanger & Coulson, 1975; Frederick Sanger et al., 1977). The above mentioned ddNTPs may be fluorescently or radioactively labelled with a probe such that it may be distinguished by a particular detection system (Fred Sanger & Coulson, 1975; Smith et al., 1985).

A DNA sample of interest is used in four separate reactions, each using one ddNTP per reaction; the ddNTPs may be ddGTP (dideoxyguanine triphosphate), ddATP (dideoxyadenine triphosphate), ddCTP (dideoxycytosine triphosphate) or ddTTP (dideoxythymine triphosphate). Conventional unmodified deoxynucleosides (dNTPs) (adenine (A), thymine (T), guanine (G) and cytosine (C)), primers sequenced that are complementary to the target DNA, and DNA polymerase are added to each of the four respective reactions that contain one ddNTP. Unmodified dNTPs are added in a 100-fold higher concentration to allow both full sequence transcription and chain terminated fragments to be produced. Primers serve as a starting point for DNA transcription (Frederick Sanger et al., 1977). The resulting DNA fragments from each of the reactions are then heat denatured and size separated via agarose gel electrophoresis. By detecting the signal that is produced by the probe linked to the ddNTPs, and utilising the different fragment sizes, a DNA sequence may be inferred.

The principle of Sanger sequencing was refined and optimised in subsequent years, and became the gold standard of DNA sequencing. Dye-terminator sequencing allowed the utilisation of all four-chain terminating ddNTPs in a single reaction, each emitting their own district wavelengths that can be detected. DNA fragment separation was further developed and improved, separating fragments using capillary electrophoresis. This resulted in highly automated and relatively high throughput instruments that allowed the sequencing of several samples at a time, and outputting results as chromatograms for analysis and interpretation. Sanger sequencing however does have its limitations. Sanger sequencing has the capacity to provide only a consensus sequence of a mixed sample, with a limit of detection for a mutation at ~20%. In the context of HIV-1, the percentage at which a mutation can be identified is limited to ~20% of the overall viral population. Below 20% of the viral population that is sequenced, the mutation would be considered undetectable (Church et al., 2006; B. Larder et al., 1993; Palmer et al., 2005). Drug resistant mutations that are not detected may have significant clinical implications, if appropriate ARV's are not selected for by the attending clinician.

Sanger sequencing however has been the gold standard for HIV-1 ARV drug resistance detection for well over a decade. Despite its limitations, this method has proved extremely valuable to accurately assign a correct ART regimen for an individual that harbours drug resistant HIV-1, and subsequently limit the progression to virological failure.

1.6.2 Next Generation Sequencing (NGS)

High throughput sequencing, also known as NGS, is the umbrella term used to describe modern high capacity sequencing technologies. Driven by the need to reduce cost per sample and increase output, NGS has revolutionised the sequencing field as a whole. Massively parallel sequencing methods utilised by NGS technologies have addressed and overcome scalability of conventional Sanger sequencing by hybridising DNA molecules to beads or solid surfaces such that millions of micro-reactions may occur in parallel (Reis-Filho, 2009). As opposed to utilising longer read lengths in Sanger sequencing, NGS generates millions of shorter read lengths which effectively allows for mutation detection in a small subpopulation that is amongst a large population of WT sequences (Reis-Filho, 2009; Schuster, 2007). This is especially useful in the context of HIV-1 genotyping. Minority variants that harbour mutations that constitute a very small percentage of the total viral population can be accurately detected.

Overall, there are currently four major platforms each utilising unique chemistry, which may be used for the purpose of NGS shown below in **Table 1.2**:

Table 1.2: Summary of commonly used next generation sequencing technologies (adapted from Gibson et al., 2014)

	454 (GS Jr.)	**Illumina (MiSeq)**	**Ion Torrent (S5)**	**PacBio (RS II)**
Amplification method	Emulsion PCR on beads	Bridge PCR *in situ*	Emulsion PCR on beads	Linear
Principle (chemistry)	Synthesis (pyrosequencing)	Sequencing by Synthesis (SBS) (reversible termination)	Synthesis (H+ detection)	Single molecule, real-time synthesis
Average read length (bp)	400–700	150 - 300	200 - 400	4,200–8,500
Primary error (error rate)	Indel ~1 %	Substitution ~0.1 %	Indel ~1 %	Indel ~13 %
Main advantage(s)	Long read length, maturity	Easy work flow, maturity	Low cost, fast run	Longest reads
Main disadvantage(s)	Homopolymer misreads, high cost per Mb	Shortest reads	Homopolymer misreads	High error rate, expensive

1.6.2.1 Roche 454 Pyrosequencing

The Roche 454 was the first NGS sequencing platform made commercially available and utilised the chemistry principle termed pyrosequencing. Chemiluminescent signals are produced in the event of a complementary nucleotide reaction which allows for the determination of a nucleotide sequence (Balzer, Malde, Lanzén, Sharma, & Jonassen, 2010).

Pyrosequencing on the 454 platform is performed by fragmenting DNA of interest by either mechanical or enzymatic methods into fragments of approximately 400 to 600 base pairs (bp) (Rothberg & Leamon, 2008). Adapters are ligated to the ends to the fragments, where they are heat denatured to single stranded DNA fragments. Resin beads are added to the DNA mixture, which are covered by oligonucleotides on the bead surfaces that are complementary to the adapters at the ends of the DNA fragments. The DNA fragments bind to the bead surface, ideally one fragment per bead (Rothberg & Leamon, 2008). An emulsion oil is added to the bead-DNA solution, which forms droplets that separate the beads from one another. An emulsion PCR is then performed, whereby each bead surface is covered by amplification to result in identical copies of the single DNA fragment that hybridized to the bead initially. Beads are added to a pico-titer plate, such that each bead occupies one well of the plate. DNA polymerase and primers are then added to the wells. The primers bind to specific primer binding sites on the adapters, and incorporate nucleotides that are added to the wells in waves. One of the four nucleotides (adenine, thymine, cytosine and guanine) is added per wave. The incorporation of a nucleotides to the template DNA fragment leads to the release of a pyrophosphate, which after a chain of subsequent reactions, causes chemiluminescent reaction that is detected as light emission by a specialised camera (Rothberg & Leamon, 2008). The intensity of the light emission is proportional to the number of nucleotides incorporated into a nucleotide sequence; i.e. a stretch of 3 adenines would have three times the intensity of a single adenine. Pyrosequencing is however limited by the number of homopolymers it can accurately sequence. The light intensities are linked to the nucleotide wave patterns, and the sequence of DNA fragments per well are determined by way of computational processing, and converted to a DNA sequence file.

1.6.2.2 llumina Paired End Sequencing & Sequencing by Synthesis

Paired end sequencing chemistry, or sequencing by synthesis (SBS) describes the sequencing chemistry that is utilised on the Illumina platform, and more specifically the Illumina Miseq

platform that was utilised in this study. Input DNA is fragmented and uniquely identified, by hybridizing a unique combination of adaptors to the 5' and 3' ends of the fragments.

Illumina sequencing by synthesis, consists of three basic steps. The three steps are: (i) sample/library preparation (ii) cluster generation and (iii) sequencing.

Input DNA is randomly fragmented by an enzymatic reaction utilising the Illumina transposome, adapters are thereafter attached to the ends of the DNA fragments by means of reduce cycle amplification. Indexes that uniquely label each sample are ligated to the adapters, such that pooling and multiplexing of samples may occur. A specific size of fragments is selected for, which will be used for cluster generation and sequencing (Illumina, 2016b).

The clustering step is a process in which DNA fragments are passed over the flow cell surface, hybridize, and are isothermally amplified. The flow cell is a glass slide that is comprised of several lanes, each coated with a lawn of two types of oligonucleotides that are complimentary to the oligonucleotide motifs at the end of the generated fragments. A schematic illustration of clustering by bridge amplification is shown below in **Figure 1.7**.

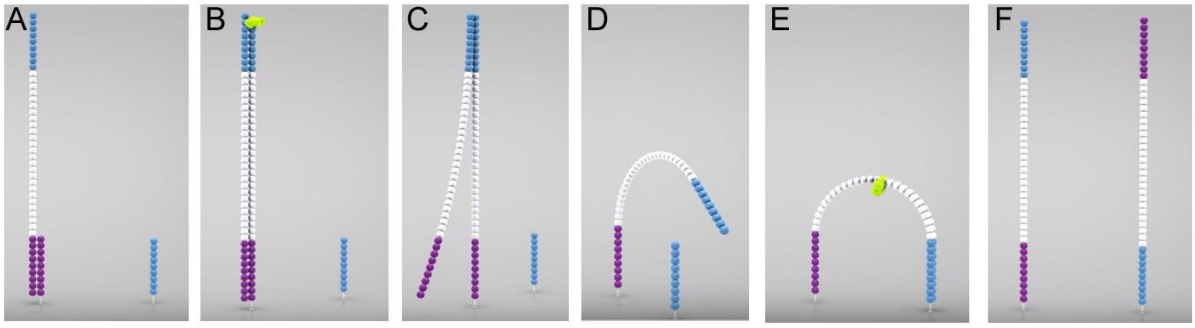


Figure 1.7: A schematic representation of bridge amplification. Bridge amplification is the process in which fragments are isothermally amplified multiple times on the surface of the sequencing flow cell (Adapted from Illumina).

Hybridisation is enabled by the first of the two types of oligonucleotides on the surface of the flow cell. The 3' end of the fragment is blocked, and cannot bind to its complementary oligo on the surface of the flow cell. The 5' end of fragments bind via their complementary oligo motif to the flow cell surface (**A**). Following binding, a polymerase is passed over the surface of the flow cell which creates a complement of the hybridised fragments (**B**). The resulting double stranded

molecule is then denatured, and the original fragment is washed away (**C**). The 3' ends are now unblocked. The remaining fragments on the flow cell surface are then clonally amplified via Illumina's bridge amplification. The stands on the surface of the flow cell to fold over, and the oligo motifs on the opposite end of the strand hybridise to the second type of oligo on the surface of the flow cell (**D**). A polymerase again generates a complementary strand which results in a double stranded bridge (**E**). The bridge is denatured, and thus results in two single copies of the molecule that is tethered to the flow cell (**F**). The process is thereafter repeated until a required cluster density is obtained (Illumina, 2016b).

Once approximately thousand cycles of bridge amplification are completed, the reverse strands are cleaved and washed away. This ensures that reads are all obtained in the same 5' to 3' direction for read one. The sequencing reaction begins by the extension of the first sequencing primer, which binds to the index on the 3' end (as seen in **Figure 1.8**).



Figure 1.8: The 5' to 3'direction of the first sequencing read. The first read sequencing primer binds to the primer binding site, and produces the first read by incorporating nucleotides in the sequencing cycles that follow.

With each cycle, fluorescently tagged nucleotides compete for the addition to the growing chain. A single nucleotide is incorporated based on the sequence of the template strand. Once a single nucleotide is incorporated, the last incorporated nucleotide with fluorochrome attached is excited by a light source, which results in a characteristic florescent single to be emitted and detected,

which thus determines the base call. The number of cycles, in which nucleotides are incorporated, determines the read length of the reaction. Hundreds of millions of clusters are sequenced simultaneously, in a massively parallels process. Upon completion of the first read, the read product is denatured and washed away. An index read then follows once the read product is washed away. A primer binds to a sequence on the 5' end of the DNA fragment attached to the flow cell, which extends to sequence the index on the 5' end (termed p5). As the fragment remains in same location on the flow cell, the first index sequence is assigned to the read that was produced in the previous step, and mapped to the same location on the flow cell. The strands bend over and again by hybridizing to the complementary oligonucleotide sequences on the surface of the flow cell, as shown in **Figure 1.9 (A)**.

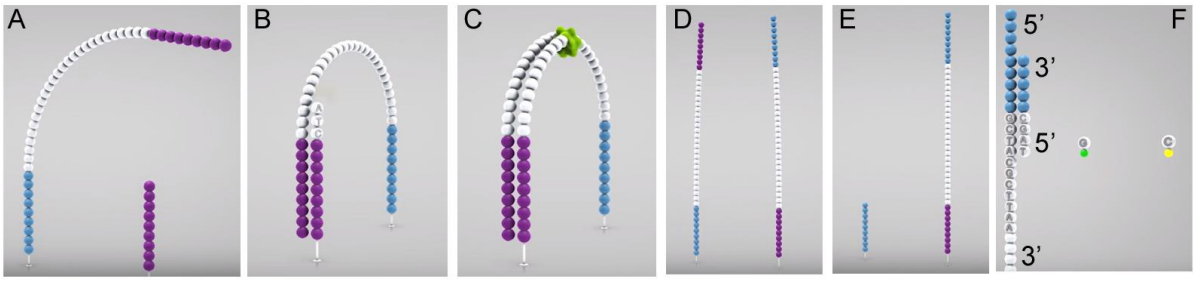


Figure 1.9: The production of the second sequencing read. The second read is produced by reading the second index primer first, and then proceeding with the sequencing reaction.

First the second index read is performed (**B**). A primer attaches to the binding site on the opposite end of the fragment, and extends to sequence the index second index (termed p7) (**B**). As the strand is still bent over, a polymerase is added to the reaction and proceeds to extend the bent over fragment (**C**). Once the original forward strand is copied and the stands are denatured (**D**). The original forward strand is cleaved and washed away (**E**). The second read is now performed. As in the first read, the number of cycles are repeated until the read length is obtained (**F**) (Illumina, 2016b).

1.6.2.3 Thermo Fisher Ion Torrent Platform

The Thermo Fisher Ion torrent systems, and more specifically the most recent instrument termed the S5, utilises semi-conductor technology. A semi-conductor chip containing millions of wells is

used to capture chemical information that is generated during the DNA sequencing process. The release of a hydrogen ion upon successful incorporation of a nucleotide is detected and measured, and converted into digital information that provides data to determine a base call.

The amplification method used by this platform is similar to that of Roche 454 pyrosequencing. Input DNA is fragmented and denatured, followed by the addition of adapter sequences to end of the fragments. The sequence of the adapters allows for the fragments to bind to complementary oligonucleotides on the surface of a resin bead, ideally one fragment at a time (Rothberg et al., 2011). An emulsion oil is added, and results in one oil droplet to contain one bead with one fragment. The oil droplet contains all the PCR components that are required to amplify the single fragment, to cover the entire surface of the bead. Millions of different beads are thus covered with identical copies per bead of millions of different fragments (Rothberg et al., 2011).

The beads are then passed over the semi-conductor chip, which is pitted with millions of wells. One bead is deposited into one well. The chip is thereafter flooded with one of the four different DNA nucleotides every 15 seconds. When a nucleotide is incorporated to a complementary nucleotide on the single stranded DNA fragments on the surface of the bead, a hydrogen ion is released (Rothberg et al., 2011). The release of the hydrogen ion leads to a potential of hydrogen (pH) change in the well, which is detected and read directly on the sequencing chip by an ion sensitive layer at the bottom of each well. The pH change is converted to voltage and recorded, which indicates an incorporation of a nucleotide, and subsequently calling the base (Rothberg et al., 2011). If a nucleotide is not incorporated, no ion is released, and no voltage change is detected and recorded. When two or more nucleotides are incorporated, a voltage change proportional to the number of nucleotides incorporated is recorded (Rothberg et al., 2011). This platform can perform a sequencing reaction faster than other light-based sequencing platforms, and is moreover vastly scalable. The platform however is not as accurate in base calling as other platforms, with increased homopolymer misreads (Rothberg et al., 2011).

1.6.2.4 Pacific Biosciences PacBio

The PacBio platform utilises a DNA sequencing technology termed single molecule real time (SMRT) DNA sequencing. This technology utilises the DNA polymerase to incorporate propriety phospholinked nucleotides. The polymerase activity is visualised, by using nucleotides that are each fluorescently labeled specific to the four different nucleotides. The labeled nucleotides differ

from other light-based sequencing platforms in that the PacBio phospholinked nucleotides carry their fluorescent labels on the terminal phosphate, rather than the nucleotide base (Fabbri, 2013). The fluorescent label is cleaved away once the nucleotide has been incorporated to a complementary nucleotide in the template DNA strand. A completely natural DNA strand results once the labeled nucleotide has been incorporated and cleaved, thus allowing for the exploitation of the inherent properties of DNA polymerase's activity. These properties allow for high speed of sequencing, significantly longer read length and high fidelity (Fabbri, 2013). Although the PacBio platform allows for significantly higher read lengths, the technology is more expensive and has a significantly higher error rate relative to other light-based sequencing platforms.

1.6.3 NGS Error Correction

The process in which genetic material of interest is prepared for sequencing, known as library preparation, introduces artificial errors in downstream amplified and processed genetic material. The introduction of errors may be problematic especially in the context of NGS, as the technology has the capacity to produce deep coverage of target sequences, and thus the detection of low frequency variants. The sequencing instruments may additionally read individual base pairs in genetic sequences incorrectly. The above is considered and addressed bioinformatically, however it is important to note that not all artificially introduced errors, as well as sequencing errors, may be corrected with bioinformatic tools. Thus, a cut off of 1% of the sequenced sample is used, as at this percentage, base calls at a high confidence may be made.

1.6.3.1 PCR Amplification Bias and Sequencing Error

The conversion of viral RNA to cDNA is performed by reverse transcription polymerase chain reaction (RT-PCR). Despite using high fidelity RT-PCR enzymes, errors may be introduced in the reverse transcription reaction. The amplification of cDNA with conventional PCR in the subsequent step that follows RT-PCR, introduces yet another potential step in which artificial errors may be introduced. PCR re-sampling may occur in the amplification steps, and involves the artificial inflation of the prevalence of one variant. An error may be introduced in the first round of PCR, and may be particularly problematic as the erroneous sequence may emerge as the dominant sequence.

The innate properties of the sequencing chemistry may additionally result in nucleotides being read incorrectly by the sequencer, and is termed sequencing error. The innate properties of the SBS chemistry has a defined error rate of 0.1%. A specific sequencing reaction may have its own quality metric, known as the Q value. The Q30 value is the 0.001% probability that that a nucleotide will be called incorrectly. Thus, the higher the percentage of reads at or above Q30, the higher the quality of the library. Reads with less than 100% Q30 are simply discarded in the bioinformatic pipeline. Alternatively, the bioinformatic pipeline may be altered based on the Q30 scores, depending on the sensitivity required.

It is very difficult to quantify the errors introduced in the final sequenced product by PCR error. Sequencing error is easier to identify as the error profiles of instruments are known, and the number of reads required to confidently call a mutation can be based on this error profile. However, due to the error prone nature of the amplification steps that precedes the library preparation, and sequencing error, a mutation cut-off of 1% is used. Mutations frequencies below 1% should not be interpreted with any clinical regard.

1.7 Rationale, Aims and Objectives of the study

The increasing availability of ART in sub-Saharan Africa, in conjunction with WHO recommendations for earlier initiation of treatment for HIV-1infected individuals has led to an increase in HIV-1 drug resistance. A consequence of increased drug resistance is an increasing prevalence of TDR. Individuals identified with baseline drug resistance in the International AIDS Vaccine Initiative (IAVI) Early Infections Cohort (Protocol C) provided a unique research opportunity to study HIV-1 transmitted drug resistance. Of particular interest is the persistence and/or reversion of TDRMs in drug naïve individuals.

Longitudinal samples with varying periods of follow-up were made available for 23 the individuals identified with TDR. Of the 23 individuals, only one initiated ART during the follow up period; all other individuals were ART naïve. Ten of the 23 individuals initiated treatment outside of the follow-up period of this study. Understanding the population dynamics of TDR HIV-1 in the absence of drug pressure is essential for clinical management and public health strategies.

The study aim was to monitor and study the evolution of TDRMs in the individuals identified with baseline resistance.

i. To extract viral RNA from longitudinally collected plasma from the antiretroviral drug-naïve recently infected participants with documented transmitted HIV-1 drug resistance.

ii. Develop an NGS drug resistance assay on the Illumina Miseq platform to detect minority variants that may harbour transmitted drug resistance mutations (TDRMs) with a prevalence of 1% utilising NGS on the Illumina Miseq platform in the longitudinally collected plasma.

iii. To measure, analyse, and compare the decay (reversion) or persistence of TDRMs in the longitudinally collected plasma.

CHAPTER 2: Materials and Methods

2.1 International Aids Vaccine Initiative (IAVI) Protocol C Cohort

Participant samples from the IAVI protocol C cohort were used in this study, and included acutely HIV-1 infected and antiretroviral drug naïve participants from Kenya, Uganda, Rwanda, Zambia and South Africa (Price et al., 2011). The Protocol C project was initially launched in 2006, and to date 613 participants have been voluntarily enrolled (http://www.iavi.org). Population based Sanger sequencing performed in our laboratory identified 29 participants with TDRM. These sequences were available for comparative purposes in this study.

Longitudinal blood plasma samples (n=160) were available from 23 of the 29 individuals in which TDRMs mutations were identified. The number of sample available for each of the 23 participants, in shown in **Table 1** in Appendix D. Samples for 6 of the 29 individuals were not available for use in this study. Plasma samples were shipped to Johannesburg and stored at -80°C until used. Human ethical clearance was obtained from the University of the Witwatersrand for research on human subjects (Clearance number M160691; Appendix A). A 159 of 160 samples were sequenced using NGS on the Illumina Miseq platform.

2.2 Viral RNA Extraction, Reverse Transcription and Nested Polymerase Chain Reaction (RT-PCR) Amplification

2.2.1 Viral RNA Extraction

Viral RNA was extracted from participant blood plasma samples utilising the NucliSENS® easyMAG® total nucleic acid automated extractor (BioMérieux, France). Briefly, 500 µl of plasma was added to an 8 well NucliSENS® easyMAG® disposable cartridge (BioMérieux, France). The viral on-board lysis program was initiated on the easyMAG® instrument. Once lysis was complete, 50 µl of NucliSENS® easyMAG® magnetic silica beads (BioMérieux, France) was added to each well containing lysed sample and re-suspended. Standard protocol procedures were followed on the NucliSENS® easyMAG® instrument and extracted viral RNA was eluted in a final volume of 25 µl. Viral RNA was immediately stored at -80°C to minimize RNA degradation, until required.

2.2.2 Reverse Transcription PCR (RT-PCR)

The HIV-1 *polymerase* (*pol*) gene, encompassing the viral genes *protease* (*pro*) and *reverse transcriptase* (*RT*), was RT-PCR amplified. A one step RT-PCR protocol was followed to reverse transcribe viral RNA to cDNA and PCR amplify an approximately 2.38 kilobase (kb) fragment.

The first-round RT-PCR reaction utilised the SuperScript® III One-Step RT-PCR System with Platinum Taq High Fidelity PCR kit (Thermo Fisher Scientific, Massachusetts, USA). The outer primers used were IN3 (5'- GCA AGA GTT TTG GCT GAA GCA ATG AG -3') (Integrated DNA Technologies (IDT), USA) as the forward primer and G25REV (5'- TCT ATC CCA TCT AAA AAT AGT ACT TTC CT -3') (Integrated DNA Technologies (IDT), Iowa, USA) as the reverse primer. The PCR reaction volumes for a single reaction were 0.5 µl Superscript III RT/Platinum Taq Mix, 12.5 µl 2x Reaction Mix, 0.5 µl each of IN3 and G25REV (10 pM stock concentration), 1.5 µl of $MgSO_4$ (5 mM stock concentration), 1.5 µl of sterile nuclease free water and 8 µl of extracted viral RNA, for a total reaction volume of 25 µl.

The master mix was then placed in a GeneAmp PCR system 2700 (Applied Biosystems, USA) and the following cycling conditions were used: cDNA synthesis for 60 minutes (min) at 50°C, initiation for 2 min at 94°C, followed by 35 cycles of denaturation for 15 seconds (sec) at 94°C, annealing for 30 sec at 54.5°C, and elongation for 2 min 30 sec at 68°C. A second elongation was initiated after the 35 cycles were completed for 10 min at 68°C, followed by a 4°C hold.

2.2.3 Nested PCR of RT-PCR Products

A nested PCR protocol was implemented to improve the amplicon yields. The first round was followed by a nested PCR reaction to amplify an approximately 1.8 kb pol fragment.

The second-round utilised Platinum Taq DNA Polymerase High Fidelity PCR kit (Thermo Fisher Scientific, Massachusetts, USA) with primers PolM4 (5'- CTA TTA GCT GCC CCA TCT ACA TA -3') (Integrated DNA Technologies (IDT), USA) as forward primer and AV150 (5'- GTG GAA AGG AAG GAC ACC AAA TGA AAG -3') (Integrated DNA Technologies (IDT), USA) as reverse primer. The PCR reaction volumes for a single reaction were 0.2 µl Platinum Taq DNA High Fidelity Polymerase, 6 µl 10x high fidelity buffer, 0.5 µl each AV150 and POLM4 (at 20

µM stock concentration), 2 µl MgSO4 (50 mM stock concentration), 37.8 µl sterile nuclease free water and 3 µl cDNA from first round RT-PCR, for a final reaction volume of 50 µl.

The second-round DNA-master mix mixtures were then placed in a GeneAmp PCR system 2700 (Applied Biosystems, USA) and the following cycling conditions were used: initiation for 2 min at 95°C, followed by 35 cycles of denaturation for 30 sec at 95°C, annealing for 30 sec at 51°C, and elongation for 2 min 30 sec at 68°C. A second elongation was initiated after the 35 cycles were completed for 10 min at 68°C, followed by a 4°C hold.

The resulting PCR products were purified utilising the GeneJet Purification Kit (Thermo Scientific Fisher, USA) according to manufacturer's instruction (Appendix B). The purified PCR products were verified by gel electrophoresis on a 1% (w/v) agarose gel (appendix B) and visualised under UV trans-illumination (Biorad, USA). The purified PCR products were stored at -20°C until required in subsequent steps.

2.3 Next Generation Sequencing on the Illumina Miseq Platform

The Nextera XT DNA sample preparation kit, Nextera XT Index Kit (all Illumina, California, USA) were used to generate a sequencing library as per manufacturer's instructions, with several exceptions as outlined below. The MiSeq Reagent Kit V3 (Illumina, California, USA) was used to perform the sequencing reaction. Samples one to 38 were sequenced in run one, 39 to 79 in run two, 80 to 125 in run three and 126 to 159 in run four.

2.3.1 Amplicon Quantification

Purified PCR amplicons were quantified on the Qubit 2.0 fluorometer using the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Massachusetts, USA) (Appendix B). The PCR products were thereafter diluted with nuclease-free water (Thermo Fisher Scientific, Massachusetts, USA) to a final concentration of 0.2 ng/µl. To limit pipetting error, a constant volume of 2 µl per sample was added to an appropriate diluting volume of nuclease free water (Thermo Fisher Scientific, Massachusetts, USA).

2.3.2 Sample/Library preparation

The Nextera XT DNA sample preparation kit and Nextera XT Index kits were used for sample/library preparation following the Nextera XT library preparation guide. The library preparation consisted of four main steps, specifically: (i) input DNA fragmentation (tagmentation), (ii) sample indexing, (iii) fragment size selection and (iv) library normalisation.

2.3.2.1 Input DNA Fragmentation (Tagmentation)

A total of 1 ng (5 µl total volume per sample at 0.2 ng/µl) of each participant sample was added onto a 96 well PCR plate (Biorad, USA), to which 10 µl of tagment DNA buffer (TD) and 5 µl of amplicon tagment mix (ATM) was added to each well. The ATM contains the transposome that enzymatically fragments the input DNA, whilst tagging the fragments with a transposon primer at the ends of the fragments. The plate was then sealed using a microseal B adhesive seal (Biorad, USA), and subjected to centrifugation at 280 x g at 20°C for 1 min, and placed in the C1000 Touch Thermal Cycler (Bio-Rad, USA). The thermocycler was run at 55°C for 5 min, followed by a hold at 10°C. Once 10°C was reached, the plate was removed, and 5 µl of neutralize tagment (NT) buffer was added to each well to neutralise the transposome activity.

.

2.3.2.2 Sample Indexing

Once the input DNA was fragmented (tagmented), the DNA was amplified in a limited cycle PCR that adds indexes and sequences to the end of the fragments that are required for cluster generation. Indexing of samples serves as a method to multiplex and identify each sample after pooling and post sequencing. Unique combinations of indexes (i7 & i5 – 6- 12 base pairs (bp) sequences; provided in the Nextera XT Index kit) were hybridized to the ends of the DNA fragments, to which the transposon attached adapter sequences. The index sequences as a whole consist of 3 motifs: (i) a sequence that is complementary to oligonucleotide sequences on the flow cell surface (ii) i7/i5 indexes and (iii) read primer that is complementary to the transposon primer that was added in the tagmentation step. The motifs are graphically represented in **Figure 2.1**. The combination of index primers was determined utilising the Illumina Experiment Manager (Version 1.9.1). Indexing of samples was performed as per the Nextera XT library preparation guide.

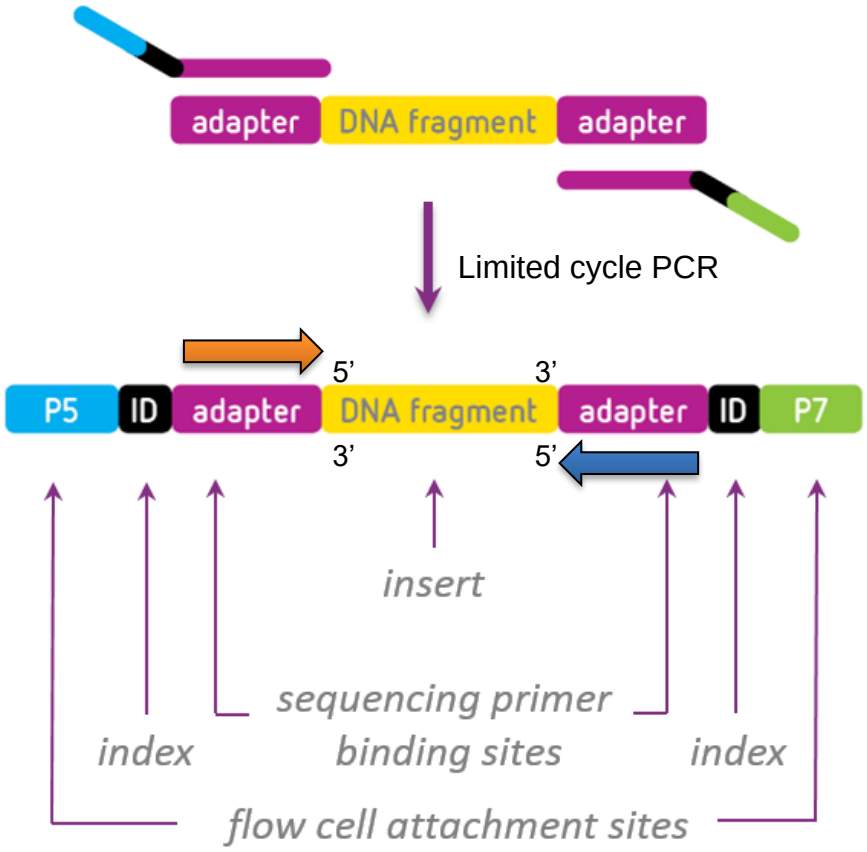


Figure 2.1: Schematic representation of Illumina paired end sequencing adapters added to both ends of the cDNA fragments. Forward reads are indicated by an orange arrow moving in a 5' to 3' direction. Reverse reads are indicated by a blue arrow moving in a 5' to 3'on the complementary strand (Adapted from http://www.gendx.com).

Fifteen microliters of Nextera PCR master mix (NPM) was added to each well of the plate that contained the fragmented DNA. Five microliters of read primer 1 (i7) and 5 µl of read primer 2 (i5) was added to each well, in a unique combination that was determined by means of the Illumina Experiment Manager. The contents of each well were mixed thoroughly by pipetting the solution up and down five to seven times, followed by sealing the plate with a Microseal A film (Bio-Rad, USA). The plate was subjected to centrifugation at 280 x g at 20ºC for 1 minute, and placed in the C1000 Touch Thermal Cycler. The following program was performed on the thermocycler: 72°C for 3 min; 95°C for 30 sec; 12 cycles of: 95°C for 10 sec, 55°C for 30 sec and 72°C for 30 sec; followed by 72°C for 5 min and a final hold at 10°C.

2.3.2.3 Fragment Size Selection

After the samples were fragmented and uniquely indexed, they were subjected to a size selection. Solid phase reverse immobilisation (SPRI) AMPure XP beads (Beckman Coulter, USA), stored at 4ºC, were brought to room temperature. A fresh solution of 80% ethanol was prepared by diluting ethanol absolute (Merck, USA) with nuclease free water. Forty microliters of the indexed PCR products from the previous indexing step was transferred to a new clean 96 well plate. The AMPure XP beads were vigorously vortexed to ensure a homogenous bead solution. Bead volumes relative to DNA volumes (bead ratios) was determined by using work performed by Connolly et al. (2010). The resulting purified fragments obtained by Connolly et al. (2010) were resolved on a 2% (w/v) agarose gel. The gel image obtained by Connolly et al. (2010) is shown in Figure 2.2 below:

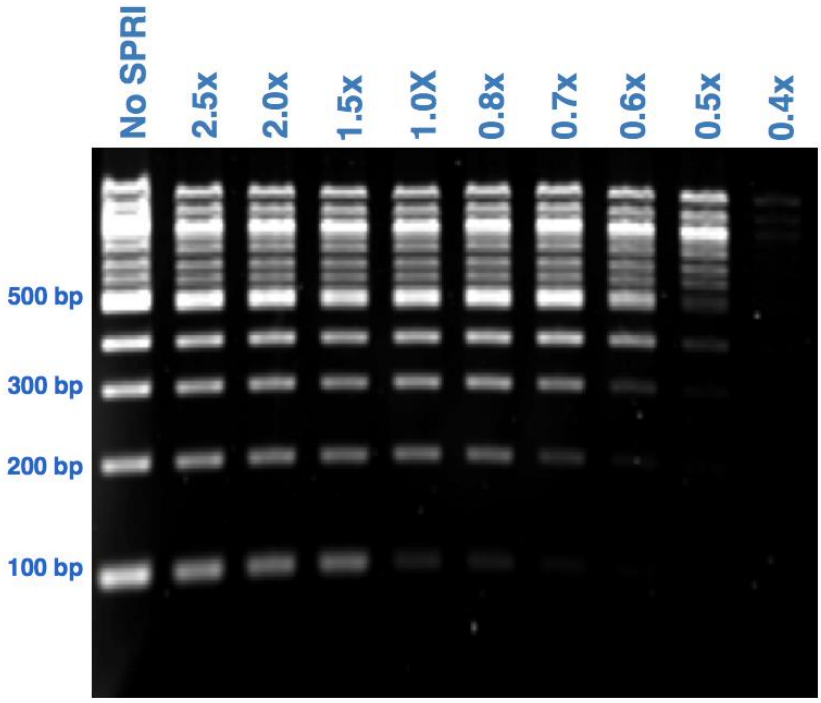


Figure 2.2: A range of SPRI bead ratios (0.4 to 2.5 times SPRI bead volume to DNA sample volume) were used to purify fragmented DNA samples, and the resulting purified fragments were resolved on a 2% agarose gel (Connolly et al., 2010).

SPRI beads were specifically developed by the Whitehead institute (Massachusetts, USA) in 1995 to eliminate smaller DNA fragments, and thus to isolate and purify amplicons as a post-PCR clean up (DeAngelis, Wang, & Hawkins, 1995). Each SPRI bead is composed of a polystyrene core, covered by a layer of magnetite. The magnetite is crucial for the function of SPRI beads, as it

results in the surface of the bead covered in carboxyl groups (DeAngelis et al., 1995). The magnetite additionally gives the beads the property of being paramagnetic, which prevents clumping and falling out of solution in the absence of a magnetic field. The carboxyl groups reversibly bind to negatively charged DNA in the presence of NaCl and polyethylene glycol (PEG), which acts as an aggregation agent. The concentration of PEG and NaCl is proportional to the initial amount of DNA immobilised per bead, and thus the volume of beads to DNA is critical.

DNA size is proportional to the charge of the DNA molecule, i.e. a larger DNA molecule (a larger number of base pairs in sequence) will be more negatively charged relative to a smaller molecule. Due to this property, larger DNA molecules are more likely to bind to the beads due to electrostatic interaction with the beads surface, and thus displace smaller DNA fragments. Thus, the lower the volume of beads relative to DNA, the higher the DNA fragment will bind to the bead surface (DeAngelis et al., 1995). This ratio can subsequently be manipulated to obtain a spectrum of fragment sizes of interest.

SPRI Size Selection Ratios

Four Miseq sequencing runs were performed. For each run, a library preparation was performed, in which a SPRI size selection step was done. Below are the details for each run as to how the SPRI selection was completed and how the selections differed from each run.

SPRI to DNA Ratio for Run One

A double size selection method was utilised in the first run. A double size selection is the use of SPRI beads to eliminate smaller fragments, and thereafter performing an additional size selection to eliminate larger fragments. For the first run an average library fragment size of ~250 basepairs (bp) was desired.

The first SPRI bead ratio to DNA was 1x beads:DNA (50 µl beads to 50 µl DNA). This ratio would eliminate all fragments smaller than 200 bp (fragments approximately $\geq$200 bp would bind to the beads, and fragments approximately $\leq$200 bp would remain in the supernatant). A volume of 50 µl SPRI beads were added to the wells containing sample. The bead-DNA mixture was thoroughly but gently mixed by pipetting the solution up and down ten times, followed by placing the plate on a micro plate shaker set at 1800 revolutions per minute (RPM) for 2 minutes. The plate was

thereafter placed onto a magnetic stand for 2 minutes, and visually inspected to ensure the solution had cleared. The supernatant was carefully removed, so as not to disrupt the beads, and discarded. Fresh 80% ethanol, 200 µl, was added to each well. The ethanol was allowed to incubate in the wells containing the beads for 30 seconds where after it was removed and discarded from each well. This ethanol wash step was repeated. Following the second ethanol wash, the plate was allowed to air dry in a laminar flow cabinet for 15 minutes. The beads were thereafter suspended in 55 µl re-suspension buffer (RBS), and placed back onto the magnetic stand. Once the solution had cleared, 50 µl of the supernatant was collected and transferred to a new 96 well plate.

A second size selection step was performed, in which a 0.5x ratio of AMPure XP beads to DNA (25 µl beads to 50 µl DNA) was added to each well. This step would eliminate all larger fragments (~500 bp and above fragments would bind to the SPRI beads). The bead-DNA mixture was thoroughly but gently mixed by pipetting the solution up and down ten times, followed by placing the plate on a micro plate shaker set at 1800 RPM for 2 minutes. The plate was then placed onto a magnetic stand for 2 minutes, and visually inspected to ensure the solution had cleared. A volume of 50 µl supernatant was carefully collected and transferred to a new 96 well plate.

SPRI to DNA Ratio for Runs Two to Four

Following results obtained in the first run, the size selection ratios were altered. An additional fragment analysis quality control step was introduced.

The desired average fragment size was modified to 750 bp. A review of literature resulted in the modification of the average library fragment size. As per Dudley et al. (2014), a fragment size of 774 bp (range 519-1060 bp) yielded optimal sequencing read lengths.

A single size selection was used in subsequent runs 2 to 4. A ratio of 0.5 x SPRI beads to DNA was used (25 µl SPRI beads to 50 µl of DNA). This would result in binding of fragments ~500 bp and above. The beads were washed twice with 80% (v/v) ethanol and allowed to dry as described in the initial bead size selection above. Once dry, the beads were re-suspended in 15 µl RBS. Ten microliters were transferred to a new clean 96 well plate.

2.3.2.4 Library Fragment Analysis

The size selected, and indexed fragments were quantified using the Qubit dsDNA High Sensitivity Assay Kit as per manufacturer's instruction (Appendix B). Fragment sizes were verified by fragment analysis for runs 2 to 4. Initially the fragment sizes for samples in run one was not determined, as it was believed the inferred sizes would be adequate, however, samples from run one were retrospectively fragment analysed.

The average fragment size of the library was obtained by means of an Agilent 2100 Bioanalyzer utilising the High Sensitivity DNA Analysis Kits (Agilent Technologies, USA) as per the manufacturer's instructions. The Agilent 2100 Bioanalyzer software, 2100 Expert (version B.02.08.SI648 (SR3), Agilent Technologies, USA) was utilised to obtain average fragment size values and visualise graphs. In the software, a sample is selected, which displays an electropherogram illustrating the spread of fragment DNA sizes measured in the respective sample. A region selecting the entire library fragment distribution, from the smallest to the largest DNA fragment was highlighted, and the "region table" tab is selected. Details of the region of interest is displayed, including sizes of the smallest and biggest DNA fragments respectively, concentration of the sample, and the average size of the sample.

2.3.2.5 Library Normalisation

All samples were quantified utilising the Qubit® 2.0 fluorometer with the Qubit® dsDNA High Sensitivity Assay Kit as per manufacturer's instructions (Appendix B). Concentrations were obtained (in ng/µl) and converted to molarity (nmol/l). Molarity was calculated using a calculation provided by Illumina, by using the average fragment sizes of the library that was obtained by means an Agilent 2100 Bioanalyzer (Agilent Technologies, California, United States) and sample concentrations. Samples were diluted to a working concentration of 2 nmol/l (M).

The following calculation was used for the molar conversion:

$$Molarity\ \left(\frac{nmol}{l}\right) = \frac{(concentration\ in\ \frac{ng}{\mu l})\ (10^6)}{(660\frac{g}{mol})\ (average\ library\ fragment\ size\ in\ bp)}$$

2.3.3 Library pooling and denaturing

A series of dilutions were performed in order to obtain the correct final library concentration of 7 pM for optimal cluster density generation. **Figure 2.3** is a schematic representation of the dilution series that was performed:

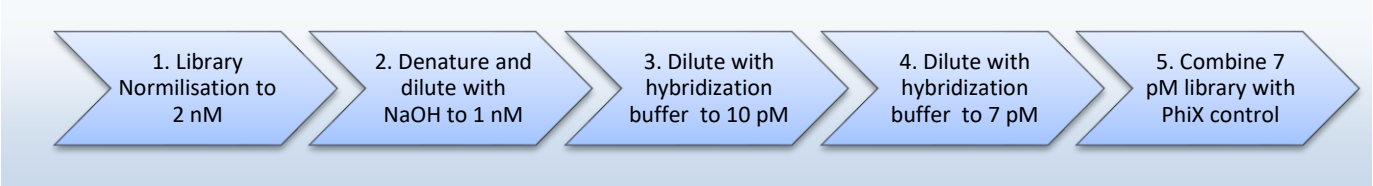


Figure 2.3: The series of library dilutions, starting at 2 nM, performed to obtain the final concentration of 7 pM suitable for Miseq loading and subsequent sequencing.

Five microliters of each sample were added to an Eppendorf microfuge tube. The pooled samples were mixed by pipetting the solution up and down ten times. Five microliters of the pooled sample solution were taken and combined with 5 µl of 0.2 M NaOH (Sigma-Aldrich, USA). The NaOH was prepared by diluting 1 M NaOH with nuclease free water to a final concentration of 0.2 M. The solution was briefly vortexed, and incubated at room temperature for 5 min. This resulted in a 1 nM pooled library solution. A volume of 990 µl of pre-chilled Hybridization Buffer (Nextera XT DNA sample preparation kit) was added to 10 µl of the denatured DNA solution, resulting in a 10 pM solution. The solution was further diluted to 7 pM by mixing 420 µl of the 10 pM DNA solution with 180 µl of Hybridization Buffer.

The PhiX Sequencing Control V3 (Illumina, USA), which is a reliable adapter-ligated library derived from the small, well-characterized PhiX genome, was used as a sequencing control. The PhiX genome was used as quality control for cluster generation, sequencing, and alignment, and a calibration control for cross-talk matrix generation, phasing, and pre-phasing.

Two microliters of the 10 mM PhiX solution provided in the PhiX Sequencing Control V3 kit was diluted with 3 µl nuclease free water. The 5 µl solution was combined with 5 µl of 0.2 M NaOH, vortexed and incubated for 5 minutes. The denatured PhiX solution was diluted to a final concentration of 20 pM by combining 10 µl of PhiX solution with 990 µl of pre-chilled Hybridization Buffer.

The Nextera XT protocol advises that the final volume that is to be loaded into the reagent cartridge for sequencing, is 600 µl. A volume of 480 µl 7 pM library was thus combined with 120 µl 20 pM PhiX (20% of the final loading volume).

2.3.4 Clustering and Sequencing

The Illumina V3 chemistry (MiSeq Reagent Kit V3, Illumina, USA) was utilised for sequencing of all samples on the Illumina Miseq benchtop sequencer. The V3 chemistry was preferable in the context of HIV-1 sequencing, as the V3 chemistry is characterised by longer read lengths (300 bp).

The pooled library (600 µl) was loaded into the designated well on the reagent cartridge provided in the MiSeq Reagent Kit V3. Instrument onscreen instructions were followed to set up the sequencing reaction. When prompted by the Miseq instrument, the cartridge was loaded into the instrument, followed by the incorporation buffer and the flow cell. Special care was taken to rinse the flow cell with nuclease free water followed by visually inspection to ensure no precipitates or residues were present, before loading it into the instrument. The sequencing reaction parameters of 600 cycles (300 forward cycles, and 300 reverse cycles) were specified on the instrument. The 600 cycle parameters were selected as this is the cycle capacity of the Miseq Reagent Kit V3. The sample sheet, previously prepared in the Illumina Experiment Manager (Version 1.9.1), was uploaded into the instrument when prompted in subsequent steps. The sample sheet contains the unique combination of indexes assigned to each sample. A pre-check was initiated on the Miseq instrument, verifying that all microfluidics were operational, after which the sequencing reaction commenced.

Once the sequencing run was initiated, preliminary QC criteria displaying cluster density (k/mm^2), passing filter and Q scores were shown on the instrument's display. Following the completion of a sequencing run, several sequencing reaction metrics were analysed to determine to quality of the run. The metrics are described below:

- Cluster density describes the density of clusters (defined as K), for each tile, on the surface of the flow cell in thousands per mm^2. The Illumina Nextera V3 chemistry flow-cells utilised for all four runs, had 38 tiles per flow cell. The final library preparation was diluted to obtain cluster densities within the desired range of 800 – 1200 K/mm^2. Clusters below 800 K/mm^2 results in a poor under clustered run; cluster densities over 1200 K/mm^2 results

in an over clustered flow cell in which individual base calls cannot be distinguished from one another resulting in poor output.

- Clusters Passing Filter (PF) describes the number of clusters which passed the defined passing filter for each tile in millions. The passing filter is defined during cycles 1 – 25 of the first read. A chastity filter proceeds to remove the least reliable clusters from the image extraction results. A cluster is permitted to pass the filter if no more than one base call has a chastity value below a defined cut-off in the first 25 cycles. The chastity filter is the ratio of the brightest base intensity detected, divided by the sum of the brightest and the second brightest intensities (Illumina, 2015).
- % Phasing is the average rate, in percentage per cycle, at which nucleotides in clusters are not incorporated and thus fall behind, and will continue to lag behind.
- % Pre-phasing is when more than one nucleotide is incorporated in clusters per cycle, resulting in a sequence leading ahead. The phasing/prephasing are reported as a percentage per cycle (Ledergerber & Dessimoz, 2011).
- Reads is simply the number of nucleotide reads detected in the sequencing reaction, and is reported in millions (M).
- A Q score is based in the Phred scale, and indicates the probability that a nucleotide is called wrong by the sequencer. A Q30 score is subsequently a 1 in 1000 (0.1%) probability of a base being called incorrectly. A % >= Q30 value indicates the percentage of bases that with a quality in which the probability of a wrong base being called is less than 0.1%.
- Yield is the amount of data that the sequencing reaction generated in GB.

2.4 Data Analysis

A bioinformatics analysis pipeline was designed to effectively process raw sequence reads generated by the Illumina Miseq instrument for each participant sample, as seen in **Figure 2.4**.

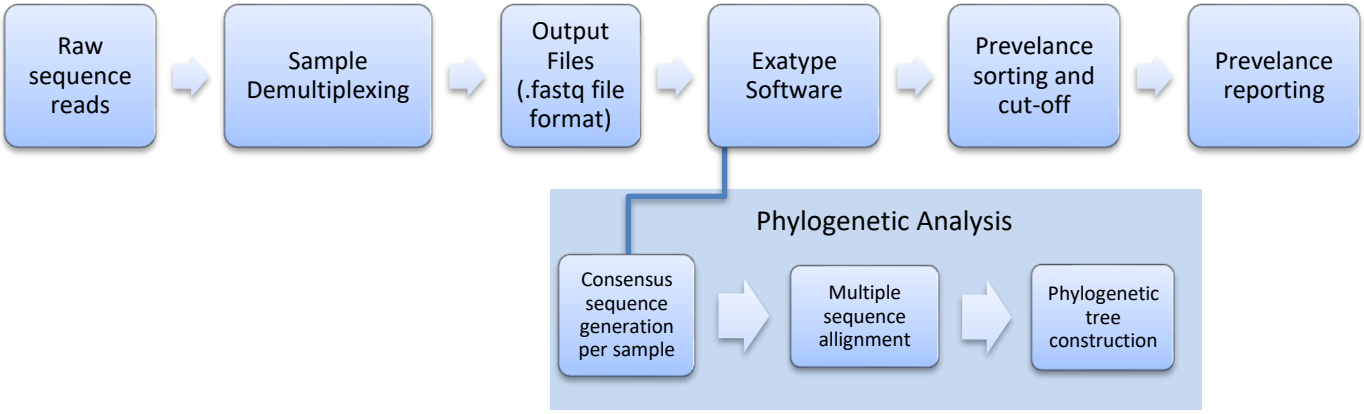


Figure 2.4: Bioinformatics pipeline utilised for processing raw HIV-1 *pol* sequence data to usable antiretroviral drug resistance mutation prevalence and construction of phylogenetic trees for phylogenetic analysis.

Raw sequences containing the unique combinations of indexes for each sample were demultiplexed to respective participant sequences, utilising the Miseq Reporter software (version 2.5.1.2) on the Illumina MiSeq instrument. This resulted in two .fastq files per sample, one for the forward read and the other for the reverse read. The .fastq files were saved and used as input sequences in downstream analyses.

The .fastq files were submitted to Exatype (version 1.11.6, https://www.exatype.com, Hyrax Biosciences, South Africa) to identify DRMs as well as the proportions of each DRM that was present in each participant sample sequenced. The Exatype software aligns reads contained in each .fastq file to a reference HIV-1 genome (HXB2), and identifies differences relative to the reference genome. A coverage requirement is specified in Exatype. A minimum coverage of 330 is required to confidently identify mutations at 1% prevalence. At 330 times coverage a mutation may be identified as a true mutation, and not as a result of sequencing error. The coverage requirement is based on the properties of the sequencing chemistry, and the probability of incorrect base pair incorporation.

Exatype (version 1.11.6, https://www.exatype.com, Hyrax Biosciences, South Africa) was used to identify and annotate nucleotide mutations relative to the reference genome, and additionally scored to determine the functional chances the mutation would induce. Nucleotide changes in codons that would lead to amino acid changes, were reported and recorded. Amino acid changes of the sequenced samples were submitted to the Stanford Calibrated Population Resistance (CPR) tool and Stanford HIV-1 Drug Resistance (Stanford version 7.0) database (https://hivdb.stanford.edu/).

The sequences were analysed for the presence of TDRMs, as well as persistence/reversion of those mutations and/or emergence of new ARV mutations over time, relative to longitudinal samples per participant. The "Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update" paper by Bennett et al. (2009) was used for TDRMs and included the 34 NRTI-resistance mutations at 15 RT positions, 19 NNRTI-resistance mutations at 10 RT positions, and 40 PI-resistance mutations at 18 protease positions.

Mutations were identified at a 1% cut-off for the virus population that was sequenced. Prevalences were sorted by filtering text (.txt) files in which the prevalences were contained. MacOS (version 10.10) terminal scripts were written and used to only keep mutations that were present at prevalences $\geq$1%. The filtered mutations obtained from NGS were thereafter additionally compared to those obtained from Sanger sequencing.

Phylogenetic tree analysis was performed in order to confirm the genetic relatedness of each participant sample and to additionally confirm sample subtypes. A consensus sequence was obtained from each the NGS sequenced sample timepoint, for each participant in Exatype. A text file was created that contained a combination of the consensus sequences, the matched participant baseline Sanger sequences, subtype reference genomes (subtype A to K, 2016, https://www.hiv.lanl.gov) as well as 100 sequences randomly selected for subtypes A, B and C (https://www.hiv.lanl.gov) and was saved as a .fasta file. A multiple sequence alignment was performed using CLUSTAL X (software version 2.0; www.clustal.org), and were thereafter imported and visually inspected and manually edited utilising Aliview (software version 1.18; http://www.ormbunkar.se). The alignment file was imported into CLC Genomics Workbench 8 (software version 8.5.1; QIAGEN, Germany) and a neighbour-joining tree, utilising the Kimura-

80 parameter model with 1000 bootstrap replications, was constructed. The tree was exported as a Nexus file (.nxs) and viewed as a radial tree in FigTree (software version 1.4.2; tree.bio.ed.ac.uk).

CHAPTER 3: Results

3.1 IAVI Protocol C Cohort Demographic and Clinical Data

Demographic and clinical data of the 23 IAVI protocol C participants is shown in **Table 3.1** and **Table 1** in Appendix D. The participants were followed for an average of 26.4 months, ranging from 1.31 up to 65.72 months. The cohort was comprised of 17 males and 6 female participants, with a median age of 33 years. The majority of participants identified with TDRMs were from Uganda.

Table 3.1: Demographic and clinical data of 23 participants from the IAVI Protocol C Cohort identified to harbour HIV-1 with TDRMs.

Variable	
Number of participants	23
Average age at cohort enrolment, (median) (min, max) (IQR) (**years**)	34.78 (33) (20, 53) (18.3)
Average follow up period, (median) (min, max) (IQR) (**months**)	26.4 (22.1) (1.31, 65.72) (19.16)
Male (%)	73.9
Female (%)	26.1
Average Baseline HIV-1 plasma RNA level, (median) (min, max) (IQR) (**RNA copies/ml**)	$1.11x10^6$ ($1.48x10^5$) ($5.39x10^3$, $9.58x10^6$) ($3.49x10^5$)
Average Baseline CD4 cell count, median (median) (min, max) (IQR) (**cells/mm^3**)	522.08 (465) (189, 1079) (196,5)
Site	n = 7 (30.4%) Uganda
	n = 6 (26.1%) Kenya
	n = 5 (21.72%) Zambia
	n = 4 (17.4%) Rwanda
	n = 1 (4.32%) South Africa

3.2 Viral RNA Extraction and RT-PCR Amplification

Viral RNA was extracted, and the approximately 1.8 kb *pol* region of the samples was successfully RT-PCR amplified for 159 of the 160 longitudinal samples (**Figure 3.1**). Samples that did not successfully amplify in an initial extraction and amplification attempt, was re-extracted and RT-PCR amplified. Despite several attempts, sample 57 from participant nine (baseline) failed to amplify.

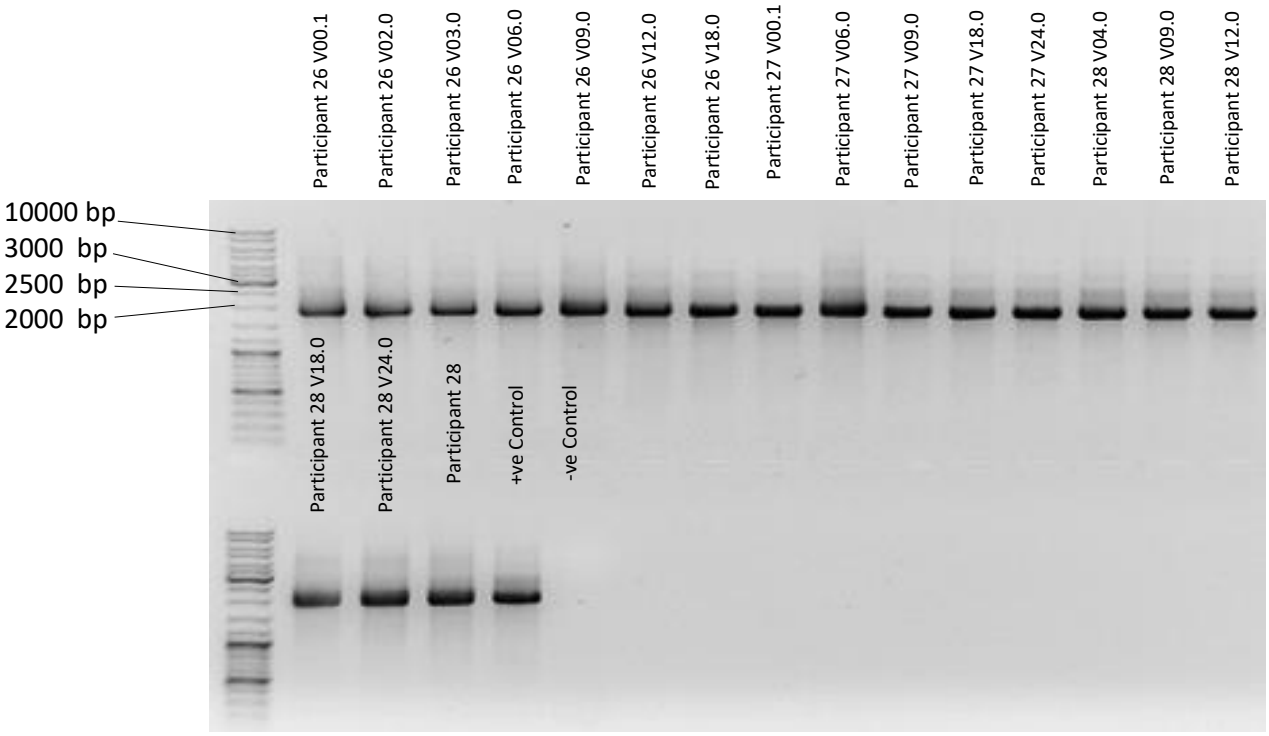


Figure 3.1: RT-PCR amplification of the pol region (~1800 bp) from representative extracted HIV-1 viral RNA from the IAVI Protocol C cohort, electrophoresed on a 1% agarose gel under ultraviolet (UV) trans-illumination. Amplicon sizes were determined utilising the molecular weight marker (MWM, O'GeneRuler DNA Ladder Mix #SM1173, Thermo Fisher Scientific, Massachusetts, USA) basepairs (bp) depicted on the left lane. Participant number and visit dates are assigned to each respective lane.

3.3 Illumina Miseq Nextera XT Library preparation

All 159 samples that were successfully amplified were successfully sequenced on the Illumina Miseq platform.

3.3.1 Amplicon Quantification

Amplicons for all 159 samples were successfully quantified utilising the Qubit 2.0 and diluted to a final concentration of 0.2 ng/µl.

3.3.2 DNA Fragmentation and Size Selection

Amplicons were fragmented, and adaptor sequences were added. The fragments were then subject to size selection. Fragment selection is one of the most crucial steps in library sample preparation for sequencing by synthesis next generation sequencing.

Fragment Size Selection Results for Run One

A double size selection was used to narrow the spectrum of fragments for run one. An average fragment size of ~250 bp was desired for the first run. At the time of the first library preparation, fragment analysis by means of the Agilent Bioanalyzer was not performed. However, the samples were retrospectively analysed. The graph shows the signal intensity in fluorescent units to the corresponding fragment size in base pairs. A representative sample was fragment analysed (sample five from participant one). The analysis of the sample demonstrated that the average library fragment size obtained from the specific double size selection bead ratios was of 280 bp. The graph is shown below in **Figure 3.2**.

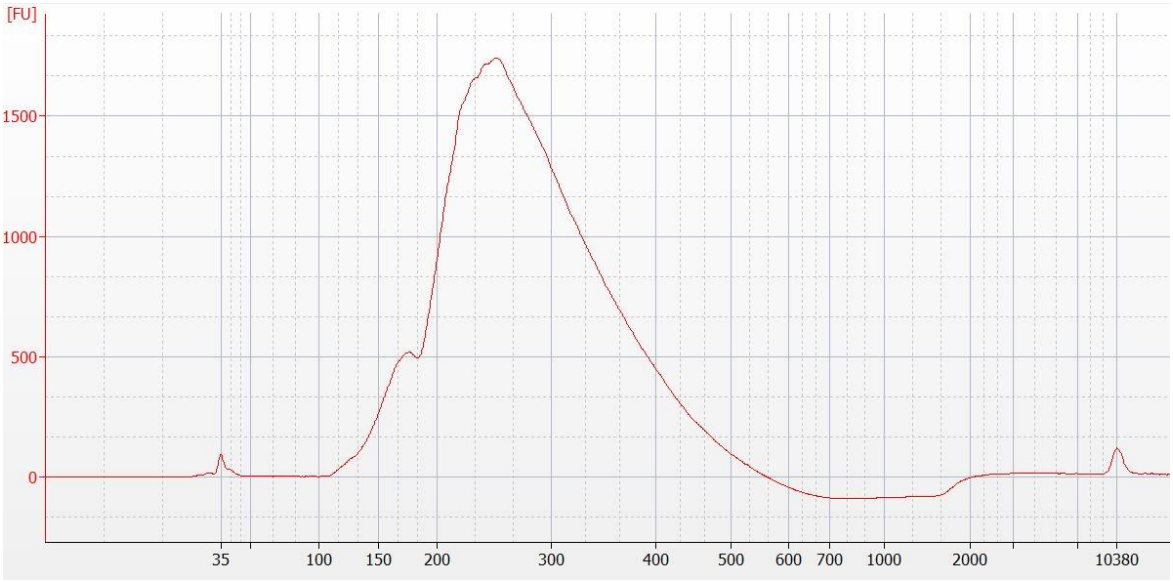


Figure 3.2: Electropherogram showing library fragment sizes for sample five from participant one. The electropherogram graphically illustrates the spread of fragments selected for by a double size selection using a 1x and 0.5x SPRI to DNA bead ratio, respectively. The ratios yielded a peak fragment size of 252 bp and an average library fragment size of 280 bp.

Fragment Size Selection Results for Runs two to four

Low quality metrics were obtained in run one. The fragment size selection was subsequently altered based on literature review. For runs two to four a single size selection was used. A representative sample (sample 47 from participant seven) was analysed. The result, shown below in **Figure 3.3**, demonstrated that the average library fragment size selected for by the single bead ratio selection was ~746 bp.

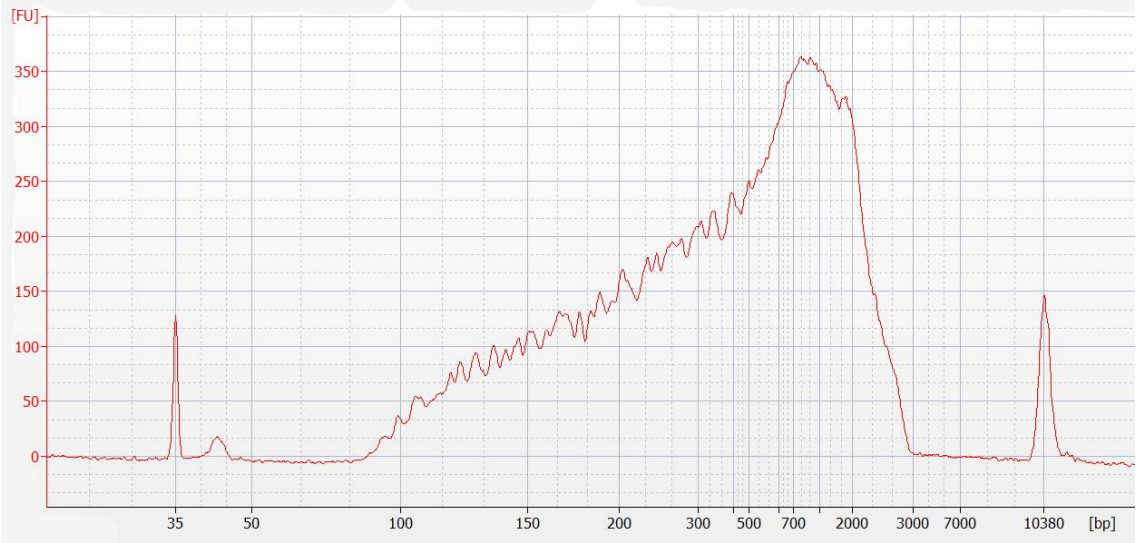


Figure 3.3: Electropherogram showing library fragment sizes for sample 47 from participant seven. The electropherogram graphically illustrates the spread of fragments selected for by a single size selection using a 0.5x SPRI to DNA ratio. The single selection yielded a peak fragment size of ~724 bp and an average library fragment size of 712 bp.

Fragment sizes may be manipulated by using different SPRI bead rations. In run one a double size selection was used. The first selection step used a 1x bead to DNA ratio. This specific bead ratio would bind fragments ~200 bp and above. Unbound smaller fragments were eliminated by discarding the supernatant. A second size selection was then performed, to narrow the spectrum of fragments. A ratio of 0.5x SPRI beads to DNA was used. This ratio would bind fragments ~500 bp and above. Smaller fragments less than ~500 bp would be left in the supernatant, and larger fragments would be bound to the beads. The larger fragments were eliminated, by keeping the supernatant and discarding the beads. The resulting spectrum of fragments had an average size distribution of 280 bp.

In runs 2 to 4, a single size selection was used. A single SPRI ratio of 0.5x beads to DNA was used to bind fragments ~500 bp and above to the beads. Smaller fragments, less than ~500 bp were discarded by eliminating the supernatant. A single size selection resulted in an average size distribution for the three library preps (runs 2 to 4) to be ~746 bp.

3.4 Illumina Miseq Next Generation Sequencing

A total of four Illumina Miseq sequencing reaction runs were performed utilising the Illumina Nextera XT V3 chemistry. With each run, library preparation was successively optimised to obtain ideal cluster densities. Miseq sequencing metrics for the four runs results are shown below in **Table 3.2**.

Table 3.2: Miseq sequencing metrics recorded for the sequencing runs one to four

Run	Read	Cluster Density (K/mm^2)	Clusters PF (%)	Phas/ Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Yield (GB)	Error Rate (%)
1	1	248 +/- 36	98.01 +/- 0.53	0.171 / 0.054	6.22	6.09	64.57	1.83	4.73 +/- 0.32
	2	248 +/- 36	98.01 +/- 0.53	0.193 / 0.026	6.22	6.09	55.88		7.29 +/- 0.73
2	1	1408 +/- 37	77.52 +/- 5.20	0.142 / 0.042	32.48	25.21	70.6	7.56	3.50 +/- 0.42
	2	1408 +/- 37	77.52 +/- 5.20	0.099 / 0.005	32.48	25.21	55.95		6.10 +/- 0.69
3	1	775 +/- 5	90.69 +/- 2.12	0.147 / 0.050	19.19	17.4	91.45	3.48	1.16 +/- 0.07
	2	775 +/- 5	90.69 +/- 2.12	0.113 / 0.030	19.19	17.4	85.58		1.30 +/- 0.11
4	1	1107 +/- 25	85.34 +/- 2.29	0.150 / 0.062	26.13	22.32	88.5	4.46	1.29 +/- 0.10
	2	1107 +/- 25	85.34 +/- 2.29	0.130 / 0.060	26.13	22.32	80.73		1.81 +/- 0.18

- **K/mm^2** = number of clusters per square millimetre (in thousands).
- **PF**= The number of clusters passing filter for each tile (in millions).
- **Phas/Prephas percentage** = The estimated percentage of molecules in a cluster for which sequencing falls behind (phasing) or jumps ahead (prephasing) the current cycle within a read.
- **Reads** = Number of reads produced in the sequencing reaction (in millions).
- **Reads PF** = Number of reads that passing filter (in millions).
- **% >= Q30** = The percentage of bases with a quality score of > 30.
- **Yield** = Amount of data produced in sequencing reaction (in gigabytes).
- **Error Rate** = The calculated error rate, as determined by the spiked in PhiX control sample.

Overall, each run successively resulted in improved metrics each time, as library preparation was optimised and refined. Additionally, final dilutions were adjusted to yield ideal cluster densities.

3.5 Exatype Results

The Exatype software produced two diagrams per run, which served as a metric to illustrate the overall quality of the run. **Figure 3.4** shows sequencing reads obtained per sequencing run per sample. **Figure 3.5** shows the average coverage of the *pol* gene obtained from all the samples sequenced per run.

Figure 3.4 illustrates the number of sequencing reads that were generated for each sample, for runs 1 to 4.

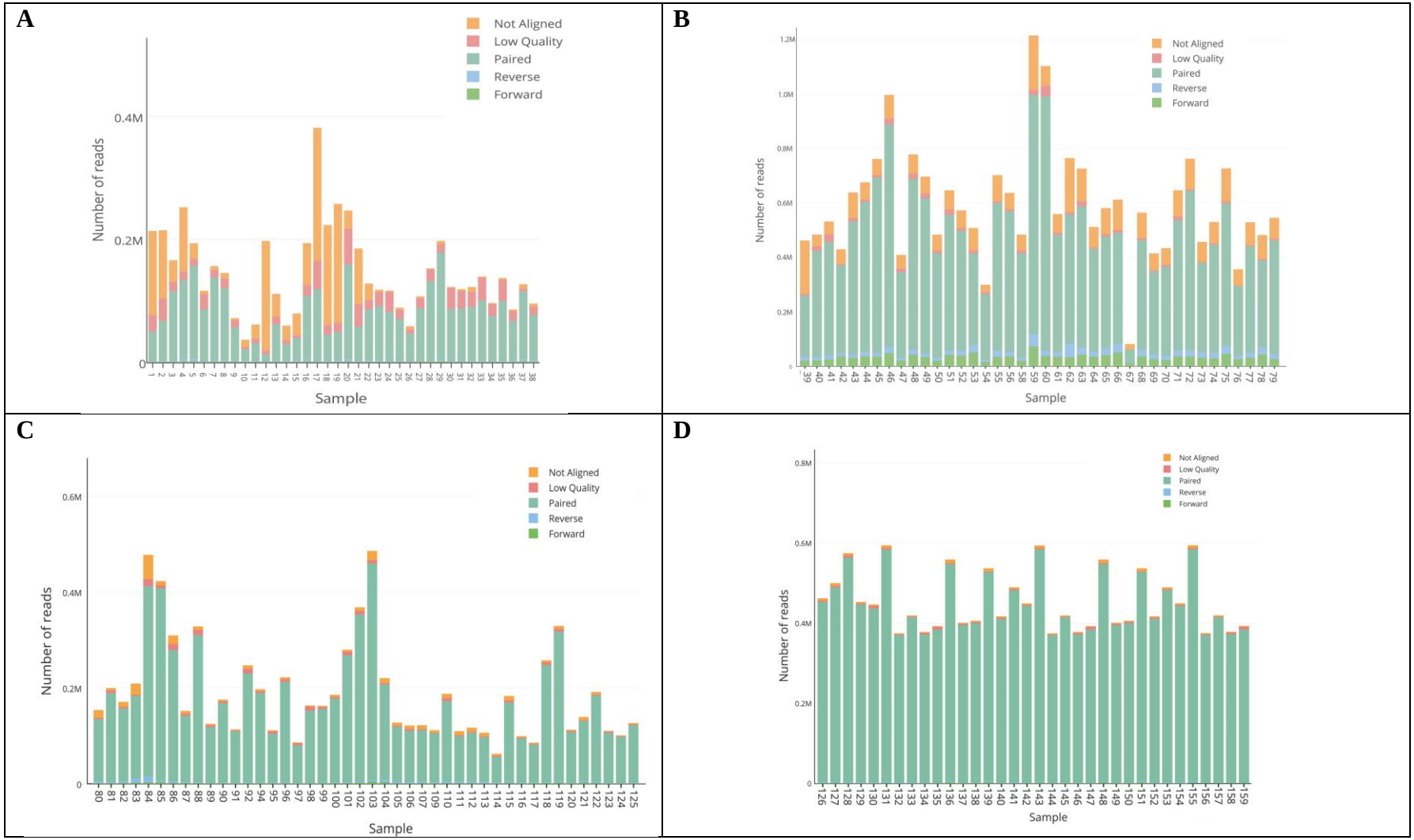


Figure 3.4: Miseq sequencing reads generated in runs 1 to 4, for samples 1 to 159. The four graphs illustrate the number of forward and reverse sequencing reads, as well as unaligned and low-quality sequencing reads generated for each sample, per run. Run one to four are shown in graphs **A** to **D,** respectively.

The variance observed in the peaks, and thus number of reads per sample can be attributed to unequal amounts of sample that are added to each sequencing reaction. The library normalisation step, in which all samples are diluted to 2 nM and pooled, aims to normalise the amount of sample added per reaction such that each sample is sequenced equally; thus, theoretically that each sample produces the same amount of reads per reaction. The library normalisation step was optimised and subsequently the peak variance range decreased with each consecutive run. **Figure 3.4 D** illustrates an acceptable read variance range; however, optimisation may yet be required in the library normalisation step to further decrease the range.

Overall, the quality of each successive run increased as the protocol was optimised, and the technique better understood. **Figure 3.4** A illustrates a large number of non-aligned and poor-quality reads per sample, indicating that the quality of the library sequenced was poor. The average number of paired forward and reverse reads for run one was approximately 100 000 reads per sample. This value increased to an average of approximately 500 000 reads per sample for run two, however a very large number of unpaired and poor reads were obtained in the second sequencing run. In run three a decrease in unaligned and poor-quality reads was observed, yet the average reads per sample was approximately 200 000 reads. In run four, poor and unaligned reads were near to the number of paired read obtained for each sample, with an approximate average of 400 000 reads per sample.

Figure 3.5. was produced by the Exatype software (Hyrax Biosciences, South Africa), and is a graphic representation of the read coverage of the *pol* gene (or vertical depth; i.e. how many times a section of *pol* was sequenced) for all samples sequenced per sequencing. A minimum coverage of 330 (shown as a purple line in **Figure 3.5**) was required to identify DRMs at 1 % of the sequenced viral population.

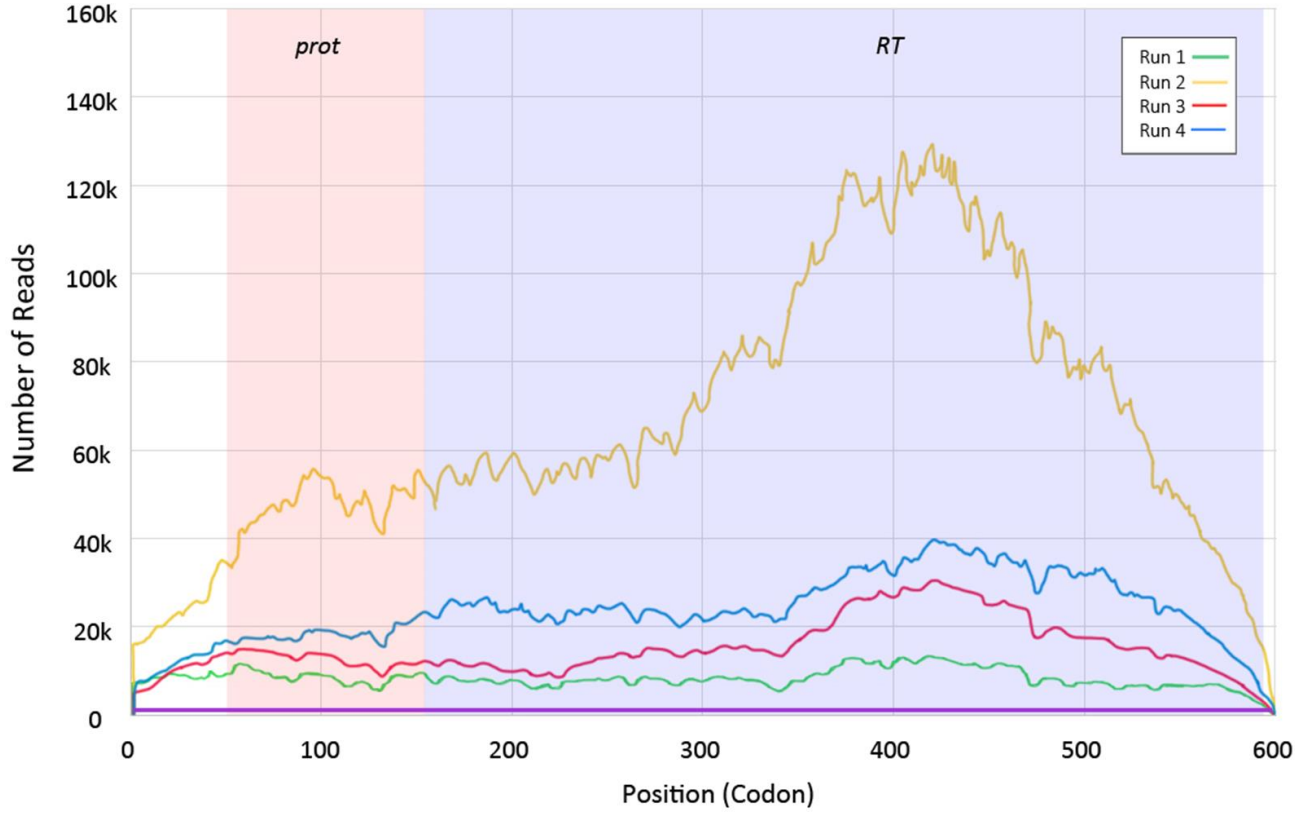


Figure 3.5: The average number of read coverage for the region of *pol* that contains *prot* and *RT*, for all samples (1 to 159) sequenced in the 4 sequencing runs. Coverage in number of reads, per run, relative to the codon position of *pol* obtained for all samples is shown. The reads that fall within the *prot* and *RT* codons are coloured in red and blue, respectively. A purple line indicates the minimum number of reads (330) required to identify mutations at 1% of the sequenced viral population.

Overall, the read coverage obtained for all runs were more than sufficient to identify TDRM mutations present at equal to or greater than 1% prevalence of the sequenced viral quasispecies. The Exatype software had a coverage requirement of 330 reads to identify mutations $\leq$ 1%.

3.6 Phylogenetic Analysis

Phylogenetic analysis was performed in order to ensure genetic relatedness of each sample, and additionally to confirm the subtype of the samples. Baseline Sanger sequences and all IAVI baseline and follow-up consensus sequences for all 23 participants were used to draw a phylogenetic tree. Each consensus sequence is generated by selecting the most common nucleotide observed in mapped reads at each position in the reference, including insertions occurring in >50% of reads. Additional subtype reference sequences and randomly selected HIV-1 sequences from the Los Alamos HIV sequences database were used to draw the tree shown in **Figure 3.6.**

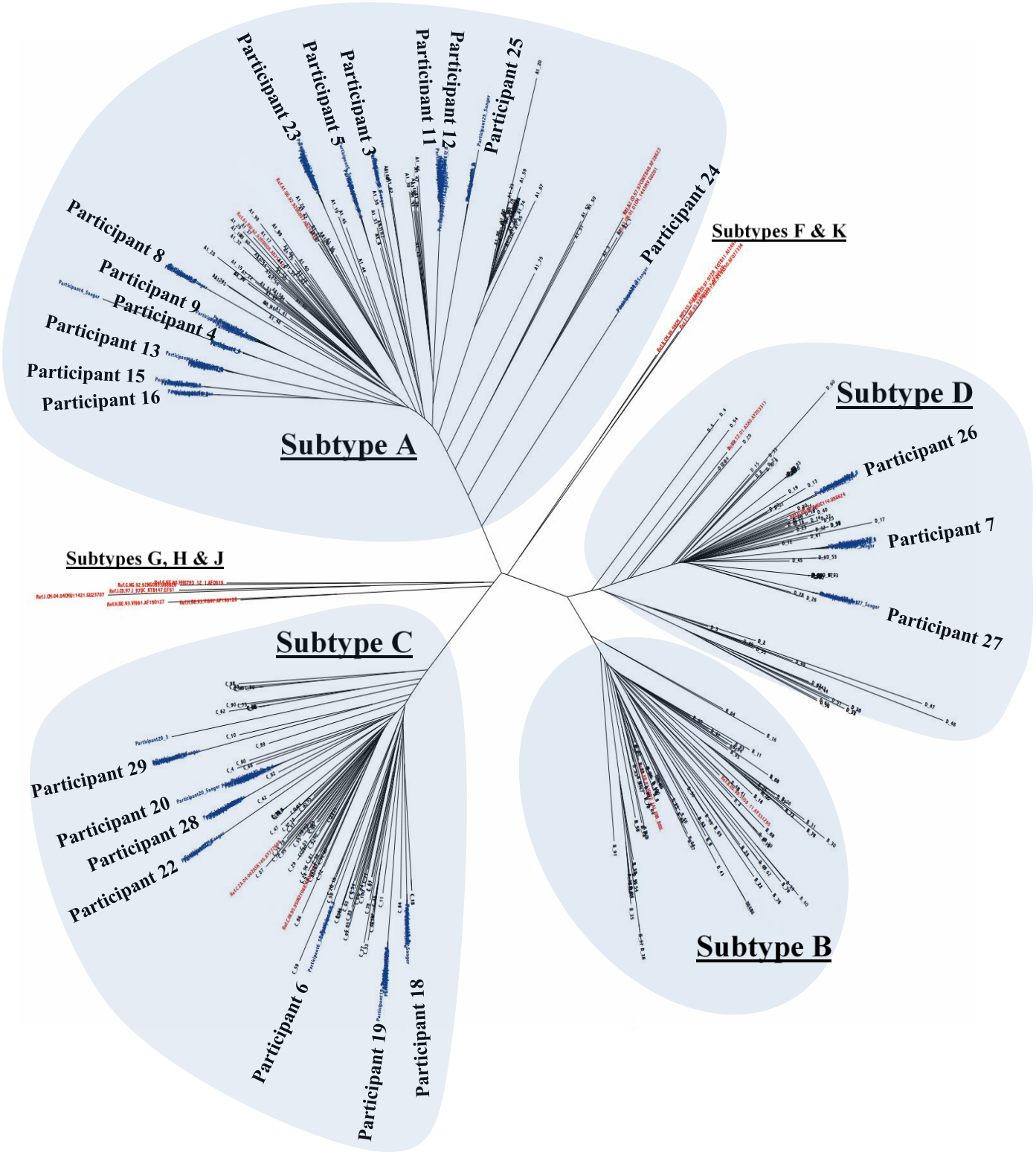


Figure 3.6: Neighbour-joining tree (Kimura-80 model with 1000 bootstrap) of all 23 participants. Participant samples are shown in blue, Subtype A-K reference genomes are shown in red, and background noise HIV-1 samples are shown in black. Sub-type branches, with each participant's cluster are annotated on the tree.

3.7 Antiretroviral Drug Resistance Mutation Results

All 159 successfully amplified samples were sequenced on the Illumina Miseq platform. The Exatype software produced resistance calls of mutations as per the Stanford University HIV-1 drug resistance database that were present at ≥1% of the sequenced viral quasispecies per sample. All mutations detected in all 23 participant's longitudinal samples are shown in **Table 2** in **Appendix D**.

3.7.1 Conformation of Sanger Data by NGS, and Persistence and Reversion of TDRMs identified at baseline

Table 3.3 below contains mutations that were detected at baseline using conventional population-based Sanger sequencing by Price et al. 2011, and mutations identified using NGS in this study. Samples in this study verified the mutations that were detected by Sanger and detected additional minority variants that were missed by Sanger sequencing. The end point sample for each participant's longitudinal sample set is shown below the baseline samples for NGS, indicating which mutations persisted, and which additional mutations were present at end point of the follow up period. As Sanger was only performed on the initial baseline sample in which TDRMs were identified, there is no Sanger data for the end-point samples.

Table 3.3: Persistence or reversion of HIV-1 transmitted drug resistance mutations in 23 participants from the IAVI Protocol C Cohort, showing clinical parameters, subtype, ARV drug resistance mutations (% of amplified sequenced virus) detected by NGS, and mutations detected by Sanger Sequencing**.**

Pct No	Visit Days	Viral Load (RNA copies/ml)	CD4 Count (cells/mm³)	Sub-Type	Sanger Baseline Mutations		NGS	
					RT DRMs	PR DRMs	RT DRMs (% Prevalence)	PR DRMs (% Prevalence)
1	B: 0	83400	456	A1	K103N	L100I	K103N (99.83)	
	L: 152	195000	320				K103N (99.01)	
2	B: 0	148000	496	A1	-	M46L	-	M46L (99.89)
	L: 313	26500	254				-	M46L (99.95)
3	B: 0	169000	589	A1	K103N, Y181CY	L10I	K103N (10.84), Y181C (71.88), Y188C (1.45), G190A (4.47)	L10I (99.75)
	L: 152	350000	250				K103N (4.46), T69N (7.11), Y181C (60.43), G190A (4.74), H221Y (5.33)	L10I (99.89)
4	B: 0	66600	899	C	K103N	-	K103N (40.90), K103T (1.30), V106A (1.82), V106M (1.19), V108I (2.96), Y181C (5.80), Y188C (9.38), N348I (1.74)	-
	L: 39	61100	698				K103N (95.91), Y188C (1.58)	-

Pct = Participant; B = Baseline sample; L = Last follow up sample; RT = Reverse transcriptase; PR = Protease; DRM = Drug resistance mutation; NO AMP = No amplification; - = No DRMs observed; Greyed out block = Sample was not sequenced by Sanger sequencing

5	B: 0	525000	459	D	-	I85V	-	I85V (98.87)
	L: 1673	90377	178					I85V (61.55)
6	B: 0	148069	490	A1	K103N	L10V	K103N (99.15)	L10V (99.89)
	L: 313	25665	365				K103N (73.59), G190A (28.85)	L10V (83.35)
7	B: 0	16000	975	A1	D67N	-	**NO AMP**	**NO AMP**
	L: 1470	6100	653				-	-
8	B: 0	165000	620	A1	K103N	L10V	K103N (98.96)	L10V (99.67)
	L: 665	28373	465				K103N (99.27), Y188H (1.92)	L10V (99.66)
9	B: 0	3E+06	386	A1	K103N	L10V	K103N (98.92)	L10V (99.64)
	L: 671	3945	400				K103N (99.54), M184I (2.31)	L10V (99.85)
10	B: 0	9E+06	327	A1	Y188C	-	Y188C (99.68)	-
	L: 586	8445	602				Y188C (39.42)	V82A (6.75)
11	B: 0	146000	913	A1	K103N, E138A	-	K103N (98.79), E138A (99.54)	-
	L: 792	52904	758				K103N (97.99), E138A (99.75)	-
12	B: 0	141000	303	A1	K103N	-	K103N (98.17)	-
	L: 749	227879	307				K103N (97.56)	-
13	B: 0	5390	464	C	-	T74S	-	T74S (99.7)
	L: 1255	38300	204				K65R (1.35)	T74S (99.61)
14	B: 0	392000	521	C	K103N	T74S	K103N (99.03)	T74S (99.81)
	L: 1140	104773	353				K103N (98.18)	T74S (99.69)

Pct = Participant; B = Baseline sample; L = Last follow up sample; RT = Reverse transcriptase; PR = Protease; DRM = Drug resistance mutation; NO AMP = No amplification; - = No DRMs observed; Greyed out block = Sample was not sequenced by Sanger sequencing

15	B: 0	452000	465	C	K103N, V108IV, M184V	-	K65R (1.03), K103N (98.55), V108I (65.64), M184V (99.17)	-
	L: 509	30799	499				K103N (73.81), K103S (23.84), G190A (1.60), K219R (1.43)	-
16	B: 0	166374	189	C	K103N	T74S	K65R (1.18), K103N (98.71)	T74S (99.67)
	L: 480	56569	199				K65R (1.38), K103N (98.92)	T74S (99.75)
17	B: 0	73200	311	A1	K103N	-	K103N (99.24)	-
	L: 1998	11107	308				K103N (96.52), K103S (2.78)	-
18	B: 0	690000	1079	A1D	-	I85V	V118I (99.50)	I85V (99.14)
	L: 106	134000	736				V118I (99.78)	I85V (96.57)
19	B: 0	61000	366	A1	L210W	-	L210W (13.84)	V82A (1.25)
	L: 295	12600	235				L210W (92.62)	-
20	B: 0	333000	362	D	K103N	-	K103N (99.29)	-
	L: 454	26300	384				K70T (37.09), K103N (99.38), F227L (1.89)	-
21	B: 0	32500	532	D	-	M46L	K219R (4.60)	M46L (87.94), I47V (4.95)
	L: 512	152000	236					
22	B: 0	121000	495	C	K103N	-	K103N (93.74), V106M (4.42)	
	L: 748	40300	665				K103N (34.36), K103S (19.55), V106M (33.59)	-
23	B: 0	73800	621	C	M41L	-	M41L (99.67), K103R (99.46)	-
	L: 828	1017	447				M41L (99.49), K103R (99.44)	-

Pct = Participant; B = Baseline sample; L = Last follow up sample; RT = Reverse transcriptase; PR = Protease; DRM = Drug resistance mutation; NO AMP = No amplification; - = No DRMs observed; Greyed out block = Sample was not sequenced by Sanger sequencing

The NNRTI K103N was the most prevalent mutation overall, with 60.9% (14 of 23) of participants presenting the mutation. Other NNRTI mutations detected were Y181C (n=1), Y188C (n=1), E138A (n=1), V108I (n=2). NRTIs that were detected are D67N (n=1), M184V (n=1) and L210W (n=1). The most prevalent PIs were T74S (n=3) and L10V (n=3); other PIs detected were M46L (n=2), L10I (n=1), L10V (n=3) and I85V (n=2). Additional minority quasispecies/variants were detected by NGS that were not detected by Sanger population sequencing. The mutations Y188C (n=2), K65R (n=2), G190A (n=1) and V118I (n=1) were detected at prevalences below 10% in baseline samples.

3.7.2 Longitudinal Persistence and Reversion Analysis of TDRMs

Figures 3.7.1, 3.7.2 and **3.8** below shows TDRM outcomes throughout the longitudinal follow-up period for the 23 participants. The figures are separated into participants in which full reversion was observed, in which a combination of persistence and reversion of TDRMs were observed, and in which full persistence of TDRMs identified at baseline are observed.

Reversion of TDRMs was observed in 17% of participants (4 out of 23) (participant three, seven, 15 and 21). A mutation is said to be reverted when the mutation was below the study cut-off of 1% of the sequenced viral population. Viral quasispecies from participants 7 and 21 (**Figure 3.7.1**) showed complete reversion of D67N and M46L over time, respectively. A combination TDRM persistence and reversion was observed in participant three and 15 (**Figure 3.7.2**). Participant three harboured HIV-1 quasispecies which showed the persistence of K103N, Y181C, G190A and L10I, and the reversion of Y188C. Participant 15 similarly harboured quasispecies which showed persistence of K103N, and reversion of V108I and M184V. **Figures 3.7.1** and **3.7.2** below graphically illustrate the viral load of participants relative to time in which TDRM reversion, and reversion and persistence respectively, was observed.

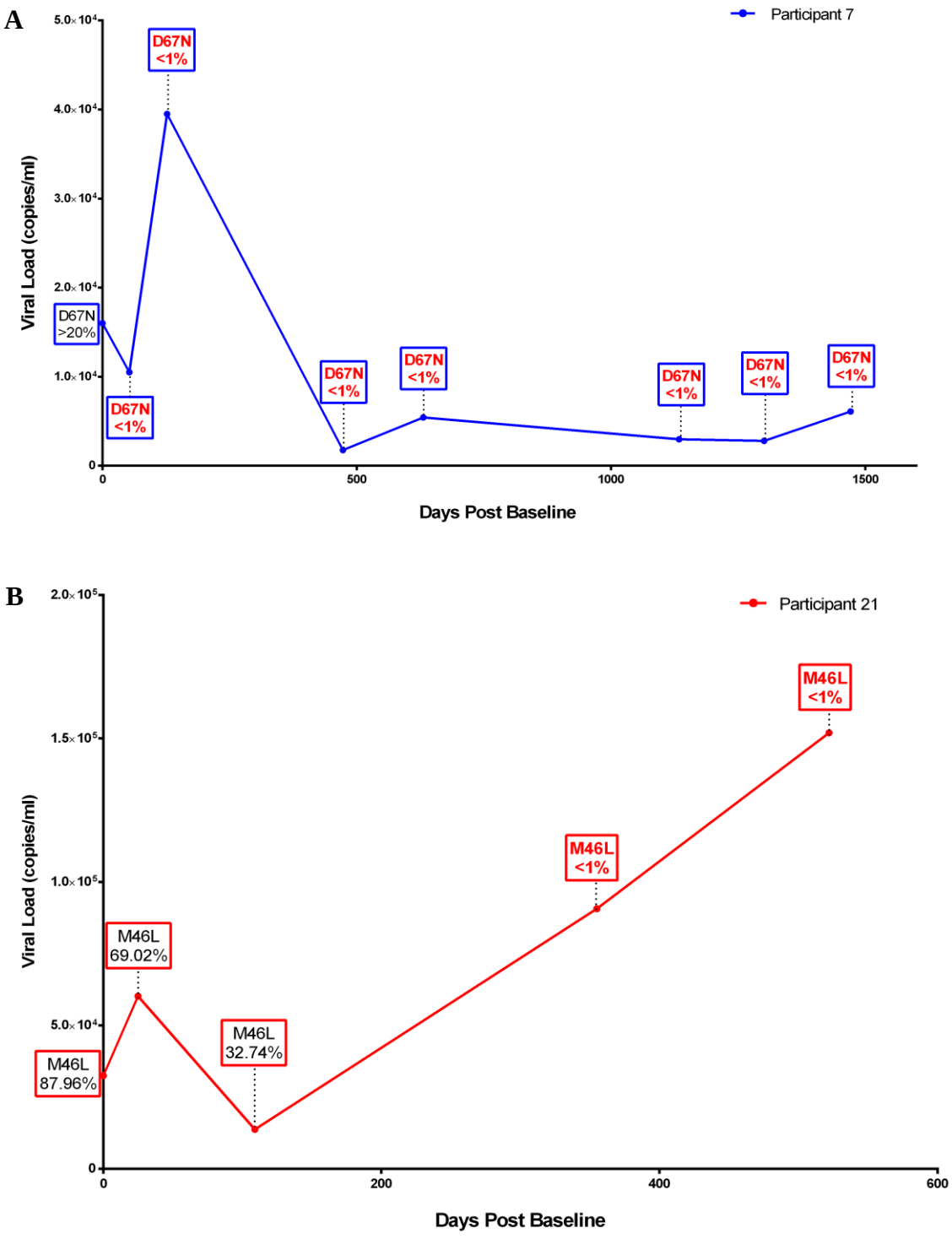


Figure 3.7.1: Viral load against follow up period in days for participants in which complete reversion was observed. Participant seven and 21 are shown in blue (**A**) and red (**B**) respectively. A TDRM is written in red where it was detected below 1%. Reversion of TDRMs for both participant seven and 21 occurred within 300 days after the baseline sample.

Figure 3.7.1 A, illiterates the longitudinal follow-up of the single TDRM D67N identified in participant seven, and shows the rapid reversion of the mutation. The baseline sample in the longitudinal set of samples failed to amplify, and thus as the mutation was detected by means of population-based Sanger sequencing by Price et al. 2011, the prevalence was assumed to be >20% of the sequenced viral population. In the remaining seven samples for participant seven, no DRMs were identified at a prevalence greater than 1% of the sequenced viral population. It was thus concluded that the D67N mutation reverted in participant 7.

Figure 3.7.1 B, illustrates the persistence/reversion outcome of the TDRM M46L across the five samples that make up the longitudinal sample set for participant 21. A gradual decline in prevalence of M46L is observed in the first 3 longitudinal samples, after which the M46L mutation was no longer detectable above 1% of the sequenced viral population for the remaining samples. It was concluded that M46L reverted in participant 21.

.

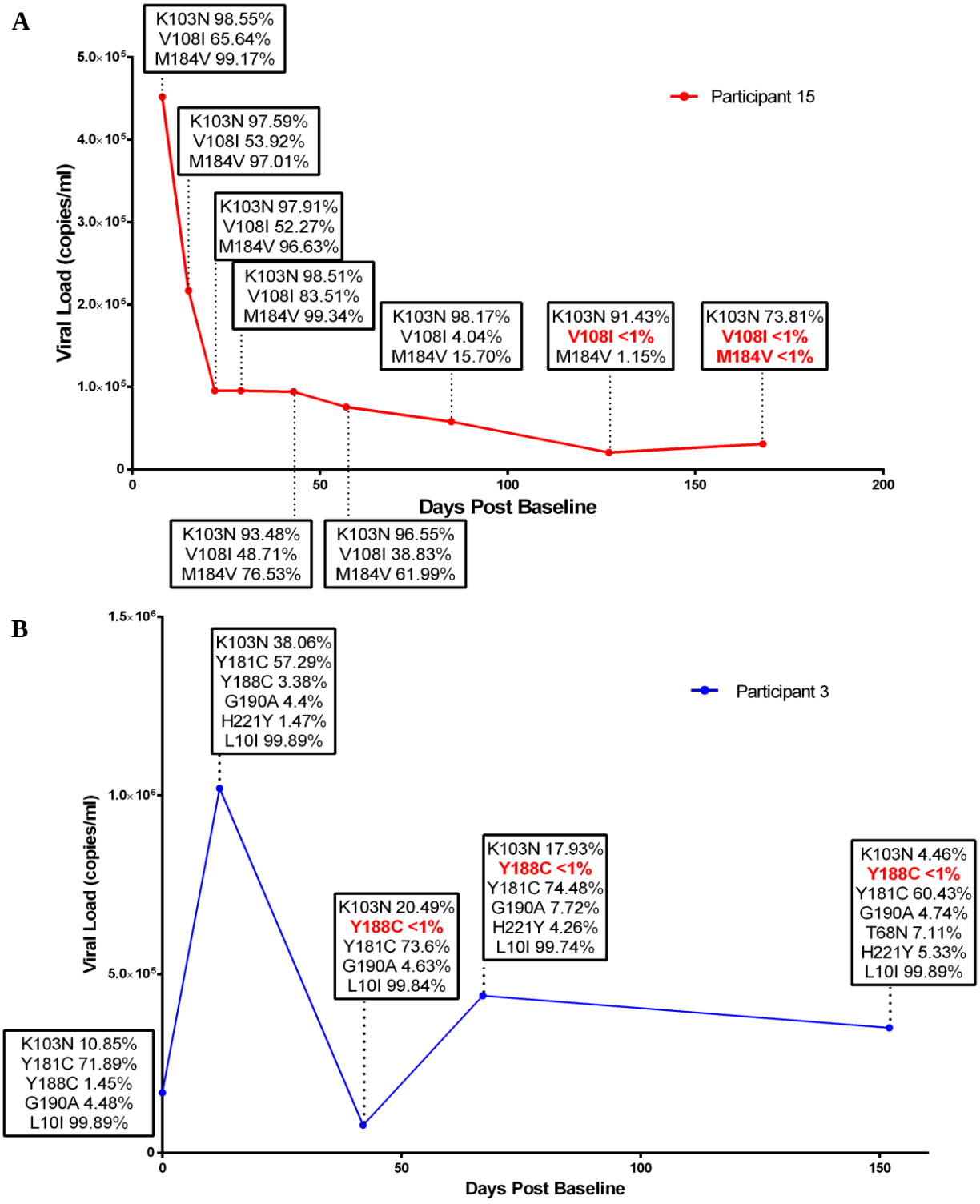


Figure 3.7.2: Viral load against follow up period in days for participants in which both persistence and reversion of TDRMs was observed. Participant 15 (**A**) and participant three (**B**) are shown in red and blue, respectively. TDRMs are written in red where detected below 1%.

Figure 3.7.2 A and B both illustrate a combination of persistence and reversion of TDRMs identified in participants three and 15. A ranger of mutations was detected for both participants. In participant 15, illustrated in **Figure 3.7.2 A**, V108I and M184V both were below 1% of the sequenced viral population by end-point follow up, whereas K103N remained a major mutation. The prevalence of V108I remained >50% of the sequenced viral population for the three longitudinal samples, and increased to >80% in the fourth. However, in the fifth a rapid decline was observed as the prevalence dropped to 4.04%. Soon after, in the sixth and the final seventh sample, the V108I mutation was no longer present at $\geq$ 1% of the sequenced viral population and was thus noted as have reverted. The M184V mutation had a similar rapid decline. For the first four follow-up samples, the prevalence was well above 97% of the sequenced viral population. In the fifth follow-up sample, a rapid decline to 15.70% was observed. The mutation soon reverted as in the sixth sample the mutation prevalence further declined to 1.15% and was ultimately no longer detected $\geq$ 1% of the sequenced viral population in the final seventh sample.

Figure 3.7.2 B illustrates the reversion of the TDRM Y188C, and persistence of K103N, Y181C, G190A, and the polymorphic L10I. Y188C was identified in participant three as a minor variant, at only 1.45% of the sequenced viral population. The mutation was not identified by Price *el* al., 2011, which used Sanger sequencing. The mutation prevalence increased to 3.38%, as a viral load spike occurred in the second follow-up sample, but thereafter declined and was no longer present $\geq$ 1% of the sequenced viral population for all subsequent follow-up samples (sample three, four and five). Y188C was thus reported as reverted.

Persistence of TDRMs, with no reversion observed was seen in 78% of participants (19 of 23 participants). **Figure 3.8** (**A** and **B**) shown below, illustrates the changes in viral load of participants which showed persistence of TDRM, relative to time. The participants were separated based on follow up time. Participants followed for less than 800 days are shown in **A**, and a follow up period of more than 800 days is shown in **B**.

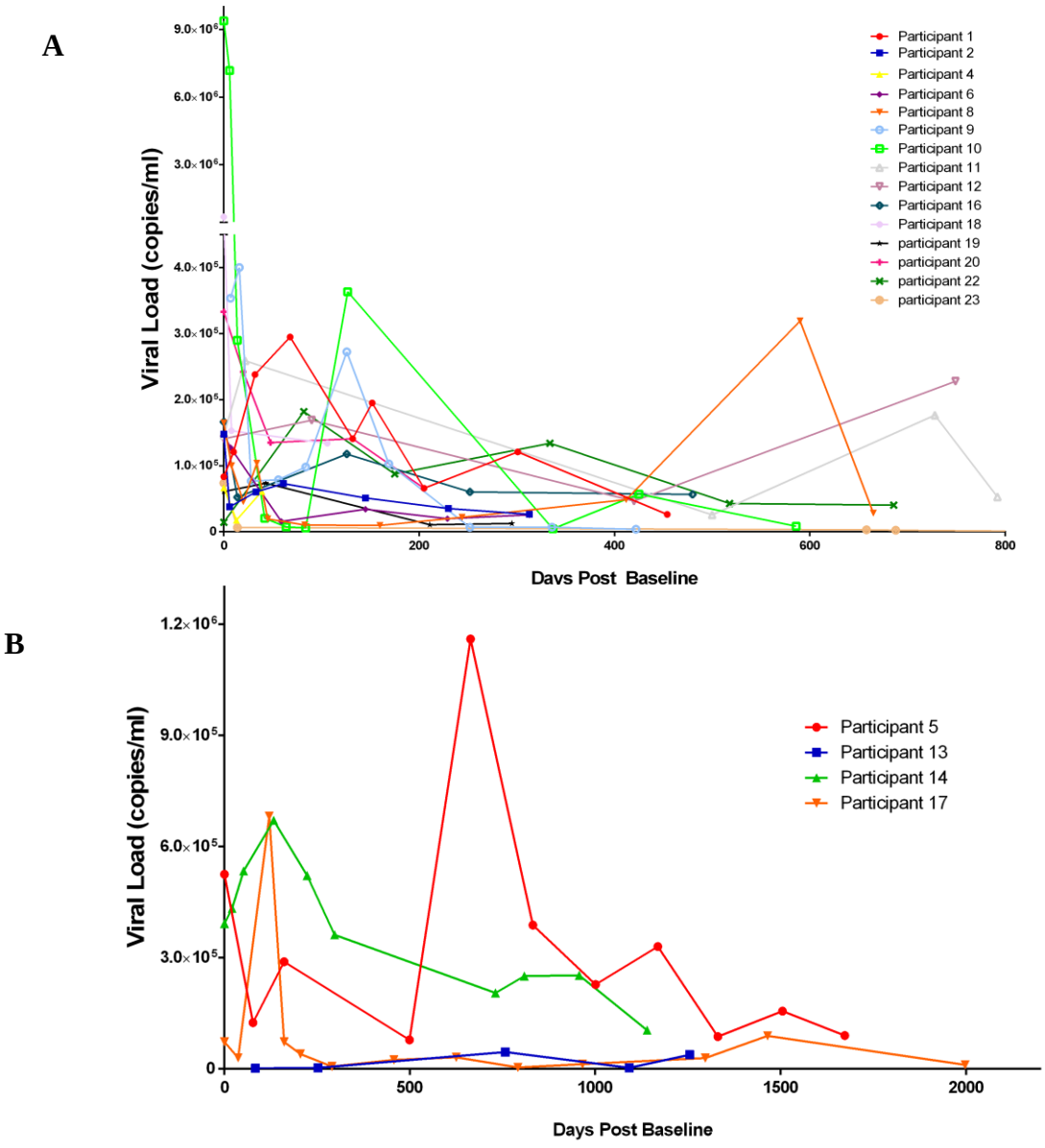


Figure 3.8: Viral load changes during the follow up period (in days) for participants in which no TDRM reversion was observed. The various participants are each shown in a different colour, and participants followed for less than 800 days are shown in **A**, and more than 800 days are shown in **B**.

Figures 3.8A and B illustrate the viral loads over time for participants in which persistence of all detected TDRMs are observed. Viral loads for the majority of participants appear to stabilise below $2x10^5$ viral copies/ml to the latter end of the longitudinal follow-ups. Despite the end-point viral loads, no consistent trend among the various samples presents itself, other than the progression from acute HIV-1 infection to the early stages clinical latency. None of the participants in this study initiated ART during the longitudinal follow-up.

Table 3.4 below summarizes the TDRMs that were detected by Sanger and confirmed by NGS, and their persistence or reversion outcomes at the end of the follow-up period. TDRMs that were not detected by Price et al. 2011, but detected in participants in this study are noted, as well as persistence or reversion outcomes.

Table 3.4: Summary table of TDRMs detected by Sanger and confirmed by NGS, and minority TDRMs detected by NGS only with observed outcomes

Participant	Mutations detected by Sanger and confirmed by NGS	TDRM Observed Outcome	Mutations detected only by NGS (as minority variant)	TDRM Observed Outcome
1	K103N	Persists		
2	M46L	Persists		
3	K103N	Persists	Y188C	**Reverts**
	L10I	Persists	G190A	Persists
	Y181C	Persists		
4	K103N	Persists	Y188C	Persists
			V108I	**Reverts**
5	I85V	Persists		
6	K103N	Persists		
	L10V	Persists		
7	D67N	**Reverts**		
8	K103N	Persists		
	L10V	Persists		

9	K103N	Persists		
	L10V	Persists		
10	Y188C	Persists		
11	K103N	Persists		
	E138A	Persists		
12	K103N	Persists		
13	T74S	Persists		
	K65R	Persists		
14	K103N	Persists		
	T74S	Persists		
15	K103N	Persists		
	K65R	Persists		
	V108I	**Reverts**		
	M184V	**Reverts**		
16	K103N	Persists	K65R	Persists
	T74S	Persists		
17	K103N	Persists		
	K103S	Persists		
18	I85V	Persists	V118I	Persists
19	L210W	Persists		
20	K103N	Persists		
21	M46L	**Reverts**		
22	K103N	Persists		
	V106M	Persists		
23	M41L	Persists		

CHAPTER 4: Discussion and Conclusion

4.1 Discussion

The use of ARV drugs has significantly reduced HIV-1 associated morbidity and mortality (Palella Jr et al., 1998). Global rollout of ART has allowed 17 million HIV-1 infected individuals to access lifesaving therapy (UNAIDS, 2016a). Unfortunately, an inevitable consequence of ART is the emergence of drug resistant HIV-1, ultimately resulting in ineffective treatment resulting in treatment failure (Deeks, Smith, Holodniy, & Kahn, 1997; McDonald & Kuritzkes, 1997). Drug pressure, sub-optimal treatment regimens and drug non-adherence, in combination with HIV-1's high replication rate, results in the unavoidable selection of drug resistance mutations (Reviewed by Clavel & Hance, 2004). It is well acknowledged that ADR may be transmitted to a new host via sexual transmission (Conlon, Klenerman, Edwards, Larder, & Phillips, 1994; Imrie, Beveridge, Genn, Vizzard, & Cooper, 1997).

In resource-rich geographical locations in which virological monitoring techniques, such as viral resistance genotyping, are readily performed, the transmission of viruses containing one or more drug resistance mutation ranges from 10 to 15% (Steinhauer et al., 1992; Walker et al., 1986). It is, however, important to note that ART has been available and more accessible in resource rich locations for a considerable period of time longer than lesser resource locations. In contrast to this, resource-limited geographical locations, such as Sub-Saharan Africa, see an estimated 5 to 10% prevalence transmission of drug resistant HIV-1 (Hamers et al., 2011; A. A. Johnson et al., 2004). Since the ART rollout in these resource-limited locations are generally not well monitored with virological genotyping, TDR HIV-1 is likely to increase, with significant clinical and public health consequences. Globally, there is a paucity of data on viral diversity and evolution of TDR HIV-1. This study describes the evolutionary mechanisms of ARV drug resistant HIV-1, after transmission to a new host, to provide insight into persistence and/or rates of reversion of the TDRMs to wild type over time. A total of 159 longitudinal samples from 23 participants from the IAVI protocol C cohort were studied.

4.1 Data Quality

The data obtained in all four of the NGS sequencing reactions were of satisfactory quality to identify mutations ≥1% of the sequenced viral population. The average coverage obtained across the 1.8 kb region of *pol* far surpassed the minimum required coverage. All metrics, post data quality control, were sufficient to proceed to downstream analysis.

The number of samples across the four sequencing reactions ranged from 38 to 45 sample per run. Based on the high coverage obtained from sequencing this number of samples, the sequencing reaction may be optimised to include a much higher number of samples that the minimum amplicon coverage to detect minority variants is still obtained. On the Illumina Miseq platform, using dual index libraries, 384 uniquely indexed samples may be sequenced together in one reaction (Illumina, 2016a). Increasing the number of samples per sequencing reaction will significantly reduce the price per sample. As NGS is cost prohibiting, a lower per-sample sequencing cost could allow clinical diagnostic laboratories to provide an NGS HIV-1 TDRM/DRM testing facility, especially in resource-limited developing countries.

4.2 Sanger Sequencing and NGS

Conventional Sanger sequencing has the capacity to identify HIV-1 mutations that are present at ≥ 20% of the sequenced viral population (Palmer et al., 2005). Palmer et al. (2005) used single-genome sequencing analysis (based on limiting-dilution assays) of viruses obtained from 26 patients with prior drug exposure to at least two ARV drug classes, and were failing treatment. Standard population-based Sanger sequencing was initially performed, and drug resistance mutations were identified. In a later experiment, the extracted viral RNA obtained from the infected patients was reverse transcribed, and a single copy of cDNA per positive PCR was diluted and obtained by using a real-time PCR method. Fifteen to 23 single sequences were obtained per patient. The single sequences were then amplified, and Sanger sequenced. The results to this study revealed that HIV-1 mutations present at ≤10% of the viral population were never identified by Sanger in the initial population-based sequencing method. Moreover, mutations present between 10% and 35% were only identified 25% of the time by Sanger sequencing (Palmer et al., 2005).

The inability of Sanger sequencing to identify low prevalence mutations presents a problem with significant clinical implications; the inability to detect HIV-1 minority variants (<20% of the

sequenced viral population). This emphasises the importance of the findings showing that NGS, on the other hand, has the capacity to identify mutations well below 1%; depending on the coverage obtained in the sequencing run (Li & Kuritzkes, 2013; Zagordi, Klein, Däumer, & Beerenwinkel, 2010). In the context of TDRMs, the capability to identify and study mutations at extremely low prevalence allows one to better understand two important factors. First, whether a mutation has truly reverted or persists at a low prevalence, and second, the clinical and treatment outcome implications a low prevalence mutation might have.

The current limitation with the Illumina Miseq sequencing by synthesis chemistry used in this study, is the short 300 bp read length. In the context of HIV-1, longer read lengths are preferred as it allows for the identification of DRMs that occur together in a unique viral sequence (i.e. combination mutations). Combination mutations have been shown to have an impact on the resistance profile of a viral strain (Liang, Mesplède, Oliveira, Anstett, & Wainberg, 2015; Tisdale, Alnadaf, & Cousens, 1997; Wensing et al., 2015). Currently, it is possible to identify mutations that occur together, but only when limited to the shorter 300 bp read. In the context of HIV-1 drug resistance, a limited number of mutations occur within 300 bp of one another. It would thus not be possible to elucidate all mutations that occur in a single viral genome by using the sequencing by synthesis chemistry. Other chemistries, such as the Pacific Biosciences single molecule real time sequencing, produces average reads > 10,000 bp, and has the capacity to produce reads > 60,000 bp. Using this method, the entire HIV-1 genome, which is approximately 9.2kb, could be sequenced, and all mutations occurring in the same genome in an individual unique strain could be identified. This would allow for the study of individual strains, and to better understand which combination of mutations lead to a survival advantage when transmitted to a drug naïve host.

4.3 Phylogenetic Analysis

Phylogenetic analysis served as a method to determine whether any sample contamination had occurred, and whether all participant longitudinal samples were genetically related. Each participant's samples clustered together significantly, indicating that all samples were indeed from the same participant, and that no significant sequence alteration was induced by PCR bias (please see **Figure 3.8**).

The phylogenetic analysis proved valuable in the context of participant seven. As discussed in 4.5, the baseline sample failed to amplify and could not be sequenced by NGS. The D67N mutation, that was detected by Sanger, could not be verified by NGS in the baseline sample, and was not detected in any subsequent longitudinal samples. Phylogenetic analysis, however, confirmed that the Sanger sequence, and all other consensus NGS sequenced, clustered significantly.

The phylogenetic analysis furthermore confirmed the HIV-1 subtypes that had previously been identified for each participant (Price et al., 2011).

4.4 PCR Bias

The PCR amplification of HIV-1 may result in the artificial inflation of mutations and the unwanted introduction of artefacts and errors (Hughes & Totten, 2003; Mansky & Temin, 1995; Perelson, Neumann, Markowitz, Leonard, & Ho, 1996). First, sequence diversity might be introduced by the misincorporation of nucleotides by the DNA polymerase. Second, PCR mediated recombination of HIV-1 viral strains may occur during the PCR amplification steps (Meyerhans, Vartanian, & Wain-Hobson, 1990; Yang, Wang, Dorman, & Kaplan, 1996). Third, as a large amount of DNA amplicons are generated from a small number of starting templates, the true original sampling of the populations may be obstructed. The starting templates may get resampled, which may lead to sequencing resampling, as opposed to the true observation of independent viral strains (Jabara, Jones, Roach, Anderson, & Swanstrom, 2011).

Errors introduced by PCR bias and PCR resampling are intrinsically extremely difficult to detect in the final sequenced product, as there is no way distinguish between what the true viral sequenced and the PCR error introduced sequence is. This is of importance when the mutations have drug resistance implications, and certain viral strains are resampled and amplified more than others. A skewed and incorrect representation of the true viral population may be obtained.

Work performed by Dudley et al. (2012) investigated the errors introduced by PCR amplification. The study PCR amplified, prepared and sequenced an HXB2 HIV-1 plasmid and HIV-1 clonal viral stock, that was not expected to have any mutations relative to the HXB2 reference genome, together with patient samples. The group reported no false drug resistance mutation above a frequency of 0.71%. Similar work done by Zagordi et al. (2010) found that in PCR amplified HIV-1 samples, the substitution error rate was approximately 0.25% across a 1.5 kb region of the

gag/pol genes, as opposed to 0.05% in non-PCR amplified samples. However, complete PCR resampling was observed at a lower percentage. In the case of significant resampling, one strain was severely inflated by PCR, which resulted in a skewed and incorrect population representation.

A limitation of this study was the fact that PCR bias and PCR resampling was not eliminated. Despite this, the fact that consistent prevalences were observed in several participant samples over time, where the samples had been extracted and RT-PCR amplified at different time points, implies that PCR bias/resampling was not a major issue.

Future research could focus on repeating the RT-PCR amplification using unique molecular identifier (UMI) tagging/molecular barcoding (Casbon, Osborne, Brenner, & Lichtenstein, 2011; Kinde, Wu, Papadopoulos, Kinzler, & Vogelstein, 2011; Van Laethem, Theys, & Vandamme, 2015). In the library preparation steps of NGS, each original DNA fragment in a sample is tagged with a unique string of unique nucleotides. In the resulting sequencing product, one would have the capacity to statistically identify the PCR induced artefacts and errors. The PCR bias affected sequencing products may then be bioinformatically discarded.

4.5 Drug Resistance Mutation Results

This study utilised next generation sequencing techniques, with unparalleled coverage and capacity, and allowed TDRMs to be identified and tracked in longitudinal samples at > 1% prevalence of the sequenced viral population. Overall, only six of the 43 (14%) total TDRMs identified in the 23 participants reverted and were undetectable at levels <1% at the last time point tested. Complete reversion of all detected TDRMs was observed in only two participants (8.7%). TDRM persistence of all detected mutations was observed in 19 of 23 (82%) participants. A combination of persistence and reversion of TDRMs were observed in two participants (8.7%).

The rates at which TDRMs revert, and thus become undetectable, are dependent on multiple factors. These include the number of required back mutations, the fitness of the mutated virus relative to the wild type, the rate at which the mutated virus replicated and thus viral turnover, and the presence of possible compensatory mutations (De Ronde et al., 2001; Goudsmit, De Ronde, Ho, & Perelson, 1996; Yerly et al., 1998). Due to HIV-1's high error rate, fitter viral progeny will inevitably arise due to random back mutation. A virus with specific mutations, or with lack of

mutations but with slightly higher replicative capacity, will compete with the mutated virus and ultimately outcompete it.

4.5.1 NNRTIs

NNRTI mutations identified in this study were K103N (n=14), G190A (n=1), Y181C (n=1), Y188C (n=1), E138A (n=1) and V108I (n=2). Of the 20 total NNRTI mutations identified, 17 (85%) persisted until end-point follow-up. NNRTI drugs that select for the above mutations are EFV (Bacheler et al., 2000; Reuman, Rhee, Holmes, & Shafer, 2010; Winslow et al., 1996), NVP (Gulick et al., 2004; Richman et al., 1991), RVP (Azijn et al., 2010; Rimsky et al., 2012) and ETR (Vingerhoets et al., 2005). Non-adherence to these drugs may lead to selection of a combination of the above-mentioned mutations, and transmission to a new host. A study by Nanfack et al. (2017) investigated the longitudinal persistence of TDRMs in 66 drug-naïve individuals by three different resistance testing methods. The group used Sanger sequencing, and thereafter compared results using an amplification-refractory mutation system (ARMS) PCR which was then sequenced with NGS. Results showed that NNRTIs were found to persist for over 3 years in the absence of drug pressure (Nanfack et al., 2017).

The NNRTIs K103N, G190A, Y181C, and Y188C were detected at baseline in participant three of this study. NVP selects for the exact combination of mutations identified in this participant. The Y181C mutation was detected at a 71.89% of the sequenced viral population, the K103N mutation at 10.85% prevalence, and the Y188C and G190A mutations were detected at 1.45% and 4.47%, respectively.

A study by Conway et al. (2001) identified this same mutation profile in patients who had received NVP, as monotherapy, or in combination with another NRTIs (AZT or DDI). Resistance mutations were studied by a phenotypic resistance assay, in which extracted viral genetic regions of interests were cloned into a pGEMT3ΔPRT plasmid and transfected into MT4 cells. The HIV-1 expressed by the cell culture was studied for drug susceptibility. The NVP resistant clones were subsequently Sanger sequenced and NVP resistance mutations identified. In one particular participant, the same combination of K103N, G190A, Y181C, and Y188C was identified (Conway et al., 2001).

Due to limitations in the Sanger sequencing technology, mutations that were identified by Conway et al. (2001) were present at a prevalence of >20% of the sequenced viral population. The mutation

prevalences identified in participant three in this study ranged from 1.45% to 71.89%. Comparing this study's findings with that of Conway et al. (2001) presents participant three as an interesting case.

A possible explanation for the resistance mutation profile observed in participant three is as followed: This male participant may have been infected with the drug resistant virus by a female who had previously received single dose nevirapine (sdNVP) in an attempt to prevent mother-to-child HIV-1 transmission. The sdNVP would have selected for the resistance profile observed, and thus transmitted to, participant three. This is a plausible explanation given that at the time of this study, sdNVP was readily available for use in Kigali, Rwanda (Delvaux et al., 2009).

The transmitting individual may have also had transmitted drug resistance, and may have been infected with a founder virus which carried Y181C. The rate of TDRM transmission was noted to be 8% for the IAVI protocol C samples used in this study (Price et al., 2011); a rate very similar to the 7.7% rate of TDRM transmission found before 2011, as found in a review of literature done by Stadeli et al. (2013) (Stadeli & Richman, 2013). This data may confer some plausibility to this first explanation.

If the transmitting individual did undergo sdNVP treatment, Y181C would have led to treatment failure (Richman et al., 1991), and furthermore selected for the additional NVP resistance mutations (K103N, G190A, and Y188C) (Conway et al., 2001). The initial viral reservoirs that were established at infection for the transmitting initial would carry the Y181C mutation, and thus would have supplemented the Y181C population. Upon transmission to a new host, Y181C would be observed as the dominant transmitted mutation, and the other NVP resistance mutations would be observed as minority mutations, as seen in participant three.

The low viral fitness costs of K103N (Collins et al., 2004; Palmer et al., 2006) and Y181C (Cong, Heneine, & García-Lerma, 2007; Martinson et al., 2007) (Dykes et al., 2001) may explain why these two mutations are identified at higher prevalences than the others. It may also explain how K103N proceeded to increase in prevalence to 38.06% and how Y181C decreased to 57.29% in 12 days after the baseline sample. The viral load in the 12-day period also increased over six-fold. Data from the transmitting individual would, however, be required to make significant conclusions with regard to this hypothesis.

4.5.1.1 K103N

The most prevalent mutation that was identified in the 23 participants from the IAVI protocol C was K103N (14/23; n= 61%), which persisted throughout study duration. The high prevalence of K103N, and persistence in the participants as a TDRM is consistent with findings by several other studies. Ferrerira et al. (2017) identified K103N at 77% of TDRMs, by population-based sequencing in a cohort of recently infected drug naïve patients. A study by Yebra et al. (2011) investigated an increase of transmitted drug resistance among HIV-1 infected individuals from Sub-Saharan Africa as a subpopulation in Spain, relative to the native Spanish population. The study, once again, identified K103N as the most frequent mutation (74% of identified TDRMs), using conventional population based Sanger sequencing (Yebra et al., 2011). Studies by Castro et al. (2013) and Little et al. (2008) both identified K103N as the predominant TDRM, at 73/313 (23%) and 11/14 (79%), respectively (Castro et al., 2013; Little et al., 2008). These studies found K103N to persist for up to between 3.4 and 3.7 years after infection (Castro et al., 2013; Little et al., 2008). As NGS has the capacity to detect mutations at a significantly lower prevalence of the sequenced viral population, the prevalence of K103N may have been higher if NGS was used.

K103N confers a relatively small fitness cost. Collins et al. (2004) created K103N clones by site directed mutagenesis; and generated and expressed several other mutant viruses (V106A, Y181C, Y188C, and G190A). Competitive assays were performed, and K103N was found to have a very small effect on the replicative capacity and viral fitness of the mutant virus (Derdeyn et al., 2004). A similar methodology was followed by Gerondelis et al. (1999) who also found that K103N has insignificant effects the viral fitness of the virus (Gerondelis et al., 1999).

K103N confers high-level resistance to EFV and NVP (Bacheler et al., 2000; Gulick et al., 2004; Margot, Lu, Cheng, & Miller, 2006). If transmitted to a treatment naïve individual, and not detected as a minority variant or tested for, the mutation will rapidly lead to treatment failure when treated with EFV or NVP regimens (Bacheler et al., 2000; Parkin, Gupta, Chappey, & Petropoulos, 2006). The mechanism of resistance for K103N involves the selective replacement of a lysine by an asparagine, that does not alter the overall complexed structure of the reverse transcriptase enzyme and the drug (NNRTIs) if bound. The asparagine, however, creates a hydrogen bond in the enzyme in the unbound drug state, which subsequently leads to the entrance of the drug binding pocket to be smaller. The smaller entry sterically hinders the NNRTIs from entering the pocket

(Clavel & Hance, 2004; Maga, Radi, Gerard, Botta, & Ennifar, 2010), causing a significant reduction in the binding of the drug to its target site. The K103N mutation thus minimally effects the enzymes activity, and viral fitness, causing K103N to emerge rapidly and to persist (Craigo et al., 2004; Dykes et al., 2001; Nicastri et al., 2003; Palmer et al., 2006).

Overall, K103N appears to be a very common mutation in the cohort used for this study, and additionally persists for a significant amount of time for all participants identified in this study. K103N additionally confers high level resistance to commonly used NNRTI drugs. It is thus of importance for any individual that is to initiate ART, to undergo an HIV-1 drug resistance test. Furthermore, it is important that a technique such as NGS is used, that has the capacity to detect K103N at both a major and minority level.

4.5.1.2 Y181C

Y181C was identified in participant three (1/23; n= 4.35%) and persisted until end point follow-up. Little et al. (2008) used conventional population-based Sanger sequencing and detected Y181C as a TDRM in 1/14 (7.14%) of the patients in the study. The study found that Y181C, as a TDRM, persisted for up to 46 weeks after the estimated date of infection of the individuals (Little et al., 2008). Another study by Castro et al. (2013) detected Y181C as a TDRM in 20 patients of a 313-patient cohort (6.4%), using population-based Sanger sequencing. Results of the study found the mutation to persist for over a median of 67.8 weeks (1.3 years). From previous reports, Y181C does not appear to be a common mutation, with reported prevalences being the approximately the same as in this study. The observations in this study are thus consistent with previous literature reporting on the persistence of this mutation.

Y181C has been shown not to reduce HIV-1's replicative capacity in any significant manner in-vitro. Work performed by De Luca et al. (2006) investigated the effect of Y181C on the replicative capacity of the virus by utilising a range of techniques. The group performed growth competition assays, replication kinetics assays, and single-cycle replication assays to determine that the fitness of a Y181C mutant was very similar to that of the unmutated wild type virus (De Luca, 2006). Growth competition assays performed by Iglesias-Ussel et al. (2002) reported that the Y181C mutation has near negligible effects on the replicative capacity and thus fitness of the virus (Iglesias-Ussel, Casado, Yuste, Olivares, & López-Galíndez, 2002).

Y181C confers high-level resistance to and selected NVP (Richman et al., 1991; Wu et al., 2012), and high-level resistance to EFV when used in conjunction with NVP (Winslow et al., 1996). Y181C is additionally selected by ETR (Vingerhoets et al., 2005) and RVP (Azijn et al., 2010). A 50-fold reduction to NVP is observed in Y181C mutants (Rhee et al., 2004), with a fivefold, threefold and twofold reduction to ETR (Vingerhoets et al., 2010), RPV (Rimsky et al., 2012) and EFV (Rhee et al., 2004) respectively.

The resistance mechanism of action for Y181C involves the selective substitution of a tyrosine to a cysteine amino acid at position 181. The NNRTI drugs, EFV and NVP, bind to the conserved tyrosine, which forms part of the p66 subunit of the reverse transcriptase enzyme near the polymerase active site. The subunit is subsequently involved in the formation of a hydrophobic pocket in a drug bound state. When EFV and/or NVP bind to the tyrosine residue, the hydrophobic pocket is created. This causes the surface area of the polymerase active site to significantly reduce, thus greatly hindering the polymerisation activity of the reverse transcriptase enzyme (Selmi et al., 2003). The lack of the tyrosine in an unbound drug state does not appear to impede the activity of the enzyme in any significant way, and thus does not seem to significantly reduce the fitness of the virus (De Luca, 2006; Selmi et al., 2003).

Overall, the persistence of Y181C over 5 months, noted in participant three, suggests that the mutation does not have a major impact on viral fitness; particularly evident by the high viral loads (Please refer to Table 2 in Appendix D). Since our study could not look at mutations that occurred in combination, future work could focus on the effects of Y181C on viral fitness, not only as a single mutation, but as part of combination mutations. In addition to this, considering the fact that Y181C persists and confers high level resistance, it is important to consider Y181C before initiating treatment.

4.5.1.3 G190A

G190A was also identified in participant three (1/23; n=4.35%) and persisted as a minority mutation (<10% of the sequenced population) throughout the follow up period of 5 months. The persistence of G190A, as observed in this study, has been reported in previous studies. Castro et al. (2013) identified G190A in 17/313 (5.43%) patients using conventional population-based Sanger sequencing. The study found that G190A persists as a TDRM and as a major mutation

(>20% of sequenced population) for over 3.6 years (Castro et al., 2013). Another study by Shet et al. (2006) used an allele-specific real-time PCR method to detect common TDRMs, such as G190A, and found the mutation in 2/112 patients (1.8%). The mutation persisted for over 3 years (Shet et al., 2006). The persistence of G190A can be attributed to the fact that the mutation has been shown to have a minimal effect on replication efficacy relative to wild type and K103N. Gerondelis et al. (1999) investigated viral fitness by observing the effects that induced mutations have on the RNase H cleavage ability of HIV-1 RT. RNase H cleavage is the process in which the original viral RNA strand is removed in order to complete the RNA to cDNA reverse transcription, prior to host genome integration (Wisniewski, Balakrishnan, Palaniappan, Fay, & Bambara, 2000). The rate at which RNase H cleavage occurs dictates the rate of reverse transcription. In a G190A mutant, a slight reduction in the rate of DNA 3′-end-directed RNase H cleavage relative to wild type RT was observed, which indicated a slightly less fit mutant (Gerondelis et al., 1999).

The G190A mutation is selected by, and confers high-level resistance to, NVP and EFV (Bacheler et al., 2000; Reuman et al., 2010). G190A reduces NVP susceptibility by over 50-fold (Rhee et al., 2004), and EFV susceptibility by five to ten fold (Huang, Gamarnik, Limoli, Petropoulos, & Whitcomb, 2003). The mechanism of action in which G190A results in resistance is caused by the selective replacement of a glycine by an alanine. The methyl group on the alanine, in proximity to the drug binding site, results in steric hinderance to the NNRTI inhibitor. In a mutated state, the inhibitor has a reduced binding efficiency, and the formation of the hydrophobic pocket which results in enzyme dysfunction is not achieved (Lyidogan & Anderson, 2014). It is important to note, however, that the alanine side chain in the mutated virus leads to hyper susceptibility to DLV, which results in a favourable interaction with the methyl group (Lyidogan & Anderson, 2014). Despite the selective glycine alanine substitution, it does not appear to hinder the function of RT, and thus does not significantly affect the fitness of the mutated virus (Gerondelis et al., 1999).

Overall, G190A persisted in participant three without any significant effect on the viral fitness of the virus. This is consistent with the literature surrounding this mutation (Gerondelis et al., 1999; Shet et al., 2006). The lower prevalence persistence of G190A, as compared to K103N and Y181C, indicates that this is a less fit mutation relative to K103N and Y181C. Despite being a less fit variant, the fitness cost of G190A was not sufficient to lead to reversion in the follow up period. Future research could focus on G190A as a combination mutation, with mutations such as K103N

and Y181C, and explore how the viral fitness is affected. Despite G190A not appearing as a commonly transmitted mutation, the mutation persists for a period after infection and confers significant drug resistance against NVP and EFV.

4.5.1.4 Y188C

Y188C was identified as a TDRM in participant three, four and 10 (3/23; n=13%). Varghese et al. (2009) used ultra-deep pyrosequencing on the Roche 454 platform to identify TDRMs in individuals that were failing ART. The study identified Y188C in 1/12 (8.3%) of the study participants (Varghese et al., 2009). Another study by Ndembi et al. (2008) investigated the prevalence of TRDMs in recently infected drug naïve individuals in Cameroon, by means of Sanger population based sequencing, and identified the Y188C mutation in 2/79 (2.53%) individuals (Ndembi et al., 2008). The two studies above did not, however, study the persistence of reversion of Y188C.

In participant three, Y188C was identified at 1.45%. as a minority mutation (< 20% of the sequenced viral population), and reverted after 12 days. Fitter viral strains, that have been shown to have higher viral fitness relative to Y188C (such as K103N, Y181C or G190A), may have outcompeted the minority variant. Research performed by Collins et al. (2004) created several HIV-1 strains which contained a single drug resistant by means of site directed mutagenesis, which were used to generate viral stocks, and used in competition assays. Results showed that Y188C was the weakest variant relative to other major NNRTIs, and that the fitness hierarchy was as follows: wild type > K103N > Y181C > G190A > Y188C (Derdeyn et al., 2004). This viral fitness hierarchy was confirmed by Gadhemsetty et al. (2010), and subsequently Y181C is considered an intermediate fit mutation relative to the other mutations identified in participant three.

In participant four, Y188C was identified as a minority mutation, and persisted below 10% of the sequenced viral population until endpoint. It is important to note, however, that the follow-up period for participant four was short (39 days), and that the prevalence that Y188C was identified at in the last sample was 1.59%. A longer follow-up period would have afforded the study an opportunity to observe Y181C in combination with other fitter mutations (such as K103N), and the persistence/reversion outcome of the mutation.

Y181C persisted as a solitary mutation in participant 10. The mutation was detected well over 99% of the sequenced viral population until 13 months post the baseline sample, and declined to 39.4% in the last sample five months later. As a solitary mutation, the population prevalence would be high for the first year. However, the weaker Y188C mutant would gradually start to decline. This is due to the fact that back mutations by random mutagenesis may allow for fitter/wild type strains to emerge. There is a possibility that PCR bias could have occurred in the last sample of participant 10 (sample 79). The sharp decline of the Y181C prevalence is not consistent with prevalences observed throughout the follow-up period. This could be due to the sudden detection of V82A, as well as the relatively low viral load in the last sample (which would cause a limited PCR template). Nonetheless, the Y188C mutation persisted for the entirety of the follow up period of 586 days.

Y188C is selected by, and confers, high-level resistance to NVP (Richman et al., 1991; Wu et al., 2012) and EFV (Winslow et al., 1996). The mutation causes at least a 50-fold reduction in susceptibility to NVP (Rhee et al., 2004), and a 20-fold reduction to EFV (Vingerhoets et al., 2010). The mechanism of resistance by which Y188C leads to resistance, is caused by the cysteine which replaces the wild type tyrosine. The cysteine lacks the aromatic side chain; a necessary component for extensive stacking interactions between the tyrosine and the bound NNRTI inhibitors (J Ren et al., 2001). The wild type tyrosine at position 188 is moreover found at the top of the NNRTI binding site and leads to interactions between the aromatic side chain to a tryptophan at position 229. This then leads to a hydrophobic sub-pocket which facilitates the binding affinity of NNRTI inhibitors (J Ren et al., 2001).. The lack of the tyrosine chain leads to a decrease in the binding affinity of NNRTIs. Ren et al. (2001) found an 83-fold reduction in the affinity of NVP to the RT drug active site in the Y188C mutant relative to wild type (J Ren et al., 2001). The cysteine also, however, causes some conformational changes in the mutant enzyme, which subsequently results in a less efficient RT enzyme and negatively affects the efficiency and the general fitness of the virus (J Ren et al., 2001).

Overall, Y188C has the potential to persist (as seen in participant 10), and additionally confer resistance to commonly used ARVs. Therefore, it is an important mutation to include in resistance screening. Although not commonly detected, the mutation may lead to treatment failure if not detected prior initiation of ART.

4.5.1.5 E138A

The E138A mutation was detected in participant 11 (1/23; n=4.35%) at baseline at a prevalence of >99% of the sequenced viral population. The mutation furthermore persisted at >99%, until the end of the follow up period of 792 days, and additionally persisted in combination with K103N (which was also identified at baseline at 98.79% of the sequenced viral population). The viral strain identified in participant 13 appears to carry both E138A and K103N mutations in the same viral genome. The addition of E138A to K103N in the same virus does not appear to hamper the viral fitness of K103N in any significant manner, as viral loads and prevalences remained high throughout the follow-up period (Please refer to table 2 in appendix D). E138A is a common polymorphic mutation which has proven not to affect the fitness of the virus. McCullum et al. (2013) performed a study in which several mutations at position 138 were generated and used in competition assays to study relative mutant fitness. The study found that among the different mutations (namely, E138A, E138G, E138K, E138Q, and E138R, E138A and E138G) were the most fit substations (McCallum et al., 2013). Additionally, the fitness relative to wild type was found to be near identical (McCallum et al., 2013).

A study by Pingen *el* al. (2014) used population-based Sanger sequencing to sequence virus isolated from newly infected, drug naïve individuals, and identified E138A as a TDRM in 1/31 patients (3.23%). In addition to this, the mutation was observed to persist for over 10 months (which was the follow-up cut off for this study)(Pingen, Wensing, et al., 2014). Further research could contribute to a better understanding of the effects of viral fitness when in combination with other mutations, such as K103N observed in this study. The mutation combination does not appear to have a negative effect on the viral fitness.

The selective replacement of glutamic acid to alanine in E138A leads to a slight affinity reduction of the NNRTI inhibitor to the drug binding hydrophobic pocket (Tambuyzer, Nijs, Daems, Picchio, & Vingerhoets, 2011). Limited research has been published on E183A mechanism of action, as the polymorphic mutation leads to low-level resistance to NNRTI drugs (Tambuyzer et al., 2011). The change from an acid side chain of glutamic acid to a hydrophobic side chain of aniline may lead to a slight conformational change. This, in turn, might alter the NNRTI the hydrophobic drug binding pocket and consequently slightly alter the binding affinity. E138A confers low-level resistance to, and is weakly selected by, ETR (Tambuyzer et al., 2011) and RPV (Haddad et al.,

2011). The mutation leads to a twofold reduction in RPV and ETR susceptibility (Haddad, Stawiski, Benhamida, & Coakley, 2010). A study by Porter et al. (2016) detected the presence of E138A in peripheral blood mononuclear cells (PBMC) DNA at baseline in a longitudinal study. The study administered ART and suppressed viral replication in infected individuals. Interestingly, regardless of the presence of E138A, four patients who received treatment containing TDF/FTC/RPV did not lead to virological rebound (Porter et al., 2016).

In line with previous findings, the high prevalence persistence of E138A does not appear to affect the viral fitness of the of the mutation in any significant manner (McCallum et al., 2013; Pingen, Wensing, et al., 2014). The effects of combination mutations (such as E183A and K103N) seen in participant 11, were not investigated in this study; thus, future research could focus on how polymorphic mutations (such as E138A) affect the persistence/reversion of mutations such as K103N. E183A does confer low level resistance and appears to persist.

4.5.1.6 Accessory mutation V108I

V108I is considered a non-polymorphic accessory mutation (Wu et al., 2012). This mutation was detected at 65.64% at baseline in participant 15, and detected as a minority variant, at 2.96%, in participant four (2/23; n=8.7%). At the time of this study, there was limited literature on the persistence of V108I as a TDRM, as it confers only low-level resistance to both EFV (Wu et al., 2012) and NVP (Reuman et al., 2010). It appears that the reversion of V108I, in combination with other NNRTI mutations, was achieved in a short amount of time. Interestingly, even though the mutation was present at high proportions in participant 15, the mutation was lost at 56 days post baseline. In participant four, it was lost within 39 days. V108I does not appear to confer any significant survival or replicative advantage to the virus; even when in combination with other mutations. This is evident in the fact that the mutant was detected as a major mutation in participant 15, and a minority mutation in participant four, and in both instances, it reverted in a short amount of time.

V108I is selected by NVP and EFV, and causes low-level resistance to both (Vermeiren et al., 2007). A twofold reduction in drug susceptibility to NVP and EFV is observed when V108I is present (Rhee et al., 2004).

It is evident that the rapid reversion of V108I observed in this study appears to hamper the replicative capacity and fitness of the virus. This is due to the fact that the variants, with fitness comparable to that of wild type (K103N), persisted and outgrew the mutation. Overall the mutation confers low-level resistance and does not appear to persist.

4.5.2 NRTIs

NRTIs that were identified in this study are K65R (n=2), D67N (n=1), M184V (n=1) and L210W (n=1). M184V was identified in participant 15 and K65R was detected as a minority variant in participants 15 and 16. NRTI TAMs detected were D67N in participant seven and L210W in participant 19. Of the 5 NRTI mutations identified, 2 (40%) of the mutations persisted, and 3 (60%) reverted below 1% of the sequenced viral population. NRTI drugs, such as 3TC (Back et al., 1996), FTC (Wang et al., 2004) and TDF (White et al., 2005), are commonly used as regimens in ART, and select for NRTI mutations identified in this study. At the time of sample collection for this study (2006 – 2011), the above-mentioned drugs were available for use as ART in the IAVI protocol C countries (Kenya; Uganda; Rwanda; Zambia; South Africa).

4.5.2.1 M184V

The M184V mutation was detected in one participant (1/23; n=4.35%), participant 15, at a prevalence of 99.2% of the total sequenced population in the first baseline sample. A study by Castro et al. (2007), explored a cohort of 313 recently HIV-1 infected and drug naïve patients, and identified M184V as a TDRM in 34 (10.86%) of these patients. The Castro study additionally used population-based Sanger sequencing to identify TDRMs and found M184V to revert in all 34 patients in which the mutation was identified. The mean time of reversion was approximately one year, although the study found that the minimum time M184V was found to be undetectable was five months. The reversion time of the mutation might have been longer, if NGS technology with a lower limit of detection, was used to longitudinally follow the persistence/reversion of mutations. A study by Little et al. (2002) identified M184V by population-based sequencing in 5 out of 258 (1.94%) recently infected drug naïve patients – no longitudinal follow-up was performed.

K103N was identified in combination with M184V at a prevalence of 98.6% at baseline in participant five. Due to the high prevalence of both M184V and K103N, it is likely that both mutations may have been contained and co-expressed in a single viral genome. It is evident that

K103N is more fit as a single mutation, as opposed to when in combination with M184V, as it's prevalence remained >90% until the last sample, where only then it dropped to 73.81%. M184V's prevalence, however, steadily declined over the follow up period until it was eventually undetectable below the cut-off of 1% in the final sample. The viral strain that co-expressed both mutations may have been the dominant strain in the quasispecies shortly after infection. However, as the viral load declined, so did the prevalence of M184V, and thus the strain which contained both K103N and M184V. Work performed by Cong et al. (2007) found that a strain which carried both K103N and M184V mutations had a 16.1 times reduction in viral fitness relative to wild type. The group generated recombinant virus with various single or combination mutations by transfecting MT-4 cells with cloned RT sequences from patients and an RT depleted proviral molecular clones. Fitness was then subsequently assessed in competition assays (Cong et al., 2007). The M184V mutation is thus generally considered an unfit mutation, and its rapid reversion due to its high fitness cost is well reported (Cong et al., 2007; Harrison et al., 2010). The V108I mutation was detected at 65.6% at baseline in participant 15, when in combination with M184V and K103N. The accessory mutation may have been co-expressed in some of M184V-K103N dual mutation strains. However, as the viral load declined, and both V108I and M184V reverted, it does not appear that V108I had any significant effect on the persistence of M184V.

M184V is selected by, and confers high-level resistance to, 3TC and FTC with a >100-fold reduced susceptibility (Grant et al., 2010; Harrigan et al., 2000; Whitcomb, Parkin, Chappey, Hellmann, & Petropoulos, 2003). The mutation may also be selected by, and cause low level resistance to, ABC and DDi (Harrigan et al., 2000; Whitcomb et al., 2003). It is important to note that M184V increases susceptibility to AZT, d4T and TDF, and hinders the formation of resistance to these drugs (B. A. Larder, Kemp, & Harrigan, 1995; Masquelier et al., 1999). The mechanism of action by which the M184V mutation causes resistance, involves the selective substitution of a methionine to a valine near the drug binding site in the RT enzyme. The smaller non-polar side-chain of valine lacks a thiol group which would otherwise be used for the formation of disulphide bonds, and thus for secondary and tertiary protein structure formation (Clavel & Hance, 2004). The altered structure of the M184V mutant protein causes an increase in drug discrimination relative to the wild type protein, which ultimately leads to a greater capacity to distinguish between dNTPs and non-extendable nucleoside analogue monophosphate NRTI drugs. In addition to this,

the mutant RT is less efficient, which imposes a fitness cost on the virus (Clavel & Hance, 2004; Harrison et al., 2010).

M184V is an overall unfit mutant that confers high resistance to 3TC/FTC. M184V is reported, and seen in participant 15, to have a reduced fitness (Harrison et al., 2010). The degree to which fitness is reduced appears to depend on the co-expression of other mutations. K103N, in combination with M184V, imposes a significant fitness cost. Further work could investigate how accessory mutations, such as V108I, affect the fitness and the persistence/reversion of the virus. Despite M184V having a high fitness cost, and low persistence potential, the mutation still confers high level resistance to 3TC/FTC (Harrigan et al., 2000), and appears to have the capacity to be transmitted and establish infection in a new host.

4.5.2.2 K65R

K65R was detected as a minority variant in two participants (2/23; n=8.7%); participant 15 and 16. It persisted in both participants until endpoint at a prevalence of between 1-3% of the sequenced viral population. K65R was not detected in baseline in participant 13, but emerged approximately 2 months after the first baseline sample. Castro et al. (2013) identified K65R in 5/313 (1.6%) of a cohort of drug naïve and newly infected HIV-1 individuals. A study by Metzner et al. (2011) identified K65R in 4/146 (2.7%) of recently HIV-1 infected drug naïve patients, by using allele-specific real-time PCR (Metzner et al., 2011). In both participants 15 and 16, K65R persisted at a low prevalence of between 1-3% throughout follow up. Work by Cong et al. (2007) used single mutation competitive assays and found the relative viral fitness of K65R relative to wild type to be low. More fit, or wild type virus would rapidly outcompete K65R mutants (Harrison et al., 2010). A study by Deval et al. (2003) moreover found *pol* recombinant viruses, that contain the K65R mutation, to have a reduced replicative capacity of up to 53%, relative to wild type. Generated recombinant viruses were subjected to single cycle virus growth assays in order to determine replicative capacity (Deval et al., 2004).

The K65R mutation is selected by TDF, ABC, d4T and DDi (García-Lerma et al., 2003; Wainberg et al., 1999), and reduces the susceptibility to these 2 fold and d4T 1.5 fold (Petropoulos et al., 2000). The mechanism of action for resistance by K65R is similar to that of M184V; K65R is a

mutation that occurs near the drug binding site and ultimately causes in increase in drug discrimination (Clavel & Hance, 2004).

As K65R is considered an unfit mutation that confers low-level resistance, the low-level persistence in both participant 15 and 16 is interesting. K65R was detected with K103N in both participants. It would appear that the co-existence of separate strains that carry K103N and K65R would allow for the less fit K65R virus to replicate at a level that is sufficient to maintain a low prevalence. The combination of K65R and K103N, as TDRMs, have been observed and reported previously in literature (Metzner et al., 2011), but this study used allele-specific real-time PCR, and thus prevalences or persistence/reversion were not reported. Future work could investigate what would allow for the low-level persistence of K65R to occur with a fitter virus, such as K103N seen in participant 15 and 16.

4.5.2.3 D67N

D67N was identified and rapidly reverted in participant seven (1/23; n=4.35%). Castro et al. (2013) identified D67N as a TDRM in 27/313 (8.6%) of recently infected patients using population based Sanger sequencing (Castro et al., 2013). In contrast to this, a much lower prevalence of D67N as a TDRM was found, also using population based Sanger sequencing, in a larger study by Yerly et al. (2007); specifically, only 9/822 (1.1%) of recently HIV-1 infected individuals (Yerly et al., 2007). In summary, Castro et al. (2013) and Yerly et al. (2007) found D67N to persist for a significant period after infection (> 6 years) in drug naïve individuals.

A study by García-Lerma et al. (2004) produced recombinant virus with patient derived RT sequences that contained D67N as a single mutation. A RT-deleted HXB2-based proviral molecular clone was used for the expression of the D67N mutant recombinant virus, which was then used in fitness assays and growth competition assays. The study found that D67N could replicate efficiently in the absence of a drug pressure, and had a potential to persist for a significant amount of time (> 5 years) (García-Lerma, MacInnes, Bennett, Weinstock, & Heneine, 2004). A similar study by Cong et al., (2007) evaluated the fitness of 11 key *RT* mutations, of which D67N was one. Mutant virus was created by site-directed mutagenesis and expressed to be used in fitness assays. Results showed that D67N was a fit virus as a single mutation, and imposed very limited fitness cost (Harrison et al., 2010).

D67N, as a type 2 TAM, is selected for and results in low-level reduced susceptibility to AZT and d4T (García-Lerma, Nidtha, Blumoff, Weinstock, & Heneine, 2001). TAMs' mechanism of resistance involves a process termed 'pyrophosphorolysis'; a reaction that removes chain-terminating NRTI drugs and reinstates an extendable primer, consequently facilitating primer unblocking (Clavel & Hance, 2004).

The baseline sample, in which D67N was identified using Sanger sequencing, by Price et al. (2011), could not be RT-PCR amplified and subsequently sequenced by NGS. However, as the mutation was identified by Sanger, and not NGS, the prevalence can be assumed to be ≥20% of the sequenced viral population (Palmer et al., 2005). To confirm genetic relatedness of the samples, phylogenetic analysis was performed (please refer to the results displayed in **Figure 3.8**). All samples for participant seven, including the baseline sample, grouped in a cluster, thus indicating that all samples are from the same patient. The D67N mutation was thus concluded to have reverted in participant seven. This outcome of D67N has not yet been observed in previous publications and warrants further investigation.

4.5.2.4 L210W

The final NRTI and TAM that was detected was L210W, and identified in participant 19 (1/23; n =4.35%). The mutation was identified by NGS at a prevalence of 13.85% of the sequenced viral population; increasing to a prevalence of 92.62% in the end point sample. L210W was also found to persist throughout follow-up. Castro et al. (2013) identified L210W, as a TDRM, in 25/313 (8%) of newly HIV-1 infected, drug naïve patients using Sanger sequencing, and found the mutation to persist for up to 4.8 years (Castro et al., 2013). A study by Boden et al. (1999) identified L210W in 1/80 (1.3%) of individuals identified to have transmitted drug resistance; however, longitudinal data on these participants was not explored A study by Little et al. (2008) found L210W in 1/14 (7%) of participants who were drug naïve and recently infected. The study found the mutation to persist for up to 170 weeks (3.3 years) after infection (Little et al., 2008).

L210W is considered a very fit mutant with a low fitness cost. Work performed by Cong et al. (2007) used single mutant analysis between HXB2 wild type and a L210W mutant, by producing clones using site directed mutagenesis. The mutants were expressed and used in growth

competition assays. Following this, the group could not identify a significant difference in viral fitness between L210W and wild type virus (Harrison et al., 2010).

L210W is considered a classical, type 1 TAM. It is selected by ABC, TDF and ZDV; conferring low level resistance to only AZT (Centers for Disease Control and Prevention, 2016; World Health Organization, 2016a). The mechanism of action includes the process of 'pyrophosphorolysis', as described in D67N's mechanism of action.

L210W does not appear to have any significant effect on viral fitness, as viral loads remained high throughout follow-up in participant 19. Although L210W does not confect great resistance to ARV drugs, the mutation has potential to persist for a significant amount of time after infection (Little et al., 2008). L210W in combination with other mutations warrant more work, to perhaps better understand how L210W as a TAM in combination with other NRTIs/NNRTIs may affect mutation persistence/reversion.

4.5.3 PIs

The PI mutations identified were M46L (n=2), and the accessory mutations identified were: T74S (n=3), I85V (n=2), L10V (n=3), and L10I (n=1). PI based regimens are not commonly administered, and are used only in individuals that are failing commonly used first line regimens (World Health Organization, 2015). The combination of TDF, 3TC/FTC and EFV, on the other hand, are commonly administered; they require less frequent administration, have fewer adverse side effects, and have better virological treatment response (World Health Organization, 2015). Despite this, the PI based treatments may have better outcomes because of the increase of drug resistance.

4.5.3.1 M46L

M46L was identified in two participants (2/23; n=8.7%); as a single mutation in participant two, and as a combination mutation in participant 21. The mutation persisted in participant two, but reverted to wild type in participant 21. Little et al. (2008) used Sanger sequencing to identify M46L as a single mutation in 2/14 (14.3%) of recently HIV-1 infected, drug naïve patients, and found the mutation to persist for up to 170 weeks (3.3 years) post infection (Little et al., 2008). Castro et al. (2013) identified M46L as a TDRM (both as single and in combination) in 16/313 (5.0%) of

newly HIV-1 infected individuals, and found the mutation as a single mutation to persist for up to 3.1 years (Castro et al., 2013).

M46L, as a lone mutation, does not appear to significantly affect the replicative capacity and viral fitness of transmitted viruses. M46L, when in combination with other common mutations such as K103N, G190A, and Y181C, does not persist and reverts within a short period of time (<1 year) (Harrigan et al., 2000). The reversion of M46L, as a combination mutation, was observed in participant 21 in this study. A study by Pingen et al. (2014) produced mutant HIV-1 by means of site directed mutagenesis contained M46L as a single mutation, as well as a combination strain that consisted of several strains of different common mutations. The study found that M46L had a significant decrease in replicative capacity and viral fitness when in combination with such as K103N, G190A, and Y181C; ultimately leading to the reversion of the mutation within 1 year (Harrigan et al., 2000).

M46L is considered a non-polymorphic PI selected mutation; it is primarily selected by IDV, NFV, FPV, ATV and LPV, and leads to potential low-level resistance to ATV and LPV (Cann & Karn, 1989; Garriga et al., 2007; Little et al., 2002).

Overall, M46L is a mutation that confers low level resistance, and may persist or revert, depending on its combinations with other mutations. In participant 21, there was no consistent combination mutation with which M46L was identified. Low prevalence mutations (<20% of the sequenced viral population), such as K219R, K103N and G190A, were identified at various time points throughout the follow up. The prevalence of M46L, however, steadily declined until no longer detected (>1% of the sequenced viral population) 1.4 years after the first sample. The time frame of the reversion of M46L, as a combination mutation, in participant 21, is consistent with the findings of Castro et al. (2013). The follow-up for participant two was, however, shorter (0.85 years), and based on previous reports, would persist as a solitary mutation (Harrigan et al., 2000). Additional work may be required to further investigate the effects of M46L in combination with other PIs, NNRTI, and NRTI mutations.

4.5.3.2 Accessory PI mutations T74S, I85V and L10I/V

T74S, I85V and L10I/V are polymorphic accessory mutations. T74S was identified in participant 13, 14 and 16 (3/23; n=13%) and persisted in all 3 participants. T74S confers no resistance to any

PI drugs currently in use. The persistence and low effect on the viral fitness of T74S as a TDRM, has been previously reported and consistent with literature (Gonzalez et al., 2008; Price et al., 2011).

I85V was identified in participants 5 and 18 (2/23; n=8.7%) and persisted until endpoint in both. I85V is a non-polymorphic PI-selected mutation that does not, in any manner, affect PI susceptibility (Rhee et al., 2004). In addition, I85V does not reduce the viral fitness of the viral strain (Robinson, Myers, Snowden, Tisdale, & Blair, 2000). The polymorphic accessory NRTI V118I mutation was detected with I85V in participant 18; both persisting well above 99% prevalence of the sequenced viral population until end point. The virus fitness does not appear to be affected by this combination of mutations. Persistence observed is consistent with literature reports (Robinson, Myers, Snowden, Tisdale, & Blair, 2000).

L10I/V were commonly detected in participants 3, 6, 8, and 9 (4/23; n=17.4%) at very high prevalences (>90% of sequenced viral population) and persisted in all participants. L10I/V mutations are PI selected, polymorphic accessory mutations, that may increase the replicative capacity of the virus when in combination with other PI selected resistance (Tzoupis, Leonis, Mavromoustakos, & Papadopoulos, 2013).

4.2 Conclusion

This study describes the population dynamics of transmitted drug resistant HIV-1, and the persistence/reversion of the transmitted drug resistant mutations, to treatment naïve individuals. The use of NGS technology provided unprecedented coverage and capacity to observe the dynamics of TDRMs confidently at low prevalence's (≤ 1% of the sequenced viral population). The use of NGS technologies enable the detection of minority drug resistance variants, otherwise missed by conventional population-based Sanger sequencing. Moreover, NGS technologies allow for a higher number of samples per resistance testing sequencing reaction; effectively reducing the price per sample per HIV-1 resistance test.

Findings confirmed that a range of ARV drug resistance mutations, previously known to have negative consequences for viral replication, do not appear to affect the transmission efficiency of the virus. These include M184 and K65R. However, mutations that impose a significant fitness cost revert in the absence of ARV drug pressure, followed by outgrowth of fitter viral strains or wild type quasispecies.

Below presents a summary of mutations that were explored in this study:

The reversion of Y188C in participant three was due to its fitness cost, relative to the other mutations with which the mutation was transmitted. Fitter mutations, such as K103N, Y181C and G190A, followed the previously reported and well established NNRTI fitness hierarchy, and out-competed the Y188C virus until it was below the limit of detection of this study.

M184V, identified in participant 15, also reverted due to the mutation's significant fitness cost; consistent with previous reports (Harrigan et al., 2000). M46L was found to persist in participant two as a single mutation and reverted in participant 21 as a combination mutation. The persistence/reversion of M46L is therefore dependent upon whether the mutation is transmitted as a single, or as a combination mutation. The reversion of V108I and D67N has not been reported extensively in previously literature. Further work may be performed to investigate the effects that the accessory mutation V108I may have on the persistence/reversion when in combination with other mutations. Future research could also explore how the mutation may influence the viral fitness and the persistence of the combination strain, and of the V108I mutation itself. D67N has

been reported to persist and not affect replicative capacity (García-Lerma et al., 2004; Harrison et al., 2010). Future studies may investigate the replicative capacity of the viruses from participant seven.

Overall, six of the 43 (14%) total TDRMs identified in the 23 participants reverted (below 1% of the sequenced viral population). Of the 43 TDRMs identified, six were not identified by conventional population-based Sanger sequencing, and four of the six persisted until end follow-up. These findings highlight the importance of detecting minority variants that are missed by conventional population-based Sanger sequencing, as missed mutations may have significant clinical consequences, such as treatment failure.

In 21 of the 23 (91%) participants studied, at least one TDRM mutation persisted throughout the follow up period. Findings confirming observations from previous studies that TDRMs, particularly NNRTI mutations such as K103N, generally persist after transmission to a new host (Geretti et al., 2009; Nishizawa et al., 2015).

A rise in transmitted HIV-1 drug resistance is currently observed in developing Sub-Saharan countries; namely, Kenya, Uganda, Rwanda, Zambia and South Africa (Hofstra, Schmit, & Wensing, 2017). These countries constitute the participant cohort of this study (Price et al., 2011). The increase in the rate of DRM transmission may be attributed to increased ART rollout, the lack of treatment adherence, and the absence of resistance testing facilities and sub-optimal treatment strategies (Hofstra et al., 2017). Rates of transmission can be classified into three main categories, according to the WHO. These include the following: low rate (<5%), intermediate rate (5-15%), and high rate (>15%). The WHO recommends that surveillance studies are performed on a regular basis in order to assess the rate of DRM transmission. The most recent reports show that the prevalence of transmitted HIV-1 drug resistance for the countries explored in this study are as follows: Kenya: 14% (intermediate) (Eneyew, Seifu, Amogne, & Menon, 2016); Uganda: 10.8% (intermediate) (Hamers et al., 2011); Rwanda: 7.7% (intermediate) (Bennett et al., 2009); Zambia: 4.4% (low) (Bennett et al., 2009; V. A. Johnson et al., 2009); and South Africa: 9% (intermediate) (Steegen et al., 2016). Recent studies have, however, reported that countries that have made ART available to HIV-1 infected individuals for >5 years are significantly more likely to observe an increase in drug resistance transmission (Stadeli & Richman, 2013). In addition to this, countries

observed in the IAVI protocol C are expected to soon increase to high rates of resistance transmission (Price et al., 2011).

Results from this study highlight the importance of understanding the population dynamics of TDR HIV-1 in the absence of drug pressure for clinical management. The intermediate rates of drug resistance transmission for countries of the IAVI protocol C cohort, used in this study, are soon to increase drastically. Due to the multitude of factors that play a role in exacerbating transmission rates, it is critical importance that drug resistance transmission are well understood, and effective and insightful public health strategies are successfully implemented such that. The results from this study could be used to modify ART regimens so as to include more affordable drugs, with much higher barriers to resistance such as dolutegravir (DTG).

5. References

Abdool Karim, S. S. (2015). Overcoming impediments to global implementation of early antiretroviral therapy. *373*(9), 875-876.

Abrahams, M.-R., Anderson, J., Giorgi, E., Seoighe, C., Mlisana, K., Ping, L.-H., . . . Wood, N. (2009). Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *Journal of virology, 83*(8), 3556-3567.

Appadurai, R., & Senapati, S. (2016). Dynamical Network of HIV-1 Protease Mutants Reveals the Mechanism of Drug Resistance and Unhindered Activity. *Biochemistry, 55*(10), 1529-1540.

Arion, D., Sluis-Cremer, N., & Parniak, M. A. (2000). Mechanism by which phosphonoformic acid resistance mutations restore 3′-azido-3′-deoxythymidine (AZT) sensitivity to AZT-resistant HIV-1 reverse transcriptase. *Journal of Biological Chemistry, 275*(13), 9251-9255.

Azijn, H., Tirry, I., Vingerhoets, J., de Béthune, M.-P., Kraus, G., Boven, K., . . . Rimsky, L. T. (2010). TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. *Antimicrobial agents and chemotherapy, 54*(2), 718-727.

Bacheler, L. T., Anton, E. D., Kudish, P., Baker, D., Bunville, J., Krakowski, K., . . . Ellis, D. (2000). Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. *Antimicrobial agents and chemotherapy, 44*(9), 2475-2484.

Back, N., Nijhuis, M., Keulen, W., Boucher, C., Essink, B. O., Van Kuilenburg, A., . . . Berkhout, B. (1996). Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme. *The EMBO journal, 15*(15), 4040.

Balzer, S., Malde, K., Lanzén, A., Sharma, A., & Jonassen, I. (2010). Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim. *Bioinformatics, 26*(18), i420-i425.

Bangsberg, D. R., Charlebois, E. D., Grant, R. M., Holodniy, M., Deeks, S. G., Perry, S., . . . Zolopa, A. (2003). High levels of adherence do not prevent accumulation of HIV drug resistance mutations. *Aids, 17*(13), 1925-1932.

Barber, T. J., Harrison, L., Asboe, D., Williams, I., Kirk, S., Gilson, R., . . . Dunn, D. (2012). Frequency and patterns of protease gene resistance mutations in HIV-infected patients treated with lopinavir/ritonavir as their first protease inhibitor. *Journal of antimicrobial chemotherapy*, dkr569.

Barbour, J. D., Hecht, F. M., Wrin, T., Liegler, T. J., Ramstead, C. A., Busch, M. P., . . . Grant, R. M. (2004). Persistence of primary drug resistance among recently HIV-1 infected adults. *Aids, 18*(12), 1683-1689.

Baxter, J., Dunn, D., White, E., Sharma, S., Geretti, A., Kozal, M., . . . Lundgren, J. (2015). Global HIV‐1 transmitted drug resistance in the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial. *HIV medicine, 16*(S1), 77-87.

Bennett, D. E., Camacho, R. J., Otelea, D., Kuritzkes, D. R., Fleury, H., Kiuchi, M., . . . Schapiro, J. M. (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PloS one, 4*(3), e4724.

Blattner, W., Gallo, R. C., & Temin, H. (1988). Hiv causes AIDS. *science, 241*(4865), 515-516.

Blower, S., Aschenbach, A., Gershengorn, H. B., & Kahn, J. (2001). Predicting the unpredictable: transmission of drug-resistant HIV. *Nature medicine, 7*(9), 1016-1020.

Brenner, B. G., Routy, J.-P., Petrella, M., Moisi, D., Oliveira, M., Detorio, M., . . . Lalonde, R. (2002). Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *Journal of virology, 76*(4), 1753-1761.

Brik, A., & Wong, C.-H. (2003). HIV-1 protease: mechanism and drug discovery. *Organic & biomolecular chemistry, 1*(1), 5-14.

Cann, A. J., & Karn, J. (1989). Molecular biology of HIV: new insights into the virus life-cycle. *Aids, 3*(1), S19-34.

Casbon, J. A., Osborne, R. J., Brenner, S., & Lichtenstein, C. P. (2011). A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic acids research, 39*(12), e81-e81.

Castro, H., Pillay, D., Cane, P., Asboe, D., Cambiano, V., Phillips, A., . . . Chadwick, D. (2013). Persistence of HIV-1 transmitted drug resistance mutations. *Journal of Infectious Diseases, 208*(9), 1459-1463.

Centers for Disease Control and Prevention. (2016). Pre-exposure prohylaxis (PreP). *https://www.cdc.gov/hiv/risk/prep/, [Accessed 05.12.2016]*.

Chiu, T. K., & Davies, D. R. (2004). Structure and function of HIV-1 integrase. *Current topics in medicinal chemistry, 4*(9), 965-977.

Chun, T.-W., Stuyver, L., Mizell, S. B., Ehler, L. A., Mican, J. A. M., Baseler, M., . . . Fauci, A. S. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences, 94*(24), 13193-13197.

Church, J. D., Jones, D., Flys, T., Hoover, D., Marlowe, N., Chen, S., . . . Jackson, J. B. (2006). Sensitivity of the ViroSeq HIV-1 genotyping system for detection of the K103N resistance mutation in HIV-1 subtypes A, C, and D. *The Journal of Molecular Diagnostics, 8*(4), 430-432.

Clavel, F., & Hance, A. J. (2004). HIV drug resistance. *New England Journal of Medicine, 350*(10), 1023-1035.

Coffin, J. M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *science, 267*(5197), 483-489.

Cohen, M. S. (1998). Sexually transmitted diseases enhance HIV transmission: no longer a hypothesis. *The Lancet, 351*, S5-S7.

Cohen, S. E., Vittinghoff, E., Bacon, O., Doblecki-Lewis, S., Postle, B. S., Feaster, D. J., . . . Estrada, Y. (2015). High interest in preexposure prophylaxis among men who have sex with men at risk for HIV infection: baseline data from the US PrEP Demonstration Project. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 68*(4), 439-448.

Collins, J. A., Thompson, M. G., Paintsil, E., Ricketts, M., Gedzior, J., & Alexander, L. (2004). Competitive fitness of nevirapine-resistant human immunodeficiency virus type 1 mutants. *Journal of virology, 78*(2), 603-611.

Condra, J. H., Holder, D. J., Schleif, W. A., Blahy, O. M., Danovich, R. M., Gabryelski, L. J., . . . Rhodes, A. (1996). Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. *Journal of virology, 70*(12), 8270-8276.

Cong, M.-e., Heneine, W., & García-Lerma, J. G. (2007). The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. *Journal of virology, 81*(6), 3037-3041.

Conlon, C., Klenerman, P., Edwards, A., Larder, B., & Phillips, R. (1994). Heterosexual transmission of human immunodeficiency virus type 1 variants associated with zidovudine resistance. *Journal of Infectious Diseases, 169*(2), 411-415.

Conway, B., Wainberg, M. A., Hall, D., Harris, M., Reiss, P., Cooper, D., . . . Lange, J. M. (2001). Development of drug resistance in patients receiving combinations of zidovudine, didanosine and nevirapine. *Aids, 15*(10), 1269-1274.

Costin, J. M. (2007). Cytopathic mechanisms of HIV-1. *Virol J, 4*(100), 1-22.

Craigie, R. (2001). HIV integrase, a brief overview from chemistry to therapeutics. *Journal of Biological Chemistry, 276*(26), 23213-23216.

De Kok, B., Widdicombe, S., Pilnick, A., & Laurier, E. (2018). Doing patient-centredness versus achieving public health targets: A critical review of interactional dilemmas in ART adherence support. *Social Science & Medicine*.

De Luca, A. (2006). The impact of resistance on viral fitness and its clinical implications.

De Ronde, A., van Dooren, M., van der Hoek, L., Bouwhuis, D., de Rooij, E., van Gemen, B., . . . Goudsmit, J. (2001). Establishment of new transmissible and drug-sensitive human immunodeficiency virus type 1 wild types due to transmission of nucleoside analogue-resistant virus. *Journal of virology, 75*(2), 595-602.

Deacon, N., Tsykin, A., Solomon, A., Smith, K., Ludford-Menting, M., Hooker, D., . . . Chatfield, C. (1995). Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *science, 270*(5238), 988-991.

DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic acids research, 23*(22), 4742.

Deeks, S. G., Smith, M., Holodniy, M., & Kahn, J. O. (1997). HIV-1 protease inhibitors: a review for clinicians. *Jama, 277*(2), 145-153.

Delvaux, T., Elul, B., Ndagije, F., Munyana, E., Roberfroid, D., & Asiimwe, A. (2009). Determinants of nonadherence to a single-dose nevirapine regimen for the prevention of mother-to-child HIV transmission in Rwanda. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 50*(2), 223-230.

Department of Health. (2015a). National Consolidated Guidelines. *For the prevention of mother-to-child transmission of hiv (PMTCT) and the management of HIV in children, adolescents and adults*.

Department of Health. (2015b). Standard Treatment Guidelines and Essential Medicines List for South Africa. *Primary Health Care Level 2015 Edition*.

Derdeyn, C. A., Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. A., . . . Hahn, B. H. (2004). Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *science, 303*(5666), 2019-2022.

Deval, J., White, K. L., Miller, M. D., Parkin, N. T., Courcambeck, J., Halfon, P., . . . Canard, B. (2004). Mechanistic basis for reduced viral and enzymatic fitness of HIV-1 reverse transcriptase containing both K65R and M184V mutations. *Journal of Biological Chemistry, 279*(1), 509-516.

Dinoso, J., Kim, S., Wiegand, A., Palmer, S., Gange, S., Cranmer, L., . . . Brennan, T. (2009). Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences, 106*(23), 9403-9408.

Douek, D. C., Brenchley, J. M., Betts, M. R., Ambrozak, D. R., Hill, B. J., Okamoto, Y., . . . Wolinsky, S. (2002). HIV preferentially infects HIV-specific CD4+ T cells. *Nature, 417*(6884), 95-98.

Dykes, C., Fox, K., Lloyd, A., Chiulli, M., Morse, E., & Demeter, L. M. (2001). Impact of clinical reverse transcriptase sequences on the replication capacity of HIV-1 drug-resistant mutants. *Virology, 285*(2), 193-203.

Eneyew, K., Seifu, D., Amogne, W., & Menon, M. (2016). Assessment of Renal Function among HIV-Infected Patients on Combination Antiretroviral Therapy at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia. *Technology and Investment, 7*(03), 107.

Eshleman, S. H., Guay, L. A., Mwatha, A., Brown, E. R., Cunningham, S. P., Musoke, P., . . . Jackson, J. B. (2004). Characterization of nevirapine resistance mutations in women with subtype A vs. D HIV-1 6–8 weeks after single-dose nevirapine (HIVNET 012). *JAIDS Journal of Acquired Immune Deficiency Syndromes, 35*(2), 126-130.

Eshleman, S. H., Mracna, M., Guay, L. A., Deseyve, M., Cunningham, S., Mirochnick, M., . . . Mofenson, L. M. (2001). Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent HIV-1 vertical transmission (HIVNET 012). *Aids, 15*(15), 1951-1957.

Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., . . . Hamill, T. (2000). HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proceedings of the National Academy of Sciences, 97*(21), 11244-11249.

Fauci, A. S. (1988). The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science, 239*(4840), 617-622.

Fink, S. L., & Cookson, B. T. (2005). Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells. *Infection and immunity, 73*(4), 1907-1916.

Fletcher, C. V., Staskus, K., Wietgrefe, S. W., Rothenberger, M., Reilly, C., Chipman, J. G., . . . Schmidt, T. E. (2014). Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proceedings of the National Academy of Sciences, 111*(6), 2307-2312.

Flys, T., Nissley, D. V., Claasen, C. W., Jones, D., Shi, C., Guay, L. A., . . . Jackson, J. B. (2005). Sensitive drug-resistance assays reveal long-term persistence of HIV-1 variants with the K103N nevirapine (NVP) resistance mutation in some women and infants after the administration of single-dose NVP: HIVNET 012. *Journal of Infectious Diseases, 192*(1), 24-29.

Fonager, J., Larsson, J. T., Hussing, C., Engsig, F. N., Nielsen, C., & Fischer, T. K. (2015). Identification of minority resistance mutations in the HIV-1 integrase coding region using next generation sequencing. *Journal of Clinical Virology, 73*, 95-100.

Frentz, D., Boucher, C., & Van De Vijver, D. (2012). Temporal changes in the epidemiology of transmission of drug-resistant HIV-1 across the world. *AIDS Rev, 14*(1), 17-27.

Gandhi, R. T., Wurcel, A., Rosenberg, E. S., Johnston, M. N., Hellmann, N., Bates, M., . . . Walker, B. D. (2003). Progressive reversion of human immunodeficiency virus type 1 resistance mutations in vivo after transmission of a multiply drug-resistant virus. *Clinical infectious diseases, 37*(12), 1693-1698.

García-Lerma, J. G., MacInnes, H., Bennett, D., Reid, P., Nidtha, S., Weinstock, H., . . . Heneine, W. (2003). A novel genetic pathway of human immunodeficiency virus type 1 resistance to stavudine mediated by the K65R mutation. *Journal of virology, 77*(10), 5685-5693.

García-Lerma, J. G., MacInnes, H., Bennett, D., Weinstock, H., & Heneine, W. (2004). Transmitted human immunodeficiency virus type 1 carrying the D67N or K219Q/E mutation evolves rapidly to zidovudine resistance in vitro and shows a high replicative fitness in the presence of zidovudine. *Journal of virology, 78*(14), 7545-7552.

García-Lerma, J. G., Nidtha, S., Blumoff, K., Weinstock, H., & Heneine, W. (2001). Increased ability for selection of zidovudine resistance in a distinct class of wild-type HIV-1 from drug-naive persons. *Proceedings of the National Academy of Sciences, 98*(24), 13907-13912.

Garriga, C., Pérez‐Elías, M. J., Delgado, R., Ruiz, L., Nájera, R., Pumarola, T., . . . Menéndez‐Arias, L. (2007). Mutational patterns and correlated amino acid substitutions in the HIV‐1 protease after virological failure to nelfinavir‐and lopinavir/ritonavir‐based treatments. *Journal of medical virology, 79*(11), 1617-1628.

Geretti, A. M. (2006). *The significance of minority drug-resistant quasispecies*. Chapter 15: London: Mediscript.

Geretti, A. M., Fox, Z. V., Booth, C. L., Smith, C. J., Phillips, A. N., Johnson, M., . . . Johnson, J. A. (2009). Low-frequency K103N strengthens the impact of transmitted drug resistance on virologic responses to first-line efavirenz or nevirapine-based highly active antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 52*(5), 569-573.

Gerondelis, P., Archer, R. H., Palaniappan, C., Reichman, R. C., Fay, P. J., Bambara, R. A., & Demeter, L. M. (1999). The P236L delavirdine-resistant human immunodeficiency virus type 1 mutant is replication defective and demonstrates alterations in both RNA 5′ -end-and DNA 3′ -end-directed RNase H activities. *Journal of virology, 73*(7), 5803-5813.

Goudsmit, J., De Ronde, A., Ho, D. D., & Perelson, A. S. (1996). Human immunodeficiency virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. *Journal of virology, 70*(8), 5662-5664.

Grant, R. M., Lama, J. R., Anderson, P. L., McMahan, V., Liu, A. Y., Vargas, L., . . . Ramirez-Cardich, M. E. (2010). Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *New England Journal of Medicine, 363*(27), 2587-2599.

Gu, Z., Wainberg, M. A., Nguyen-Ba, N., L'Heureux, L., de Muys, J.-M., Bowlin, T. L., & Rando, R. F. (1999). Mechanism of action and in vitro activity of 1′, 3′-dioxolanylpurine nucleoside analogues against sensitive and drug-resistant human immunodeficiency virus type 1 variants. *Antimicrobial agents and chemotherapy, 43*(10), 2376-2382.

Guay, L. A., Musoke, P., Fleming, T., Bagenda, D., Allen, M., Nakabiito, C., . . . Deseyve, M. (1999). Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *The Lancet, 354*(9181), 795-802.

Gulick, R. M., Ribaudo, H. J., Shikuma, C. M., Lustgarten, S., Squires, K. E., Meyer III, W. A., . . . Murphy, R. L. (2004). Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. *New England Journal of Medicine, 350*(18), 1850-1861.

Haaland, R. E., Hawkins, P. A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., . . . Shaw, G. M. (2009). Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog, 5*(1), e1000274.

Haddad, M., Napolitano, L., Paquet, A., Evans, M., Petropoulos, C., & Whitcomb, J. (2011). Impact of HIV-1 reverse transcriptase E138 mutations on rilpivirine drug susceptibility. *Antivir Ther, 16*(Suppl 1), A18.

Haddad, M., Stawiski, E., Benhamida, J., & Coakley, E. (2010). *Improved genotypic algorithm for predicting etravirine susceptibility: Comprehensive list of mutations identified through correlation with matched phenotype.* Paper presented at the 17th Conference on Retroviruses and Opportunistic Infections, San Francisco.

Hamers, R. L., Wallis, C. L., Kityo, C., Siwale, M., Mandaliya, K., Conradie, F., . . . Sigaloff, K. C. (2011). HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study. *The Lancet infectious diseases, 11*(10), 750-759.

Harrigan, P. R., Stone, C., Griffin, P., Nájera, I., Bloor, S., Kemp, S., . . . Larder, B. (2000). Resistance profile of the human immunodeficiency virus type 1 reverse transcriptase inhibitor abacavir (1592U89) after monotherapy and combination therapy. *The Journal of infectious diseases, 181*(3), 912-920.

Harrison, L., Castro, H., Cane, P., Pillay, D., Booth, C., Phillips, A., . . . Study, U. C. H. C. (2010). The effect of transmitted HIV-1 drug resistance on pre-therapy viral load. *Aids, 24*(12), 1917-1922.

Hazuda, D. J. (2009). Resistance to Inhibitors of Human Immunodeficiency Virus Type I Integration *Antimicrobial Drug Resistance* (pp. 507-517): Springer.

Henderson, G. J., Lee, S.-K., Irlbeck, D. M., Harris, J., Kline, M., Pollom, E., . . . Swanstrom, R. (2012). Interplay between single resistance-associated mutations in the HIV-1 protease and viral infectivity, protease activity, and inhibitor sensitivity. *Antimicrobial agents and chemotherapy, 56*(2), 623-633.

Hofstra, L. M., Schmit, J.-C., & Wensing, A. M. (2017). Transmission of HIV-1 Drug Resistance. *Handbook of Antimicrobial Resistance*, 455-478.

Huang, W., Gamarnik, A., Limoli, K., Petropoulos, C. J., & Whitcomb, J. M. (2003). Amino acid substitutions at position 190 of human immunodeficiency virus type 1 reverse transcriptase increase susceptibility to delavirdine and impair virus replication. *Journal of virology, 77*(2), 1512-1523.

Hunt, G. M., Coovadia, A., Abrams, E. J., Sherman, G., Meyers, T., Morris, L., & Kuhn, L. (2011). HIV-1 drug resistance at antiretroviral treatment initiation in children previously exposed to single-dose nevirapine. *AIDS (London, England), 25*(12), 1461.

Iglesias-Ussel, M. D., Casado, C., Yuste, E. s., Olivares, I., & López-Galíndez, C. (2002). In vitro analysis of human immunodeficiency virus type 1 resistance to nevirapine and fitness determination of resistant variants. *Journal of general virology, 83*(1), 93-101.

Illumina. (2015). NextSeq System denature and dilute libraries guide*., https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-denature-dilute-libraries-guide-15048776-04.pdf*, [Accessed 03.06.2015].

Illumina. (2016a). Indexed Sequencing (Overview Guide). *https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/indexed-sequencing-overview-guide-15057455-03.pdf, [Accessed 08.12.2017]*.

Illumina. (2016b). Sequencing by Synthesis Technology. *https://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.html, [Accessed 20.12.2016]*.

Imrie, A., Beveridge, A., Genn, W., Vizzard, J., & Cooper, D. A. (1997). Transmission of human immunodeficiency virus type 1 resistant to nevirapine and zidovudine. *Journal of Infectious Diseases, 175*(6), 1502-1506.

Jochmans, D. (2008). Novel HIV-1 reverse transcriptase inhibitors. *Virus research, 134*(1), 171-185.

Johnson, A. A., Marchand, C., & Pommier, Y. (2004). HIV-1 integrase inhibitors: a decade of research and two drugs in clinical trial. *Current topics in medicinal chemistry, 4*(10), 1059-1077.

Johnson, V. A., Brun-Vézinet, F., Clotet, B., Gunthard, H., Kuritzkes, D. R., Pillay, D., . . . Richman, D. D. (2009). Update of the drug resistance mutations in HIV-1: December 2009. *Topics in HIV Medicine, 17*(5), 138-145.

Katzenstein, D. A., Holodniy, M., Israelski, D. M., Sengupta, S., Mole, L. A., Bubp, J. L., & Merigan, T. C. (1992). Plasma viremia in human immunodeficiency virus infection: relationship to stage of disease and antiviral treatment. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 5*(2), 107-112.

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., . . . Li, H. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences, 105*(21), 7552-7557.

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W., & Vogelstein, B. (2011). Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences, 108*(23), 9530-9535.

Larder, B., Kohli, A., Kellam, P., Kemp, S., Kronick, M., & Henfrey, R. (1993). Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature, 365*(6447), 671-673.

Larder, B. A., Kemp, S. D., & Harrigan, P. R. (1995). Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science*, 696-696.

Lauring, A. S., & Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog, 6*(7), e1001005.

Ledergerber, C., & Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Briefings in bioinformatics*, bbq077.

Li, J. Z., & Kuritzkes, D. R. (2013). Clinical implications of HIV-1 minority variants. *Clinical infectious diseases, 56*(11), 1667-1674.

Liang, J., Mesplède, T., Oliveira, M., Anstett, K., & Wainberg, M. A. (2015). The combination of the R263K and T66I resistance substitutions in HIV-1 integrase is incompatible with high-level viral replication and the development of high-level drug resistance. *Journal of virology, 89*(22), 11269-11274.

Little, S. J., Frost, S. D., Wong, J. K., Smith, D. M., Pond, S. L. K., Ignacio, C. C., . . . Richman, D. D. (2008). Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *Journal of virology, 82*(11), 5510-5518.

Little, S. J., Holte, S., Routy, J.-P., Daar, E. S., Markowitz, M., Collier, A. C., . . . Conway, B. (2002). Antiretroviral-drug resistance among patients recently infected with HIV. *New England Journal of Medicine, 347*(6), 385-394.

Lyidogan, P., & Anderson, K. S. (2014). Current perspectives on HIV-1 antiretroviral drug resistance. *Viruses, 6*(10), 4095-4139.

Maldarelli, F., Kearney, M., Palmer, S., Stephens, R., Mican, J., Polis, M. A., . . . Rock-Kress, D. (2013). HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *Journal of virology, 87*(18), 10313-10323.

Mansky, L. M., & Temin, H. M. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *Journal of virology, 69*(8), 5087-5094.

Marcelin, A.-G. (2006). Resistance to nucleoside reverse transcriptase inhibitors. *Chapter 1, London: Mediscript*.

Margot, N., Lu, B., Cheng, A., & Miller, M. (2006). Resistance development over 144 weeks in treatment‐naive patients receiving tenofovir disoproxil fumarate or stavudine with lamivudine and efavirenz in Study 903. *HIV medicine, 7*(7), 442-450.

Martinson, N. A., Morris, L., Gray, G., Moodley, D., Pillay, V., Cohen, S., . . . Steyn, J. (2007). Selection and persistence of viral resistance in HIV-infected children after exposure to single-dose nevirapine. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 44*(2), 148-153.

Masquelier, B., Descamps, D., Carriere, I., Ferchal, F., Collin, G., Denayrolles, M., . . . Buffet-Janvresse, C. (1999). Zidovudine resensitization and dual HIV-1 resistance to zidovudine and lamivudine in the delta lamivudine roll-over study. *Antiviral therapy, 4*, 69-78.

McCallum, M., Oliveira, M., Ibanescu, R.-I., Kramer, V. G., Moisi, D., Asahchop, E. L., . . . Wainberg, M. A. (2013). Basis for early and preferential selection of the E138K mutation in HIV-1 reverse transcriptase. *Antimicrobial agents and chemotherapy, 57*(10), 4681-4688.

McDonald, C. K., & Kuritzkes, D. R. (1997). Human immunodeficiency virus type 1 protease inhibitors. *Archives of Internal Medicine, 157*(9), 951-959.

Menéndez-Arias, L. (2011). A structural frame for understanding the role of thymidine analogue resistance mutations in resistance to zidovudine and other nucleoside analogues. *Antiviral therapy, 16*, 943-946.

Metzner, K. J., Rauch, P., Braun, P., Knechten, H., Ehret, R., Korn, K., . . . van Lunzen, J. (2011). Prevalence of key resistance mutations K65R, K103N, and M184V as minority HIV-1 variants in chronically HIV-1 infected, treatment-naive patients. *Journal of Clinical Virology, 50*(2), 156-161.

Meyer, P. R., Matsuura, S. E., So, A. G., & Scott, W. A. (1998). Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proceedings of the National Academy of Sciences, 95*(23), 13471-13476.

Miller, S. (2002). HIV life cycle and potential targets for drug activity. *Southern African Journal of HIV Medicine, 3*(2).

Miller, V., Ait-Khaled, M., Stone, C., Griffin, P., Mesogiti, D., Cutrell, A., . . . Pearce, G. (2000). HIV-1 reverse transcriptase (RT) genotype and susceptibility to RT inhibitors during abacavir monotherapy and combination therapy. *Aids, 14*(2), 163-171.

Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A., & Coffman, R. L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *The Journal of immunology, 136*(7), 2348-2357.

Nanfack, A. J., Redd, A. D., Bimela, J. S., Ncham, G., Achem, E., Banin, A. N., . . . Meli, J. (2017). Multimethod Longitudinal HIV Drug Resistance Analysis in ART Naïve Patients. *Journal of clinical microbiology*, JCM. 00634-00617.

Ndembi, N., Abraha, A., Pilch, H., Ichimura, H., Mbanya, D., Kaptue, L., . . . Arts, E. J. (2008). Molecular characterization of human immunodeficiency virus type 1 (HIV-1) and HIV-2 in Yaounde, Cameroon: evidence of major drug resistance mutations in newly diagnosed patients infected with subtypes other than subtype B. *Journal of clinical microbiology, 46*(1), 177-184.

Nishizawa, M., Matsuda, M., Hattori, J., Shiino, T., Matano, T., Heneine, W., . . . Sugiura, W. (2015). Longitudinal detection and persistence of minority drug-resistant populations and their effect on salvage therapy. *PloS one, 10*(9), e0135941.

Oyugi, J. H., Byakika-Tusiime, J., Charlebois, E. D., Kityo, C., Mugerwa, R., Mugyenyi, P., & Bangsberg, D. R. (2004). Multiple validated measures of adherence indicate high levels of adherence to generic HIV antiretroviral therapy in a resource-limited setting. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 36*(5), 1100-1102.

Palella Jr, F. J., Baker, R. K., Moorman, A. C., Chmiel, J. S., Wood, K. C., Brooks, J. T., . . . Investigators, H. O. S. (2006). Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 43*(1), 27-34.

Palella Jr, F. J., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., . . . Investigators, H. O. S. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine, 338*(13), 853-860.

Palmer, S., Boltz, V., Maldarelli, F., Kearney, M., Halvas, E. K., Rock, D., . . . Metcalf, J. A. (2006). Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant HIV-1 in patients starting and stopping non-nucleoside therapy. *Aids, 20*(5), 701-710.

Palmer, S., Kearney, M., Maldarelli, F., Halvas, E. K., Bixby, C. J., Bazmi, H., . . . Dewar, R. L. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *Journal of clinical microbiology, 43*(1), 406-413.

Pao, D., Andrady, U., Clarke, J., Dean, G., Drake, S., Fisher, M., . . . Tang, A. (2004). Long-term persistence of primary genotypic resistance after HIV-1 seroconversion. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 37*(5), 1570-1573.

Parkin, N. T., Gupta, S., Chappey, C., & Petropoulos, C. J. (2006). The K101P and K103R/V179D mutations in human immunodeficiency virus type 1 reverse transcriptase confer resistance to nonnucleoside reverse transcriptase inhibitors. *Antimicrobial agents and chemotherapy, 50*(1), 351-354.

Perelson, A. S., Kirschner, D. E., & De Boer, R. (1993). Dynamics of HIV infection of CD4+ T cells. *Mathematical biosciences, 114*(1), 81-125.

Petropoulos, C. J., Parkin, N. T., Limoli, K. L., Lie, Y. S., Wrin, T., Huang, W., . . . Capon, D. J. (2000). A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrobial agents and chemotherapy, 44*(4), 920-928.

Pingen, M., Sarrami-Forooshani, R., Wensing, A. M., van Ham, P., Drewniak, A., Boucher, C. A., . . . Nijhuis, M. (2014). Diminished transmission of drug resistant HIV-1 variants with reduced replication capacity in a human transmission model. *Retrovirology, 11*(1), 113-113.

Pingen, M., Wensing, A. M., Fransen, K., De Bel, A., de Jong, D., Hoepelman, A. I., . . . Poljak, M. (2014). Persistence of frequently transmitted drug-resistant HIV-1 variants can be explained by high viral replication capacity. *Retrovirology, 11*(1), 105.

Pommier, Y., Johnson, A. A., & Marchand, C. (2005). Integrase inhibitors to treat HIV/AIDS. *Nature Reviews Drug Discovery, 4*(3), 236-248.

Porter, D. P., Toma, J., Tan, Y., Solberg, O., Cai, S., Kulkarni, R., . . . Palella, F. (2016). Clinical outcomes of virologically-suppressed patients with pre-existing HIV-1 drug resistance mutations switching to rilpivirine/emtricitabine/tenofovir disoproxil fumarate in the SPIRIT study. *HIV clinical trials, 17*(1), 29-37.

Poss, M., Martin, H. L., Kreiss, J. K., Granville, L., Chohan, B., Nyange, P., . . . Overbaugh, J. (1995). Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *Journal of virology, 69*(12), 8118-8122.

Price, M. A., Wallis, C. L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., . . . Hunter, E. (2011). Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. *AIDS Research and Human Retroviruses, 27*(1), 5-12.

Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast Cancer Res, 11*(Suppl 3), S12.

Ren, J., Nichols, C., Bird, L., Chamberlain, P., Weaver, K., Short, S., . . . Stammers, D. (2001). Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *Journal of molecular biology, 312*(4), 795-805.

Ren, J., & Stammers, D. K. (2008). Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus research, 134*(1), 157-170.

Reuman, E. C., Rhee, S.-Y., Holmes, S. P., & Shafer, R. W. (2010). Constrained patterns of covariation and clustering of HIV-1 non-nucleoside reverse transcriptase inhibitor resistance mutations. *Journal of antimicrobial chemotherapy, 65*(7), 1477-1485.

Rhee, S.-Y., Liu, T., Ravela, J., Gonzales, M. J., & Shafer, R. W. (2004). Distribution of human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns in 4,183 persons undergoing genotypic resistance testing. *Antimicrobial agents and chemotherapy, 48*(8), 3122-3126.

Richman, D., Shih, C.-K., Lowy, I., Rose, J., Prodanovich, P., Goff, S., & Griffin, J. (1991). Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proceedings of the National Academy of Sciences, 88*(24), 11241-11245.

Rimsky, L., Vingerhoets, J., Van Eygen, V., Eron, J., Clotet, B., Hoogstoel, A., . . . Picchio, G. (2012). Genotypic and phenotypic characterization of HIV-1 isolates obtained from patients on rilpivirine therapy experiencing virologic failure in the phase 3 ECHO and THRIVE studies: 48-week analysis. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 59*(1), 39-46.

Robinson, L. H., Myers, R. E., Snowden, B. W., Tisdale, M., & Blair, E. D. (2000). HIV type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays. *AIDS Research and Human Retroviruses, 16*(12), 1149-1156.

Rocheleau, G., Brumme, C. J., Shoveller, J., Lima, V., & Harrigan, P. (2017). Longitudinal trends of HIV drug resistance in a large Canadian cohort, 1996–2016. *Clinical Microbiology and Infection*.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., . . . Edwards, M. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature, 475*(7356), 348-352.

Rothberg, J. M., & Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nature biotechnology, 26*(10), 1117-1124.

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology, 94*(3), 441-448.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*(12), 5463-5467.

Schock, H. B., Garsky, V. M., & Kuo, L. C. (1996). Mutational Anatomy of an HIV-1 Protease Variant Conferring Cross-resistance to Protease Inhibitors in Clinical Trials Compensatory Modulations of Binding And Activity. *Journal of Biological Chemistry, 271*(50), 31957-31963.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature, 200*(8), 16-18.

Selmi, B., Deval, J., Alvarez, K., Boretto, J., Sarfati, S., Guerreiro, C., & Canard, B. (2003). The Y181C substitution in 3′ -azido-3′ -deoxythymidine-resistant human immunodeficiency virus, type 1, reverse transcriptase suppresses the ATP-mediated repair of the 3′ -azido-3′ -deoxythymidine 5′ -monophosphate-terminated primer. *Journal of Biological Chemistry, 278*(42), 40464-40472.

Shet, A., Berry, L., Mohri, H., Mehandru, S., Chung, C., Kim, A., . . . Boden, D. (2006). Tracking the prevalence of transmitted antiretroviral drug-resistant HIV-1: a decade of experience. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 41*(4), 439-446.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., . . . Hood, L. E. (1985). Fluorescence detection in automated DNA sequence analysis. *Nature, 321*(6071), 674-679.

Ssemwanga, D., Lihana, R. W., Ugoji, C., Abimiku, A. l., Nkengasong, J., Dakum, P., & Ndembi, N. (2014). Update on HIV-1 acquired and transmitted drug resistance in Africa. *AIDS reviews, 17*(1), 3-20.

Stadeli, K. M., & Richman, D. D. (2013). Rates of emergence of HIV drug resistance in resource-limited settings: a systematic review. *Antiviral therapy, 18*(1), 115.

Stanford University. (2017). HIV Drug Resistance Database. *https://hivdb.stanford.edu, [Accessed 10.12.17]*.

Statistics-SA. (2014). Mid-Year Population Estimates. *Statistical Release, P0302*.

Steegen, K., Carmona, S., Bronze, M., Papathanasopoulos, M. A., van Zyl, G., Goedhals, D., . . . Stevens, W. S. (2016). Moderate levels of pre-treatment HIV-1 antiretroviral drug resistance detected in the first South African national survey. *PloS one, 11*(12), e0166305.

Steinhauer, D. A., Domingo, E., & Holland, J. J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene, 122*(2), 281-288.

Tabb, Z. J., Mmbaga, B. T., Gandhi, M., Louie, A., Kuncze, K., Okochi, H., . . . Dow, D. E. (2018). Antiretroviral drug concentrations in hair are associated with virologic outcomes among young people living with HIV in Tanzania. *Aids, 32*(9), 1115-1123.

Tambuyzer, L., Nijs, S., Daems, B., Picchio, G., & Vingerhoets, J. (2011). Effect of mutations at position E138 in HIV-1 reverse transcriptase on phenotypic susceptibility and virologic response to etravirine. *JAIDS Journal of Acquired Immune Deficiency Syndromes, 58*(1), 18-22.

Temesgen, Z., & Siraj, D. S. (2008). Raltegravir: first in class HIV integrase inhibitor. *Therapeutics and clinical risk management, 4*(2), 493.

Tisdale, M., Alnadaf, T., & Cousens, D. (1997). Combination of mutations in human immunodeficiency virus type 1 reverse transcriptase required for resistance to the carbocyclic nucleoside 1592U89. *Antimicrobial agents and chemotherapy, 41*(5), 1094-1098.

Tzoupis, H., Leonis, G., Mavromoustakos, T., & Papadopoulos, M. G. (2013). A comparative molecular dynamics, MM–PBSA and thermodynamic integration study of saquinavir complexes with wild-type HIV-1 PR and L10I, G48V, L63P, A71V, G73S, V82A and I84V single mutants. *Journal of chemical theory and computation, 9*(3), 1754-1764.

UNAIDS. (2016a). Fact Sheet 2016: Global Statistics 2015. *http://www.unaids.org/en/resources/fact-sheet*.

UNAIDS. (2016b). UNAIDS 2016 [Fact sheet]. *AIDS by the numbers.*

Van Laethem, K., Theys, K., & Vandamme, A.-M. (2015). HIV-1 genotypic drug resistance testing: digging deep, reaching wide? *Current opinion in virology, 14*, 16-23.

Varghese, V., Shahriar, R., Rhee, S.-Y., Liu, T., Simen, B. B., Egholm, M., . . . Babrzadeh, F. (2009). Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase

inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. *Journal of acquired immune deficiency syndromes (1999), 52*(3), 309-315.
Vermeiren, H., Van Craenenbroeck, E., Alen, P., Bacheler, L., Picchio, G., Lecocq, P., & Team, T. V. C. R. C. (2007). Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *Journal of virological methods, 145*(1), 47-55.
Vingerhoets, J., Azijn, H., Fransen, E., De Baere, I., Smeulders, L., Jochmans, D., . . . de Béthune, M.-P. (2005). TMC125 displays a high genetic barrier to the development of resistance: evidence from in vitro selection experiments. *Journal of virology, 79*(20), 12773-12782.
Vingerhoets, J., Tambuyzer, L., Azijn, H., Hoogstoel, A., Nijs, S., Peeters, M., . . . Picchio, G. (2010). Resistance profile of etravirine: combined analysis of baseline genotypic and phenotypic data from the randomized, controlled Phase III clinical studies. *Aids, 24*(4), 503-514.
Wainberg, M. A., Miller, M. D., Quan, Y., Salomon, H., Mulato, A. S., Lamy, P. D., . . . Cherrington, J. M. (1999). In vitro selection and characterization of HIV-1 with reduced susceptibility to PMPA. *Antiviral therapy, 4*, 87-94.
Walker, C. M., Moody, D. J., Stites, D. P., & Levy, J. A. (1986). CD8+ lymphocytes can control HIV infection in vitro by suppressing virus replication. *science, 234*(4783), 1563-1566.
Wang, L. H., Begley, J., St. Claire III, R. L., Harris, J., Wakeford, C., & Rousseau, F. S. (2004). Pharmacokinetic and pharmacodynamic characteristics of emtricitabine support its once daily dosing for the treatment of HIV infection. *AIDS Research & Human Retroviruses, 20*(11), 1173-1182.
Wensing, A. M., Calvez, V., Günthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., . . . Richman, D. D. (2015). 2015 update of the drug resistance mutations in HIV-1. *Topics in Antiviral Medicine, 23*(4), 132-141.
Whitcomb, J. M., Parkin, N. T., Chappey, C., Hellmann, N. S., & Petropoulos, C. J. (2003). Broad nucleoside reverse-transcriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. *The Journal of infectious diseases, 188*(7), 992-1000.
White, K. L., Margot, N. A., Ly, J. K., Chen, J. M., Ray, A. S., Pavelko, M., . . . Miller, M. D. (2005). A combination of decreased NRTI incorporation and decreased excision determines the resistance profile of HIV-1 K65R RT. *AIDS, 19*(16), 1751-1760.
Winslow, D., Garber, S., Reid, C., Scarnati, H., Baker, D., Rayner, M., & Anton, E. (1996). Selection conditions affect the evolution of specific mutations in the reverse transcriptase gene associated with resistance to DMP 266. *AIDS, 10*(11), 1205-1209.
Wisniewski, M., Balakrishnan, M., Palaniappan, C., Fay, P. J., & Bambara, R. A. (2000). Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proceedings of the National Academy of Sciences, 97*(22), 11978-11983.
Wolfs, T. F., Zwart, G., Bakker, M., & Goudsmit, J. (1992). HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology, 189*(1), 103-110.
Wolinsky, S. M., & Wike, C. M. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *science, 255*(5048), 1134.
World Health Organization. (2013). Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. *[Accessed 20.03.2018]*.
World Health Organization. (2014). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations.
World Health Organization. (2015). Use of antiretrovirals for treatment and prevention of HIV infection. *http://www.who.int/, [Accessed 28.01.15]*.
World Health Organization. (2016a). Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. *Recommendations for a public health approach - Second edition*.

World Health Organization. (2016b). Consolidated Guidelines on The Use Of Antiretroviral Drugs For Treating And Preventing HIV Infection. Recommendations For a Public Health Approach. *Second Edition*.

Wu, H., Zhang, H.-J., Zhang, X.-m., Xu, H.-f., Wang, M., Huang, J.-d., & Zheng, B.-J. (2012). Identification of drug resistant mutations in HIV-1 CRF07_BC variants selected by nevirapine in vitro. *PloS one, 7*(9), e44333.

Yebra, G., de Mulder, M., Pérez-Elías, M. J., Pérez-Molina, J. A., Galán, J. C., Llenas-García, J., . . . Holguín, Á. (2011). Increase of transmitted drug resistance among HIV-infected sub-Saharan Africans residing in Spain in contrast to the native population. *PloS one, 6*(10), e26757.

Yerly, S., Rakik, A., De Loes, S. K., Hirschel, B., Descamps, D., Brun-Vézinet, F., & Perrin, L. (1998). Switch to unusual amino acids at codon 215 of the human immunodeficiency virus type 1 reverse transcriptase gene in seroconvertors infected with zidovudine-resistant variants. *Journal of virology, 72*(5), 3520-3523.

Yerly, S., von Wyl, V., Ledergerber, B., Böni, J., Schüpbach, J., Bürgisser, P., . . . Günthard, H. F. (2007). Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. *AIDS, 21*(16), 2223-2229.

Young, S. (2001). Inhibition of HIV-1 integrase by small molecules: the potential for a new class of AIDS chemotherapeutics. *Current opinion in drug discovery & development, 4*(4), 402-410.

Zagordi, O., Klein, R., Däumer, M., & Beerenwinkel, N. (2010). Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic acids research, 38*(21), 7400-7409.

Zhou, S., Jones, C., Mieczkowski, P., & Swanstrom, R. (2015). Primer ID validates template sampling depth and greatly reduces the error rate of next-generation sequencing of HIV-1 genomic RNA populations. *Journal of virology, 89*(16), 8540-8555.

Zhu, J., & Paul, W. E. (2008). CD4 T cells: fates, functions, and faults. *Blood, 112*(5), 1557-1569.

6. Appendices

Appendix A: Ethics Clearance



R14/49 Mr Dean Harris

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M160691

NAME: (Principal Investigator)	Mr Dean Harris
DEPARTMENT:	Molecular Medicine and Haematology University of the Witwatersrand Faculty of Health Sciences School of Pathology
PROJECT TITLE:	Population Dynamics in HIV-1 Transmitted Antiretroviral Drug Resistance
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-Study
SUPERVISOR:	Prof Maria Papathanasopoulos
APPROVED BY:	Professor P. Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	07/07/2016

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

Appendix B: DNA Work

1. GeneJet DNA (Thermo Fisher Scientific, Massachusetts, USA) purification kit protocol

1. Add a 1:1 volume of Binding Buffer to completed PCR mixture (e.g. for every 100 µL of reaction mixture, add 100 µL of Binding Buffer). Mix thoroughly. Check the color of the solution. A yellow color indicates an optimal pH for DNA binding. If the color of the solution is orange or violet, add 10 µL of 3 M sodium acetate, pH 5.2 solution and mix. The color of the mix will become yellow.

2. Optional: if the DNA fragment is ≤ 500 bp, add a 1:2 volume of 100% isopropanol (e.g., 100 µL of isopropanol should be added to 100 µL of PCR mixture combined with 100 µL of Binding Buffer). Mix thoroughly.

Note. If PCR mixture contains primer-dimers, purification without isopropanol is recommended. However, the yield of the target DNA fragment will be lower

3. Transfer up to 800 µL of the solution from step 1 (or optional step 2) to the GeneJET purification column. Centrifuge for 30-60 s. Discard the flow-through.

Notes. If the total volume exceeds 800 µL, the solution can be added to the column in stages. After the addition of 800 µL of solution, centrifuge the column for 30-60 s and discard flowthrough. Repeat until the entire solution has been added to the column membrane. Close the bag with GeneJET Purification Columns tightly after each use!

4. Add 700 µL of Wash Buffer (diluted with the ethanol as described on p. 3) to the GeneJET purification column. Centrifuge for 30-60 s. Discard the flow-through and place the purification column back into the collection tube.

5. Centrifuge the empty GeneJET purification column for an additional 1 min to completely remove any residual wash buffer.

Note. This step is essential as the presence of residual ethanol in the DNA sample may inhibit subsequent reactions.

6. Transfer the GeneJET purification column to a clean 1.5 mL microcentrifuge tube (not included). Add 50 µL of Elution Buffer to the center of the GeneJET purification column membrane and centrifuge for 1 min. Note

• For low DNA amounts the elution volumes can be reduced to increase DNA concentration. An elution volume between 20-50 µL does not significantly reduce the DNA yield. However, elution volumes less than 10 µL are not recommended.

• If DNA fragment is >10 kb, prewarm Elution Buffer to 65 °C before applying to column. • If the elution volume is 10 µL and DNA amount is ≥5 µg, incubate column for 1 min at room temperature before centrifugation.

7. Discard the GeneJET purification column and store the purified DNA at -20 °C

---

2. Agarose Gel Preparation

- 1% Agarose gel (200 ml) Add 2 g of agarose to 200 ml of 1 × TAE buffer.

- 50×TAE buffer (1 L) Weigh out 242 g tris base. Add 57.1 mL glacial acetic acid and 100 mL 0.5 EDTA. Add 800 mL of distilled water. Adjust pH to 8.0 with HCl. Top up to 1L with dH2O. Store at room temperature.

---

3. Qubit 2.0 Fluorometer High Sensitivity Assay

1.1 Set up the required number of 0.5-mL tubes for standards and samples. The Qubit dsDNA HS Assay requires 2 standards.

Note: Use only thin-wall, clear, 0.5-mL PCR tubes. Acceptable tubes include Qubit assay tubes (Cat. no. Q32856) or Axygen® PCR-05-C tubes (part no. 10011-830).

1.2 Label the tube lids.

Note: Do not label the side of the tube as this could interfere with the sample read. Label the lid of each standard tube correctly. Calibration of the Qubit® Fluorometer requires the standards to be inserted into the instrument in the right order.

1.3 Prepare the Qubit® working solution by diluting the Qubit dsDNA HS Reagent 1:200 in Qubit dsDNA HS Buffer. Use a clean plastic tube each time you prepare Qubit working solution. Do not mix the working solution in a glass container.

Note: The final volume in each tube must be 200 µL. Each standard tube requires 190 µL of Qubit working solution, and each sample tube requires anywhere from 180–199 µL. Prepare sufficient Qubit working solution to accommodate all standards and samples.

1.4 Add 190 µL of Qubit working solution to each of the tubes used for standards.

1.5 Add 10 µL of each Qubit standard to the appropriate tube, then mix by vortexing 2–3 seconds. Be careful not to create bubbles.

Note: Careful pipetting is critical to ensure that exactly 10 µL of each Qubit® standard is added to 190 µL of Qubit® working solution.

1.6 Add Qubit® working solution to individual assay tubes so that the final volume in each tube after adding sample is 200 µL.

Note: Your sample can be anywhere from 1–20 µL. Add a corresponding volume of Qubit® working solution to each assay tube: anywhere from 180–199 µL.

1.7 Add each sample to the assay tubes containing the correct volume of Qubit working solution, then mix by vortexing 2–3 seconds. The final volume in each tube should be 200 µL.

1.8 Allow all tubes to incubate at room temperature for 2 minutes.

Proceed to reading standards and samples

2. On the Home screen of the Qubit 3.0 Fluorometer, press DNA, then select dsDNA High Sensitivity as the assay type. The “Read standards” screen is displayed. Press Read Standards to proceed.

Note: If you have already performed a calibration for the selected assay, the instrument prompts you to choose between reading new standards and running samples using the previous calibration. If you want to use the previous calibration, skip to step 2.4. Otherwise, continue with step 2.2.

2.1 Insert the tube containing Standard #1 into the sample chamber, close the lid, then press Read standard. When the reading is complete (~3 seconds), remove Standard #1.

2.2 Insert the tube containing Standard #2 into the sample chamber, close the lid, then press. Read standard. When the reading is complete, remove Standard #2. The instrument displays the results on the Read standard screen.

2.3 Press Run samples.

2.4 On the assay screen, select the sample volume and units:

a. Press the + or – buttons on the wheel to select the sample volume added to the assay tube (from 1–20 μL).

b. From the dropdown menu, select the units for the output sample concentration.

2.5 Insert a sample tube into the sample chamber, close the lid, then press Read tube. When the reading is complete (~3 seconds), remove the sample tube. The instrument displays the results on the assay screen. The top value (in large font) is the concentration of the original sample. The bottom value is the dilution concentration.

2.6 Repeat step 2.6 until all samples have been read.

Appendix C: Nextera XT library preparation guide

1. Fragmentation of Input DNA

During this step, input DNA is tagmented (tagged and fragmented) by the Nextera XT transposome. The Nextera XT transposome simultaneously fragments the input DNA and adds adapter sequences to the ends, allowing amplification by PCR in subsequent steps.

Consumables:

Item	Quantity	Storage	Supplied
ATM (Amplicon Tagment Mix)	1 tube	-25°C to -15°C	Illumina
TD (Tagment DNA Buffer)	1 tube	-25°C to -15°C	Illumina
NT (Neutralize Tagment Buffer)	1 tube	Room Temperature	Illumina
Input DNA (0.2 ng/µl)	-	-25°C to -15°C	User
96-well hard shell TCY plate	1 plate	Room Temperature	User
Microseal 'B' adhesive film	-	Room Temperature	User

Preparation

1. Remove the ATM, TD, and input DNA from -25°C to -15°C storage and thaw on ice.
2. Visually inspect NT to make sure that there is no precipitate. If there is precipitate, vortex until all particulates are re-suspended.
3. After thawing, mix reagents by gently inverting the tubes 3–5 times, followed by a brief spin in a micro centrifuge.

Make NTA

Make sure that the reaction is assembled in the order described for optimal kit performance. You do not need to assemble the reaction on ice.

1. Label a new 96-well TCY plate NTA (Nextera XT Tagment Amplicon Plate).
2. Add **10 µl** TD Buffer to each well to be used in this assay. Change tips between samples.

Calculate the total volume of TD for all reactions, and divide the volume equally among the wells of a PCR eight-tube strip. Use a multichannel pipette to dispense into the NTA plate.

3. Add **5 µl** input DNA at 0.2 ng/µl (1 ng total) to each sample well of the NTA plate.

It is critical to use the full amount of input DNA.

4. Using a multichannel pipette, gently pipette up and down 5 times to mix. Change tips between samples.
5. Add **5 µl** ATM to the wells containing input DNA and TD Buffer. Change tips between samples.

Calculate the total volume of ATM for all reactions, and divide the volume equally among the wells of a PCR eight-tube strip. Use a multichannel pipette to dispense into the NTA plate.

6. Using a multichannel pipette, gently pipette up and down 5 times to mix. Change tips between samples.
7. Seal the NTA plate with a Microseal 'B' adhesive seal.
8. Centrifuge at **280 × g at 20°C for 1 minute.**
9. Place the NTA plate in a thermal cycler and run the following program:

Make sure that the thermal cycler lid is heated during the incubation.

• 55°C for 5 minutes

• Hold at 10°C

10. When the sample reaches **10°C**, proceed immediately to Neutralize NTA as the transposome is still active.

Neutralize NTA

Calculate the total volume of NT for all reactions, and divide the volume equally among the wells of a PCR eight-tube strip. Use a multichannel pipette to dispense into the NTA plate.

1. Carefully remove the Microseal 'B' seal and add **5 µl** NT Buffer to each well of the NTA plate. Change tips between samples.
2. Using a multichannel pipette, gently pipette up and down 5 times to mix. Change tips between samples.
3. Seal the NTA plate with a Microseal 'B' adhesive seal.
4. Centrifuge at **280 × g at 20°C for 1 minute**.
5. Place the NTA plate at room temperature for 5 minutes.

---

2. Indexing (PCR Amplification)

In this step, the tagmented DNA is amplified via a limited-cycle PCR program. The PCR step also adds index 1 (i7) and index 2 (i5) and sequences required for cluster formation. It is critical to use the full amount of recommended input DNA and not add extra cycles of PCR cycles to ensure libraries that produce high-quality sequencing results.

Consumables:

Item	Quantity	Storage	Supplied
NPM (Nextera PCR Master Mix)	1 tube	-25°C to -15°C	Illumina
Index 1 primers (N7XX)	1 tube each	-25°C to -15°C	Illumina
Index 2 primers (S5XX)	1 tube each	Room Temperature	Illumina
Microseal 'A' film	-	Room Temperature	User

Preparation

Utilising the samples sheet prepared in Illumina Experiment Manager, index primers will be added to each sample. One N7X and one S5X primer. The Illumina Experiment Manager will verify whether the combination of primers will result in a unique and usable set of primers.

Remove NPM and the index primers (i5 and i7) from -25°C to -15°C storage and thaw on a bench at room temperature. Allow approximately 20 minutes to thaw NPM and index primers. After all reagents are thawed, gently invert each tube 3–5 times to mix and briefly centrifuge the tubes in a microcentrifuge. Use 1.7 ml Eppendorf tubes as adapters for the microcentrifuge.

Amplify NTA

1. Add **15 µl NPM** to each well of the NTA plate containing index primers. Change tips between samples.

Calculate the total volume of NPM for all reactions, and divide the volume equally among the wells of a PCR eight-tube strip. Use a multichannel pipette to dispense into the NTA plate.

2. Add **5 µl** index primer 2 primers **(white caps)** to each well.

3. Add **5 µl** index primer 1 primers **(orange caps)** to each well.

Each primer added as per the sample sheet.

4. Using a multichannel pipette, gently pipette up and down 5 times to mix. Change tips between samples to avoid index and sample cross-contamination.
5. Cover the NTA plate with Microseal 'A' film and seal with a rubber roller.
6. Centrifuge the NTA plate at **280 × g at 20°C for 1 minute**.
7. Perform PCR using the following program on a thermal cycler:
   Make sure that the thermal cycler lid is heated during the incubation.

• 72°C for 3 minutes
• 95°C for 30 seconds
• 12 cycles of:
— 95°C for 10 seconds
— 55°C for 30 seconds
— 72°C for 30 seconds
• 72°C for 5 minutes
• Hold at 10°C

SAFE STOPPING POINT

If you do not plan to proceed immediately to PCR Clean-Up, there are two options for storage. The NTA plate can remain on the thermal cycler overnight or you can store it at 2°C to 8°C for up to two days.

---

3. Size Selection

This step uses AMPure XP beads to purify the library DNA, and provides a size selection step that removes short library fragments from the population.

A double size selection protocol is followed in order to omit fragment sizes that are either too small or too large.

Consumables

Item	Quantity	Storage	Supplied
RSB (Resuspension Buffer)	1 tube	-25°C to -15°C	Illumina
AMPure XP beads	User Determined	2°C to 8°C	Illumina
Freshly prepared 80% ethanol (EtOH)	-	Room Temperature	Illumina
96-well MIDI plates	-	Room Temperature	User

Preparation

1. Bring the AMPure XP beads to room temperature.
2. Prepare fresh 80% ethanol from absolute ethanol.

Always prepare fresh 80% ethanol for wash steps. Ethanol can absorb water from the air impacting your results.

Make CAN

1. Centrifuge the NTA plate at 280 × g for 1 min (20˚C) to collect condensation.

2. Label a new MIDI plate **CAA (Clean Amplified Plate).**

3. Using a multichannel pipette set to **40 µl**, transfer the PCR product from the NTA plate to the CAA plate. Change tips between samples.

4. Vortex the AMPure XP beads for 30 seconds to make sure that the beads are evenly dispersed. Add an appropriate volume of beads to a trough.

5. Using a multichannel pipette, add **23 µl** AMPure XP beads to each well of the CAA plate. (*Note this is a 0.575 ration specific to this protocol*)

6. Gently pipette mix up and down 10 times.

Alternatively the solution can be mixed by shaking the CAA plate on a microplate shaker at 1800 rpm for 2 minutes.

7. Incubate at room temperature without shaking for 5 minutes.

8. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

9. With the CAA plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant carefully. Change tips between samples.

10. With the CAA plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
a) Using a multichannel pipette, add **200 µl freshly prepared** 80% ethanol to each sample well. Do not resuspend the beads yet.
b) Incubate the plate on the magnetic stand for **30 seconds.**
c) Carefully remove and discard the supernatant.

11. With the CAA plate still on the magnetic stand, allow the beads to air-dry for 15 minutes.

12. Elute in **55 µl RBS.**

13. Transfer **50 µl** of DNA to **new plate** (CAA2).

14. Add **37.5 µl** AMPure XP beads *(Note this is a 0.75 ratio specific to this protocol)*

15. Gently pipette mix up and down 10 times.

Alternatively the solution can be mixed by shaking the CAA plate on a microplate shaker at 1800 rpm for 2 minutes.

16. Incubate at room temperature without shaking for 5 minutes.

17. Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

18. With the CAA plate on the magnetic stand, use a multichannel pipette to remove and discard the supernatant carefully. Change tips between samples.

19. With the CAA plate on the magnetic stand, wash the beads with freshly prepared 80% ethanol as follows:
a) Using a multichannel pipette, add 200 µl freshly prepared 80% ethanol to each sample well. Do not resuspend the beads yet.
b) Incubate the plate on the magnetic stand for 30 seconds.
c) Carefully remove and discard the supernatant.

20. With the CAA plate still on the magnetic stand, allow the beads to air-dry for 15 minutes.

21. Elute in **15 µl RBS**, and transfer **10 µl** for Quantification.

Safe Stopping point
The CAN plate can be sealed and stored at -20°C for up to a week.

---

4. Library Normalisation

A non-bead based normalisation will be used for library normalisation.

Quantify ALL eluted DNA samples, and dilute to **2 nM** – Utilising **Qubit HS DNA assay**.

It is important to note however that the base pair length of the DNA molecule will ultimately affect the molarity of the final DNA solution. Thus, in order to obtain a final concertation of 2nM, the following calculations must be performed:

- Obtain concentration via Qubit of DNA in ng/µl.
- With assumption that fragments sizes are 770 bp (0.75 ratio of beads), calculate molecular weight of DNA.
- Calculate molarity based on length of fragments

1) $\left[(No\ of\ BasePairs) \times \left(607{,}4\frac{g}{mol}\right)\right] + 158\frac{g}{mol} = molecular\ weight\ of\ DNA\ fragment$

2) $\frac{Concentration\left(\frac{ng}{\mu l}\right)}{MW\ of\ DNA} \times 10^6 = Concentration\ in\ nM$

Preparation

Prepare 0.2 N NaOH fresh. In this instance, 0.2 N = 0.2 M (For a monoprotic base such as NaOH molar and normal is the same)

Pooling of samples and making Library

1. Once all samples are at an appropriate concentration, all the samples are **pooled** at equimolar concentrations. **5 µl of each sample is added to the pool.** This is now referred to as the **Library.**
2. Combine **5 µl** of the library with **5 µl** of 0.2 N NaOH. (This step denatures the dsDNA to ssDNA. Some alternative Illumina protocols may suggest heating the DNA followed by immediate cooling to prevent re-annealing; in this protocol however, NaOH denaturation is sufficient).
3. Vortex briefly to mix the sample solution, then use benchtop centrifuge to collect sample.
4. Incubate for 5 min at room temperature.
5. Add the following volume of **pre-chilled HT1** to the tube containing denatured DNA:
    - Denatured DNA – **10 µl**
    - Pre-chilled HT1 – **990 µl**

This results in a 10 pM denatured library in 2 mM NaOH.

6. Place on ice until ready to continue to final dilution.

For this protocol, the DNA will be further diluted down to **7 pM.** Thus to obtain this concentration combine the following:

- **420 µl DNA**
- **180 µl HT1**

Invert the tube several times to mix, and pulse centrifuge the solution.

---

5. PhiX preparation

PHIX is a control that is used.

In this protocol, PhiX will be used at 20%, at a concentration of 20 pM.

1. Combine **2 µl** of 10 mM PhiX with **3 µl** 10 mM Tris-Cl, pH 8.5 with 0.1% Tween/Sabax water.

2. Combine **5 µl** diluted PhiX library with 5 µl 0.2 N NaOH.

3. Vortex to mix & use benchtop centrifuge to collect sample.

4. Incubate at room temperature for 5 minutes to allow the library to denature into single strands

5. Add the following volume of pre-chilled HT1 to the tube containing denatured Phix library to result in a **20 pM Phix library**:

- **10 µl** Phix Library
- **990 µl** pre-chilled

This 20 pM Phix library is stable for up to 3 weeks at -15°C to -25°C

6. Mixing library and PhiX library & Loading Sample onto Cartridge

1. PhiX will be added to the library at a 20% concentration. Thus, for a total volume of 600 µl combine **120 µl** of Phix and **480 µl** of the 7 pM pooled DNA Library.

2. Make sure that the MiSeq reagent cartridge is fully thawed and prepared. When the reagent cartridge is ready for use, load the prepared libraries onto the cartridge.

3. Using a separate, clean, and empty 1 ml pipette tip, pierce the foil seal over the reservoir labelled **Load Samples**.
*Do not pierce any other reagent positions. Other reagent positions are pierced automatically during the sequencing run*4. Pipette 600 µl of prepared libraries into the **Load Samples** reservoir. Avoid touching the foil seal as you dispense the sample.

Appendix D: IAVI Protocol C Demographics and DRM Data

1. Demographic and clinical data of the 23 IAVI protocol C participants

Table 1: IAVI Protocol C Clinical and Demographic Data with baseline mutations detected by Sanger sequencing

Participant Number	Gender	Number of Samples	Site	Follow Up Period (Months)	Viral Load (RNA copies/ml) (baseline) (week 1)	CD4 (cells/mm^3)(baseline) (week 1)	ART Start Date	Sub-Type by *pol*	Baseline Mutations		
									RT - NNRTI	**RT** - NRTI	**PR** - PI
1	Male	5	Kigali	48.69	$8{,}34 \times 10^4$	456	2008/06/25	A1	L100I, K103N	_	_
2	Male	7	Kigali	10.32	$1{,}48 \times 10^5$	496	2008/12/01	A1	_	_	M46L
3	Male	5	Kigali	17.03	$1{,}69 \times 10^5$	589	N/A	A1	K103N, Y181CY	_	L10I
4	Male	3	Kigali	**1.31**	$6{,}6 \times 10^4$	899	N/A	C	K103N	_	
5	Male	11	Masaka	**55.03**	$5{,}25 \times 10^5$	459	2011/05/25	D		_	I85V
6	Male	7	Masaka	13.61	$1{,}48 \times 10^5$	490	2014/06/05	A1	K103N		L10V
7	Male	8	Kilifi	48.36	$1{,}6 \times 10^4$	975	N/A	A1	_	D67DN	_
8	Male	11	Kilifi	22.15	$9{,}76 \times 10^5$	310	N/A	A1	K103N	_	L10V
9	Male	11	Kilifi	22.10	$3{,}00 \times 10^6$	386	N/A	A1	K103N	_	L10V
10	Female	11	Kilifi	19.29	$9{,}00 \times 10^6$	327	N/A	A1	Y188C	_	_
11	Male	5	Kangemi	31.59	$1{,}46 \times 10^5$	913	N/A	A1	K103N, E138A	_	_
12	Male	4	Kangemi	24.65	$1{,}41 \times 10^5$	303	2014/04/16	A1	K103N	_	_
13	Male	6	Lusaka	56.38	$5{,}39 \times 10^3$ (week 30)	464 (week 30)	2010/11/12	C	_	_	T74S
14	Male	10	Lusaka	37.51	$3{,}92 \times 10^5$	521	2013/08/22	C	K103N	_	T74S
15	Male	9	Lusaka	20.515	$4{,}52 \times 10^5$	465	N/A	C	K103N, V108IV	M184V	_
16	Female	5	Lusaka	15.81	$1{,}66 \times 10^5$	189	N/A	C	K103N	_	T74S
17	Male	12	Entebbe	65.72	$7{,}32 \times 10^4$	311	2010/10/01	A1	K103N	_	_
18	Female	3	Entebbe	3.52	$6{,}9 \times 10^5$	1079	N/A	A1D	_	_	I85V
19	Female	4	Entebbe	9.73	$6{,}1 \times 10^4$	366	N/A	A1	_	L210W	_
20	Male	7	Entebbe	14.96	$3{,}33 \times 10^5$	362	2009/03/09	D	K103N	_	_
21	Female	5	Entebbe	17.16	$3{,}25 \times 10^4$	532	2010/06/07	D	_	_	M46L
22	Female	6	Cape Town	24.62	$1{,}21 \times 10^5$ (week 40)	495 (week 40)	2010/07/15	C	K103N	_	_
23	Male	5	Ndola	27.25	$7{,}38 \times 10^4$	621	N/A	C	_	M41L	_

2. Drug resistance mutations detected by next generation sequencing for all 23-participant baseline and all follow up samples

Table 2: IAVI Protocol C Drug Resistant Mutation Results Detected for All Samples

Participant Number	Sample Number	Visit Date	RT DRMs - NGS	PR DRMs -NGS	Viral Load (RNA copies/ml)	CD4 Count cells/mm^3
1	1	2007/11/17	K103N (99.84)	-	83400	456
	2	2007/11/27	K103N (99.72)	-	121000	524
	3	2007/12/19	K103N (99.56)	-	238000	360
	4	2008/01/24	K103N (99.68)	-	295000	458
	5	2008/04/17	K103N (99.02)	-	195000	320
2	6	2008/04/02	-	M46L (99.89)	148000	496
	7	2008/04/08	-	M46L (99.92)	37600	242
	8	2008/05/05	-	M46L (99.58)	60300	317
	9	2008/06/02	-	M46L (99.13)	73000	334
	10	2008/08/25	-	M46L (99.85)	51100	305
	11	2008/11/18	-	K20I (1.96), M46L (99.94)	35200	296
	12	2009/02/09	-	M46L (99.95)	26500	254

3	13	2008/05/22	K103N (10.85), Y181C (71.89), Y188C (1.45), G190A (4.47)	L10I (99.75)	169 000	589
	14	2008/06/03	K103N (38.06), Y181C (57.29), Y188C (3.38), G190A (4.40), H221Y (1.47)	L10I (99.89)	1020000	479
	15	2008/07/03	K103N (20.49), Y181C (73.59), G190A (4.63)	L10I (99.84)	78100	522
	16	2008/07/28	K103N (17.93), Y181C (74.48), G190A (7.72), H221Y (4.26)	L10I (99.74)	440000	363
	17	2008/10/21	T69N (7.11), K103N (4.46), Y181C (60.43), G190A (4.74), H221Y (5.33)	L10I (99.89)	350000	250
4	18	2008/06/27	K103N (40.90), K103T (1.31), V106A (1.83), V106M (1.19), V108I (2.96), Y188C (9.38), N348I (1.75)	-	66600	899
	19	2008/07/10	K103N (49.68), V108I (1.44), Y188C (6.15)	-	17600	390
	20	2008/08/05	K103N (95.91), Y188C (1.59)	-	61100	698
5	21	2006/08/22	-	I85V (98.87)	525000	459
	22	2006/11/07	-	I85V (90.44)	125000	514
	23	2007/01/30	-	I85V (81.40)	289000	456
	24	2008/01/03	-	I85V (96.10)	78400	466
	25	2008/06/16	-	M46L (1.29), I85V (97.51)	1160000	443

	26	2008/12/01	-	I85V (90.48)	388000	469
	27	2009/05/19	-	I85V (91.72)	228000	399
	28	2009/11/03	-	I85V (89.72)	330000	406
	29	2010/04/14	-	I85V (87.53)	87400	447
	30	2010/10/05	K103R (1.04)	I85V (72.63)	156000	208
	31	2011/03/22		I85V (61.55)	90377	178
6	32	2011/08/26	K103N (99.15)	L10V (99.89)	148069	490
	33	2011/09/02	K103N (98.31)	L10V (99.57)	127336	537
	34	2011/10/24	K103N (98.39)	L10V (99.51)	15729	524
	35	2012/01/18	K103N (96.78), V118I (1.65)	L10V (99.72)	34220	454
	36	2012/04/11	K103N (98.18), V179T (1.14)	L10V (99.33)	19803	424
	37	2012/07/04	K103N (73.59), G190A (28.85)	L10V (83.35)	25665	365
	38	2013/09/24	K103N (99.42)	L10V (99.80)	-	-
7	39	2006/08/11	**NO AMP**	**NO AMP**	16000	975
	40	2006/10/02	D67N (0.020)	-	10500	1010
	41	2006/12/15	D67N (0.019)	-	39500	688
	42	2007/11/26	D67N (0.017)	-	1770	324
	43	2008/05/02	D67N (0.017)	-	5430	980

	44	2009/09/14	D67N (0.026)	-	2980	1020
	45	2010/03/03	D67N (0.030)	-	2800	490
	46	2010/08/20	D67N (0.029)	-	6100	653
8	47	2010/05/25	K103N (98.97)	L10V (99.67)	165000	620
	48	2010/06/02	K103N (95.94), G190A (1.61)	L10V (99.66)	100000	552
	49	2010/06/14	K103N (99.39), Y188H (21.99)	L10V (95.22)	45700	551
	50	2010/06/28	K103N (99.42)	L10V (99.77)	104000	534
	51	2010/07/09	K103N (99.20), M230I (1.32)	L10V (99.66)	19400	506
	52	2010/08/16	K103N (75.45), G190A (19.69)	L10V (69.16)	10200	596
	53	2010/11/01	K103N (98.90)	I84V (1.96), L10V (57.50)	9760	580
	54	2011/01/24	K103N (96.89), K103S (2.44), K219R (3.90)	L10V (99.66)	21859	-
	55	2011/07/11	K103N (99.15), P225H (1.04)	M46L (2.03), L10V (99.05)	48734	297
	56	2012/01/05	K103N (98.21)	N83D (1.73), L10V (99.71)	319345	424
	57	2012/03/20	K103N (99.27), Y188H (1.92)	L10V (99.66)	28373	465
9	58	2010/11/17	K103N (98.92)	L10V (99.63)	2640000	386
	59	2010/11/24	K103N (98.96)	L10V (99.62)	353714	645

	60	2010/12/03	K103N (99.09)	L10V (99.83)	400178	553
	61	2010/12/15	K103N (98.99)	L10V (99.62)	76663	504
	62	2011/01/12	K103N (99.13)	L10V (99.67)	78795	772
	63	2011/02/09	K103N (98.44)	L10V (99.68)	98128	624
	64	2011/03/23	K103N (97.59)	L10V (84.53)	272579	704
	65	2011/05/05	K103N (86.58), G190A (6.50)	L10V (99.67)	102952	427
	66	2011/07/27	K103N (97.73)	L10V (99.62)	6769	685
	67	2011/10/19	K103N (99.38), H221Y (2.47)	L10V (99.61)	7047	489
	68	2012/01/13	K103N (99.54), M184I (2.31)	L10V (99.85)	3945	400
10	69	2011/01/06	Y188C (99.68)	-	9377185	327
	70	2011/01/12	Y188C (99.66)	-	7176879	218
	71	2011/01/20	Y188C (99.79)	-	289930	-
	72	2011/02/03	Y188C (99.51)	-	794402	521
	73	2011/02/17	Y188C (99.69)	-	20551	-
	74	2011/03/11	Y188C (99.78)	I50V (1.84)	7348	-
	75	2011/03/31	Y188C (99.73)	-	5820	759
	76	2011/05/13	Y188C (99.78)	-	363549	612
	77	2011/12/09	Y188C (99.67)	-	4405	546
	78	2012/03/06	Y188C (99.71)	-	57085	540

	79	2012/08/14	Y188C (39.42)	V82A (6.75)	8445	602
11	80	2010/03/15	K103N (98.79), E138A (99.54)	-	146000	913
	81	2010/04/06	K103N (99.37), E138A (99.74)	-	259000	882
	82	2011/07/28	K103N (82.44), E138A (99.83)	-	25636	696
	83	2012/03/12	K103N (97.08), K103S (1.30), E138A (99.52)	-	176680	654
	84	2012/05/15	K103N (97.99), E138A (99.76)	-	52904	758
12	85	2010/02/02	K103N (98.17)	-	141000	303
	86	2010/05/03	K103N (97.55)	-	169000	253
	87	2011/03/29	K103N (98.16)	-	46473	247
	88	2012/02/21	K103N (97.56)	-	227879	307
13	89	2007/05/03	-	T74S (99.69)	5390	464
	90	2007/07/26	K65R (1.17)	I50V (1.02), T74S (99.67)	1590	664
	91	2008/01/11	K65R (1.01)	T74S (99.68)	3360	654
	92	2009/05/30	K65R (1.23)	T74S (99.61)	45900	265
	93	2010/04/29	K65R (1.18)	T74S (99.63)	2990	179
	94	2010/10/09	K65R (1.35)	T74S (99.61)	38300	204
14	95	2009/05/02	K103N (99.03)	T74S (99.81)	392000	521
	96	2009/05/23	K103N (98.52)	T74S (99.48)	433000	386

	97	2009/06/23	K103N (98.34)	T74S (99.50)	534000	503
	98	2009/09/12	K103N (98.23)	T74S (99.37)	670312	354
	99	2009/12/11	K103N (98.72)	T74S (98.41)	521836	469
	100	2010/02/23	K103N (98.23)	T74S (99.52)	362000	291
	101	2011/05/03	K103N (98.34)	T74S (99.65)	205223	310
	102	2011/07/20	K103N (97.65)	T74S (99.53)	250336	336
	103	2011/12/15	K103N (98.30)	T74S (99.37)	252678	447
	104	2012/06/15	K103N (98.18)	T74S (99.69)	104773	353
15	105	2009/07/25	K65R (1.03), K103N (98.55), V108I (65.64), M184V (99.17)	-	452000	465
	106	2009/08/01	K103N (97.59), V108I (53.92), M184V (97.01), K65R (1.02)	-	217000	463
	107	2009/08/08	K65R (1.08), K103N (97.92), V108I (52.27), M184V (96.63)	-	95473	637
	108	2009/08/15	K103N (98.52), V108I (83.50), M184V (99.34), K65R (1.12)	-	95473	558
	109	2009/08/22	K103N (93.48), V108I (48.71), M184V (76.52), G190A (3.71), K65R (1.02)	-	94172	567
	110	2009/09/05	K103N (96.54), V108I (38.83), M184V (61.98),	-	75619	601

			K65R (1.01)			
	111	2009/09/19	K65R (1.42), K103N (98.17), V108I (4.04), M184V (15.70)	-	57876	445
	112	2009/10/17	K103N (91.42), K103S (6.96), M184V (1.14), K65R (3.24)	-	20411	205
	113	2009/11/28	K103N (73.81), K103S (23.84), G190A (1.60), K65R (1.43)	-	30799	499
16	114	2011/01/25	K65R (1.18), K103N (98.72)	T74S (99.66)	166374	189
	115	2011/02/08	K65R (2.23), K103N (98.65)	T74S (99.82)	52219	156
	116	2011/05/31	K65R (1.18), K103N (98.73)	V11I (1.89), T74S (99.67)	117650	162
	117	2011/10/04	K65R (1.12), K103N (98.63)	V11I (5.77), T74S (99.14)	60370	166
	118	2012/05/19	K65R (1.38), K103N (98.92)	T74S (99.75)	56569	199
17	119	2006/10/10	K103N (99.24)	-	73200	311
	120	2006/11/16	K103N (99.11)	-	29800	381
	121	2007/02/08	K103N (98.85)	-	682000	391
	122	2007/05/03	K103N (99.11)	V82A (1.45)	40100	314
	123	2007/07/27	K103N (99.10), V118I (6.29)	-	6620	362

	124	2008/01/10	K103N (99.02)	-	24300	200
	125	2008/06/26	K103N (99.34)	-	31600	416
	126	2008/12/10	K103N (87.16), K103S (12.21)	M46I (9.78)	4260	409
	127	2009/06/02	K103N (98.27), K103S (1.01)	-	13300	295
	128	2010/04/29	D67G (3.11), K103N (97.88)	-	29200	327
	129	**2010/10/14**	K103N (95.56), K103S (3.69), G190E (1.18)	-	89500	358
	130	2012/03/30	K103N (96.52), K103S (2.78)	-	11107	308
18	131	2006/12/06	V118I (99.50)	I85V (99.13)	690000	1079
	132	2006/12/14	V118I (99.53)	I85V (99.51)	153000	1098
	133	2007/03/22	V118I (99.78)	I85V (96.57)	134000	736
19	134	2007/03/26	L210W (13.84)	V82A (1.25)	61000	366
	135	2007/05/08	D67G (2.30), L210W (28.90)	-	73600	432
	136	2007/10/23	D67G (1.51), L210W (27.65)	-	10600	298
	137	2008/01/15	L210W (92.62)	-	12600	235
20	138	2007/10/16	K103N (99.29)	-	333000	362
	139	2007/11/05	K103N (99.21)	-	242000	287
	140	2007/12/03	K103N (95.70), T215A (14.07)	-	135000	264
	141	2008/02/25	K103N (91.3), K103S (8.23)	-	141000	274
	142	2008/05/08	K103N (47.06), K103S (23.64), V118I (30.48)	-	65900	307

	143	2008/08/12	K103N (91.19), K103S (8.44)	-	121000	248
	144	2009/01/12	K70T (37.09), K103N (99.38), F227L (1.88)	-	26300	384
21	145	2008/08/04	K219R (4.60)	M46L (87.93), I47V (4.94)	32500	532
	146	2008/08/28	K103N (13.86)	M46L (69.02)	60200	605
	147	2008/11/20	G190A (16.56), F227L (2.15)	M46L (32.74)	13800	473
	148	2009/07/24	-	V82A (1.23)	90700	249
	149	2010/01/07			152000	236
22	150	2007/10/18	K103N (0.01)	M46L (1.42)	14400	865
	151	2008/01/08	K103N (93.74), V106M (4.42)	-	182000	692
	152	2008/04/10	K103N (60.266), V106M (36.75)	-	87600	675
	153	2008/09/16	K103N (33.55), K103S (2.09), V106M (16.73)	-	134000	471
	154	2009/03/19	K103N (10.03), V106M (2.93), V118I (11.00)	-	42800	554
	155	2009/09/03	K103N (34.36), K103S (19.55), V106M (33.59)	-	40300	665
23	156	2009/08/12	M41L (99.67), K103R (99.46)	-	73800	621
	157	2009/08/26	M41L (99.62), K103R (99.41)	-	6520	593
	158	2011/06/01	M41L (99.77), K65R (1.08), K103R (99.37)	L90M (1.73)	3097	468
	159	2011/07/01	M41L (99.49), K103R (99.29)	-	2436	-

	160	2011/11/18	M41L (99.49), K103R (99.44)	-	1017	447

Appendix E: Turnitin Report

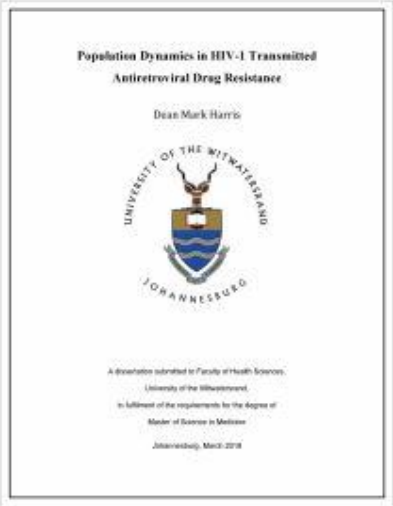
turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Dean Harris
Assignment title:	Dean H
Submission title:	441554:Dean_Harris_FINAL.docx
File name:	ents_015013af-4c2f-45bb-8ada-4...
File size:	8.17M
Page count:	151
Word count:	43,933
Character count:	246,607
Submission date:	19-Mar-2018 05:31PM (UTC+0200)
Submission ID:	932670937



Population Dynamics in HIV-1 Transmitted Antiretroviral Drug Resistance

Dean Mark Harris

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

A dissertation submitted to Faculty of Health Sciences, University of the Witwatersrand, In fulfilment of the requirements for the degree of Master of Science in Medicine

Johannesburg, March 2018

Copyright 2018 Turnitin. All rights reserved.

441554:Dean_Harris_FINAL.docx

ORIGINALITY REPORT

16%	14%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	tools.thermofisher.com Internet Source	2%
2	www.liai.org Internet Source	2%
3	Submitted to University of Witwatersrand Student Paper	1%
4	web.uri.edu Internet Source	<1%
5	eprints.ucm.es Internet Source	<1%
6	open.uct.ac.za Internet Source	<1%
7	repub.eur.nl Internet Source	<1%
8	scripties.umcg.eldoc.ub.rug.nl Internet Source	<1%
9	www.ncbi.nlm.nih.gov Internet Source	<1%